Book of Proceedings

of the 7th edition of the International Conference on Quality Engineering and Management

ICQEM 2026

TECHNICAL RECORD

Title

Proceedings book of the 7th International Conference on Quality Engineering and Management, 2026

Authors/Editors

Carvalho, André; Sampaio, Paulo; Domingues, Pedro; Salvadorinho, Juliana; Morgado, Teresa; Casadesús, Martí; Marimon, Frederic; Pires, António Ramos; Saraiva, Pedro

Publisher

International Conference on Quality Engineering and Management

Date

July 2026

Cover Design

ICQEM

ISBN

978-989-54911-3-1

ISSN

2184-3481

1st Edition

200 units

This edition is published by the International Conference on Quality Engineering and Management.

Portuguese National Library Cataloguing in Publication Data

Proceedings book of the 7th International Conference on Quality Engineering and Management edited by Carvalho, André; Sampaio, Paulo; Domingues, Pedro; Salvadorinho, Juliana; Morgado, Teresa; Casadesús, Marti; Marimon, Frederic; Pires, António Ramos; Saraiva, Pedro

ISBN 978-989-54911-3-1

ISSN 2184-3481

Publisher: International Conference on Quality Engineering and Management

Book in 1 volume, 520 pages

This book contains information obtained from authentic sources.

Reasonable efforts have been made to publish reliable data information, but the authors, as well as the publisher, cannot assume responsibility for the validity of all materials or for the consequences of their use.

Neither this book nor any part may be reproduced or transmitted in any form or by the means, electronic or physical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without prior permission in writing from the International Conference on Quality Engineering and Management Chair.

All rights reserved.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

International Conference on Quality Engineering and Management

Paulo Sampaio - Conference Chair

School of Engineering, Systems and Production Department

Campus Gualtar, 4710-057 Braga, Portugal
http://icqem.dps.uminho.pt/
icqem@dps.uminho.pt

© 2026 by International Conference on Quality Engineering and Management
ISBN 978-989-54911-3-1
ISSN 2184-3481

Organizing Committee

Chair

Paulo Sampaio, University of Minho, Portugal

Local Chair

André M. Carvalho, NOVA School of Science and Technology, Portugal

Members

Ana Sofia Matos, NOVA School of Science and Technology, Portugal

Ana Teresa Gabriel, NOVA FCT, Portugal

Helena Carvalho, NOVA FCT, Portugal

Juliana Salvadorinho, NOVA FCT, Portugal

Paulo Sampaio, University of Minho (UMINHO), Portugal

Pedro Domingues, UMINHO, Portugal

Pedro Espadinha da Cruz, NOVA FCT, Portugal

Radu Godina, NOVA FCT, Portugal

Rogério Puga Leal, NOVA FCT, Portugal

Sofia Pereira, University of Minho, Portugal

Susana Duarte, NOVA FCT, Portugal

Teresa Morgado, Lisbon Higher Institute of Engineering (ISEL), Portugal

Victor Azamfirei Ionita, NOVA FCT, Portugal

Scientific Committee

Aida Szilagyi, Polytechnic University of Timisoara (UPT), Romania

Aleksandar Anastasovski, International Balkan University (IBU), Skopje, North Macedonia

Aleksandar Erceg, J.J. Strossmayer University of Osijek (UNIOS), Croatia

Leonor Teixeira, University of Aveiro (UA), Portugal

Lígia Conceição, Polytechnic Institute of Setúbal, Portugal

Luís Lourenço, University of Beira Interior (UBI), Portugal

Ana Rita Oliveira, South East Technological University (SETU), Ireland

Angela Sofia Leal Neves, DEMGI, Polytechnic Institute of Viseu (IPV), Portugal

Antonio Henriques de Araújo Junior, State University of Rio de Janeiro (UERJ), Brazil

Antonio Padovano, DIMEG, University of Calabria (UNICAL), Italy

Apostolos Michopoulos, University of Cyprus (UCY), Cyprus

Attila Kajos, Corvinus University of Budapest (UNI Corvinus), Hungary

Biljana Činčurak Erceg, J.J. Strossmayer University of Osijek (UNIOS), Croatia

Carina Maria Oliveira Pimentel, University of Minho (UMINHO), Portugal

Duško Pavletić, University of Rijeka, Faculty of Engineering (UNIRI), Croatia

Duarte Dinis, NOVA FCT, Portugal

Eduard Laurențiu Nițu, National University of Science and Technology POLITEHNICA Bucharest (UPB), Romania

Elvis Krulčić, University of Rijeka, Faculty of Engineering (UNIRI), Croatia

Emmanuel Francalanza, University of Malta (UM), Malta

Estela Vilhena, Polytechnic of Cávado and Ave (IPCA), Portugal

Evandro G. Lorentz, Pontifical Catholic University of Minas Gerais (PUC Minas), Brazil

Márcio Machado, Universidade Paulista (UNIP), Brazil

Margarida Saraiva, Universidade de Évora (UEVORA), Portugal

Maria de la Cruz del Río-Rama, Universidade de Vigo (UVigo), Spain

Maria João Rosa, University of Aveiro (UA), Portugal

Mariana Palhau, University of Porto (UP), Portugal

Martí Casadesus Fa, University of Girona (UdG), Spain

Maurizio Galetto, Politecnico di Torino (Polito), Italy

Merce Bernardo, University of Barcelona (UB), Spain

Mohammad Hossein Zavvar Sabegh, Organizational Excellence Specialists, Canada

Patrícia Moura e Sá, University of Coimbra (UC), Portugal

Paul Refalo, University of Malta (UM), Malta

Paulo A. Cauchick Miguel, Federal University of Santa Catarina (UFSC), Brazil

Pawel Krzeminski, Norwegian Institute for Water Research (NIVA), Norway

Pedro Alexandre Marques, Universidade Lusófona (ULusofona), Portugal

Pedro Saraiva, University of Coimbra (UC), Portugal

Peter Hines, South East Technological University (SETU), Ireland

Martha Ramírez-Valdivia, Universidad de La Frontera, Chile

Fernando Costa, Universidade Estadual Paulista (UNESP), Brazil

Fiorenzo Franceschini, Politecnico di Torino (Polito), Italy

Foivos Psarommatis, University of Oslo (UiO), Norway

Frederic Marimon, International University of Catalonia (UIC), Spain

Gaia Vitrano, Politecnico di Milano (Polimi), Italy

Hasan Volkan Oral, İstanbul Aydin University (AYDIN), Türkiye

Helena Alvelos, University of Aveiro (UA), Portugal

Henriqueta Nóvoa, University of Porto (UP), Portugal

Hugo Ferreira, Polytechnic Institute of Viseu (IPV), Portugal

Iñaki Heras-Saizabitoria, University of the Basque Country (EHU), Spain

Isabel Lopes, University of Minho (UMinho), Portugal

Jacob Kjær Eskildsen, Aarhus University (AU), Denmark

José Álvarez-Garcia, Universidad de Extremadura (UNEX), Spain

José António Cabral, University of Porto (UP), Portugal

Josep Llach Pagès, University of Girona (UdG), Spain

Juan José Tarí, University of Alicante (UA), Spain

Lance Coleman, IDEX Health & Science (IDEX-HS), USA

Roberto A. Martins, Federal University of São Carlos (UFSCar), Brazil

Samruddha Kokare, Katholieke Universiteit (KU) Leuven, Belgium

Sandro Doboviček, University of Rijeka, Faculty of Engineering (UNIRI), Croatia

Sergio Sousa, University of Minho, Portugal

Shubham Sharma, Shoolini University, India

Sotirios Panagou, Norwegian University of Science and Technology (NTNU), Norway

Tse Chiu (Andy) Wong, University of Strathclyde (STRATH), United Kingdom

Vasiliki (Vicky) Skoulou, University of Hull (HULL), United Kingdom

Virgílio Cruz-Machado, NOVA School of Science and Technology (NOVA FCT), Portugal

Vittorio Solina, University of Calabria (UNICAL), Italy

Yasemin Ülker, Bahcesehir University (BAU), Türkiye

FOREWORD

It is our great pleasure to welcome you to Lisbon, Portugal, and to NOVA University Lisbon, the venue of the seventh edition of the International Conference on Quality Engineering and Management (ICQEM). Since its inception, ICQEM has established itself as a distinguished forum for the dissemination of research, the exchange of ideas, and the advancement of knowledge in the fields of quality engineering and quality management.

The 7th International Conference on Quality Engineering and Management (ICQEM26) brings together scholars, practitioners, and industry leaders to explore the evolving role of quality in engineering, management, and organizational systems. Building on the success of previous editions, this year's conference will broaden its scope to address how quality principles can be applied to the growing complexity of technological, organizational, and societal challenges.

Quality engineering and quality management are no longer confined to traditional manufacturing or service environments — they represent critical disciplines for navigating uncertainty, driving innovation, and ensuring resilience in a rapidly changing world. By adopting a systems perspective, the 7th ICQEM seeks to highlight how quality can serve as the connective framework across diverse fields such as digital transformation, healthcare, sustainability, supply chain management, and organizational design.

The conference will provide a unique interdisciplinary platform for sharing research insights, practical experiences, and forward-looking strategies. Participants will engage in discussions on emerging methodologies, integration of advanced technologies, and the role of leadership in fostering a culture of excellence. With a focus on both rigorous academic contributions and real-world impact, ICQEM26 emphasizes the power of quality to bridge disciplines and tackle complex engineering and management challenges.

At the 7th ICQEM, our aim is to not only advance the theory and practice of quality but also to reaffirm its broader mission: building systems that are effective, efficient, sustainable, and human-centered.

NOVA University, July 2, 2026.
Sincerely,

Conference Local Chair
André Carvalho, NOVA University

Conference Chair
Paulo Sampaio, University of Minho

PAPERS INDEX

ISO Standards from Compliance to Impact - A Systematic Approach for Organizations Mohamed Aichouni and Ahmed Baha Eddine Aichouni	1
Development of an operational excellence model for research centres organization and management Aline Barros, Rui M. Sousa and José Dinis-Carvalho	11
Towards Zero Defect and Zero Waste Food Systems: Extending ZDM to Controlled Environment Agriculture Victor Azamfirei Ionita and Foivos Psarommatis	22
Acceptance sampling by variables for double specification limits: save space, and compute on the Web Miguel Casquilho, Francisco Rolo, Pedro P. Santos and Rui Galhano	31
Industry 4.0 Maturity Models: A Systematic Analysis Larissa Freire Lekakis, Fabio Neves Puglieri and Camila Fabrício Poltronieri	40
Integrating Cost Management into Maturity Models for Public Healthcare Systems: A Framework of Dimensions and Competencies Tiago Rocha, Naja Santana and Larissa Araújo	50
Where to work outside the home: mapping coworking spaces and evaluating their quality in the context of remote work Lorena Bastos, Marlene Amorim, Mário Rodrigues and Cláudia Silva	58
Digitalising Quality Audits to Enhance Data Management: Evidence from a Metallurgical Company Mariana Pinto, Leonor Teixeira and Helena Alvelos	69
Aggregation Sequence in Multiple-Inspector Sampling Plans: Preliminary Outcomes and Practical Implications Domenico A. Maisano, Lucrezia Ferrara and Fiorenzo Franceschini	77
A Data-Driven Workforce Dimensioning Tool for Grocery Retail: A Kaizen Case Study João Ferreira, André Carvalho and Radu Godina	88
The New Construction Products Regulation (EU) 2024/3110 and the challenges of digitalisation in Quality 4.0 Ana Amaral	98
Leveraging Interdeparture Times for Bottleneck Detection and Process Improvement Pedro Costa, Ana Moura, Galina Robertsone and Tatjana Tambovceva	105
A Decision Support Model for Problem-Solving Framework Selection in Industry 4.0 Environments Sofia Mendes Pedro, Juliana Salvadorinho and Ana Sofia Matos	115
Edge Vision for Monitoring Manual Assembly Motions: Robust Calibration and Sensitivity Analysis Sandro Dobovičeк, Elvis Krulčić and Duško Pavletić	125
Investigating the impact of lean digital transformation on sustainability – The role of top management and green practices in the pharmaceutical industry Evgenia Saltaferi, Maria Gianni and Argyro Fragkaki	135
Understanding Lean Manufacturing and Industry 4.0 integration: the interplay of sociotechnical structures João José Medeiros Alfineto, Kleber Francisco Esposto, Fabiane Letícia Lizarelli, Susana Duarte, Igor Roberto Amancio and Lígia Lobo Mesquita	143

Use of Lean Tools to Propose Improvements in the Transport Management Process of a Public Higher Education Institution Marcelo Rodrigues, Fabiane Lizarelli, João Alfineto and Gilberto Ganga	153
The impact of ISO 9001 implementation on transport and logistics processes Andreea Madalina Alexandru and Irina Severin	163
Quality Management System of forensic genetics laboratories in Brazil: a structured guideline of quality management and improvement program Laryssa Silva de Andrade Bezerra, Advaldo Carlos Souza Neto, Maico Roris Severino, Fabiane Letícia Lizarelli e Camila Fabricio Poltronieri	171
The Integration of Big Data Analytics and Lean Six Sigma: A Systematic Literature Review Rafaela Gaeta, Antonio Almeita Junior, João Alfineto and Fabiane Lizarelli	181
Performance Assessment and Competencies in Vocational Education: A Bibliometric Analysis Cleiton Abadia Duarte, Lillian do Nascimento Gambi, Raquel Mello and Camila Fabricio Poltronieri	192
A Dual Perspective on Service Quality: Drivers of Choice and Satisfaction in Hotel Event Services Sofia Lopes, Filipa Brandão, Marlene Amorim and Cláudia Silva	202
A Quality by Design Framework for Mobile-Based Household Food Waste Reduction Alberto Piovano, Rossella Palermo, Elisa Verna and Maurizio Galetto	211
Quality Challenges in Multi-Human Multi-Robot Collaborative Assembly Design: A Coherence-Based Methodology to Analyse Process Failures Matteo Capponi, Aurora Sofia di Matteo, Luca Mastrogiacomo and Fiorenzo Franceschini	221
Understanding Assembly Process Quality through Product Complexity and Cognitive Demand: Preliminary Experimental Evidence Riccardo Gervasi, Laura Angioletti, Michela Balconi, Luca Mastrogiacomo and Fiorenzo Franceschini	231
Sustainability Assessment of Photovoltaic Solar Parks: A Four-Dimensional Delphi-Based Approach Maria de Fátima Cabral and Adelina Baptista	241
Preparing the Workforce for Digital Quality Management: A Comparative Analysis of SAP Implementation Strategies in Two Luxury Goods Manufacturers Micael Gonçalves, Marta Ribeiro, Joana Oliveira, Ana Lobarinhas, Sérgio Sousa, Senhorinha Teixeira and Pedro Domingues	251
The future of quality in manufacturing industry: From Quality 4.0 to Quality 5.0 Inês Rocha Melo, Pedro Marques, Graça Silva and Pedro Saraiva	260
Reusing of Metallic Pieces as a Case Study for a Continuous Sustainability Improvement Process in Luxury Leather Goods Industry Micael Gonçalves, Ana Lobarinhas, Sérgio Sousa, Senhorinha Teixeira and Pedro Domingues	267
Total Quality Management Modernized Definition for the Digital Era: A Delphi Validation Ahmed Baha Eddine Aichouni	275
Micro-Foundations of Sustainable Tourism Performance: A Conceptual Quality Management Framework of Vocational Competencies and Green Innovation Saud A. Alshamari and Ahmed Baha Eddine Aichouni	286

Predictive Model for Detecting Fractures in Railway Tracks Using Machine Learning Techniques in an Imbalanced Dataset Marcelo Oliveira, Paulo Sampaio and Ana Belem	297
Beyond Outcomes: Reframing Value-Based Evaluation of Multimorbidity Healthcare Services Mariana Peyroteo, André Mendes de Carvalho and Luís Velez Lapão	308
Machine Learning and Failure Mode and Effects Analysis Integration in Manufacturing: A Systematic Literature Review Francolino Agostinho Sanca, José Pedro Teixeira Domingues, Ana Maria A. C. Rocha and Emerson José de Paiva	318
Jishu-ken: The Autonomous Self-Study Group as the Structural Driver of Lasting Organizational Change Eduardo L. García-Garcia	327
A Process-Based Measurement Framework for Teaching Quality in Higher Education: Development of a Composite Index (TQI) and Illustrative Application Aneta Kucińska - Landwójtowicz and Izabela Czabak – Górska	338
Improving the Quality of Patient Post-Discharge Care using a Machine Learning–Based Decision Support System Matilde M Santos, Mariana Peyroteo and Luís Velez Lapão	350
Enhancing Value Stream Mapping through Low-Cost Digital Data Collection: An Industrial Case Study in Cork Stopper Production Miguel Moreira and Hermilio Vilarinho	357
Subsystem-oriented condition monitoring for connected vehicles using unlabelled data Stefano Puttero, Elisa Verna, Gianfranco Genta and Maurizio Galetto	368
Preliminary Framework for Verifying the Syntactic Quality of Digital Voice-of-Customer for Product Quality Monitoring Federico Barravecchia, Luca Mastrogiacomo and Fiorenzo Franceschini	378
Analysis of Non-Conformities Raised in ISO 9001 Audits in Portuguese Municipal Chambers Roberth Pereira, Pedro Domingues and Carina Pimentel	388
Measurement System Analysis and Robustness Assessment of an AI-Based Vision Inspection System: Evidence from Automotive Production Frederico Santos, Eduardo e Oliveira, Hermilio Vilarinho and Henriqueta Nóvoa	397
Performance Measurement Systems in Continuous Improvement Programs: A Systematic Literature Review Luiz Geraldo Goessler and Roberto Antonio Martins	406
Between Compliance and Excellence: A Diagnosis of Quality Culture in Portuguese Higher Education through External Evaluation Reports Sílvia Santos, Patrícia Moura e Sá and Leonor Pais	416
Standardisation as a strategic motor for interoperability in BIM Pedro Oliveira	427
Using Design Methods for the Development of a New Framework for Root Cause Analysis Matthew Barsalou, André Carvalho and Rogério Puga-Leal	435
Effectiveness of Training Programmes for Software Engineers in Digital Transformation Contexts Georgiana-Alexandra Voicu and Irina Severin	442
Professional Training in Lean Six Sigma: An analysis of Yellow Belt and Green Belt Programs Mónica Puertas, Edgar Yalta and José Carlos Flores-Molina	450

Transformational Leadership and ISO/DIS 9001:2026: Are the Quality Managers transformational leaders? Noémia Almeida, Cláudia Sousa Silva and Magda Monteiro	461
The impact of lithium mining in Portugal: chemometric studies Inês Timóteo, Susana Jesus, Marta Martins, Inês Coelho and Ana Matos	472
From Validated Problem to First Revenue: A Quality-Grounded Design Science Research Methodology for Early-Stage Startups Luis Alberto González, Valeria de Castro and Esperanza de Marcos	480
Governing the Angenic Shift: A Quality Governance Framework for Sovereign AI in Industrial Organizations Eduardo L. García-Garcia	490
Quantitative Detectability in FMEA: A Top-Down Functional Approach to Risk Prioritization Filipe Alberto Chindula, Tiago Alexandre Narciso Silva	504
Assessment of the compliance of gas cylinder manufacturing company's laboratories with ISO/IEC 17025:2017 requirements Margarida Jesus, Hermilio Vilarinho	512
Evaluating Tourism Service Quality: A Comparative Analysis of Supply Standards and Consumer Expectations Raffaele Tardio, Chiara Cardona, Sofia Menestò, Silvia Verdura, Kasra Hasani and Francesco Picin	519
Quality 4.0 Unlocked: an implementation roadmap for its progressive adoption in Portuguese companies Diana Oliveira, Maria João Rosa and Helena Alvelos	528
A Discussion on Quality Definitions for Artificial Intelligence Systems André M. Carvalho	536
A Framework for Quality Assessment in Service SMEs Based on a Process Weighting Model Juan Canela, Esperanza Marcos, Valeria de Castro, Sofía Canela and Jaime Dionisio	541
Professional Training in Lean Six Sigma: **An analysis of Yellow Belt and Green Belt Programs** Mónica Puertas, Edgar Yalta, José Carlos Flores-Molina	548
Real-Time Visual Inspection System for Automotive Chassis: development of a framework in a Toyota Plant Marco Silva, Ana Raquel Xambre, Helena Alvelos and Carlos Rodrigues	559
Complementarity of Concepts: The Synergies between Society 5.0, Industry 5.0 and Quality 5.0 Flaviane Moblicci, Paulo Sampaio and André M. Carvalho	567

SUBMITTED PAPERS

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

ISO Standards from Compliance to Impact - A Systematic Approach for Organizations

Mohamed Aichouni [1], and Ahmed Baha Eddine Aichouni [2]

[1] Industrial Engineering Department, College of Engineering, University of Hail, Hail, Saudi Arabia; m.aichouni@uoh.edu.sa
[2] CEMMPRE, ARISE, Department of Mechanical Engineering, University of Coimbra, 3030-788 Coimbra, Portugal; aaichouni@student.uc.pt

***Abstract* — Sustainable development has become a strategic priority for organizations operating in increasingly regulated and competitive environments. This study examines the role of international ISO standards as strategic drivers of sustainability and organizational performance. Using the PRISMA 2020 protocol, a systematic review of the quality management and sustainability literature was conducted across major academic databases. Qualitative content analysis and an Importance–Performance Analysis were applied to synthesize and evaluate the practical relevance of the included studies. The review identified 42 individual high-impact international ISO standards, organized into six capability domains. The organizations capability domains are core management systems, quality tools and methods, people and organizational culture, sustainability and social responsibility, governance and compliance, and innovation and digital transformation. The paper proposes the implementation of international standards of core management systems such as ISO 9001, ISO 14001, and ISO 45001 with the integration in organizational practice of strategically critical standards, including ISO 26000, ISO 37001, and emerging AI governance frameworks ISO/IEC 42001. The proposed model suggests that the 42 ISO standards function not merely as compliance mechanisms, but as integrated strategic capabilities embedded within a national quality infrastructure. The study provides a roadmap for leveraging international standards to strengthen organizational resilience, through performance improvement, fostering innovation, reducing risks and hence contributing to the UN Sustainable Development Goals.**

Keywords – International Standards, Sustainable Development Goals, Organizational Performance, Systematic Review.

I. INTRODUCTION

Organizations currently operate under stringent expectations to institutionalize responsible, transparent, and sustainable practices that reconcile operational performance with environmental and social accountability (Denicolai, Zucchella, & Magnani, 2021). Sustainability no longer exists as a peripheral initiative; it constitutes a core strategic imperative directing corporate and macroeconomic objectives toward sustained value creation. Consequently, international frameworks established by the International Organization for Standardization (ISO) serve as the primary mechanisms for operationalizing these theoretical sustainability constructs (Moktadir et al., 2020; Womack & Jones, 1996). Containing a repository of over 40,000 active global standards, the ISO ecosystem provides codified, highly structured methodologies that allow organizations to systematize management systems, ensure statistical quality control, mitigate systemic risks, and drive empirical improvements in sustainability performance (Barni et al., 2018; Tasleem et al., 2019).

Core frameworks, including ISO 9001 (Quality Management), ISO 14001 (Environmental Management), ISO 45001 (Occupational Health and Safety), ISO 50001 (Energy Management), and ISO 26000 (Social Responsibility), establish a baseline architecture for embedding sustainable operations across diverse industrial sectors (Arana-Landín et al., 2024; Croft, 2014; Missimer et al., 2014). Institutionalizing these standards enables consistent process reliability, optimized resource utilization, improved workplace safety, and structured societal engagement. When deployed systematically, these frameworks advance the United Nations Sustainable Development Goals (SDGs) by anchoring micro-level operational processes to quantify macroeconomic, ecological, and sociological indicators (Ronalter, 2023; Mosgaard & Kristensen, 2023). Within the specific geopolitical and economic context of developing countries, adopting these ISO standards directly supports the strategic mandates of future visions, facilitating economic diversification, structural organizational excellence, and long-term sustainability and resilience (Aichouni et al., 2020).

To address existing analytical limitations, this study introduces a rigorous six-category conceptual framework that classifies ISO standards based on their precise functional contributions to sustainability parameters. Derived through a structured systematic literature review and empirical synthesis, the taxonomy categorizes standards into: (1) Core Management Systems, (2) Quality Tools and Methods, (3) People and Culture, (4) Sustainability and Social Responsibility, (5) Governance and Compliance, and (6) Innovation and Digital Transformation. Each defined classification isolates a distinct operational capability area, spanning from variance reduction and workforce competency to ethical compliance algorithms and digital infrastructure deployment. This integrated theoretical framework establishes a robust construct validity for evaluating how aggregated ISO mechanisms facilitate sustainable development and continuous process improvement (Nicoletti Junior & de Oliveira, 2019; Rafique et al., 2019; Shin et al., 2018).

Despite the prevalence of literature examining the isolated effects of specific ISO standards on organizational performance metrics, the existing body of knowledge suffers from severe methodological fragmentation and narrow epistemological scoping (Latan et al., 2020; Rezak et al., 2022; Schwartz & Tilling, 2009). Previous research designs predominantly evaluate single standards, such as ISO 9001 or ISO 14001, confined within monolithic industry strata, thereby generating context-bound data that fails to capture the cumulative operational impact of integrated management systems

(Louw & Droomer, 2019). This isolated analytical approach compromises the empirical robustness required to understand cross-functional interactions among economic, environmental, and social dimensions. Furthermore, the literature lacks mathematically grounded, review-based syntheses that categorize ISO standards by strategic utility or measure their comparative implementation efficacy. Consequently, the macro-level contribution of these standards remains theoretically underdeveloped and practically under-deployed, creating significant adoption disparities in rapidly industrializing economies like Saudi Arabia (Aichouni et al., 2023; Al-Tit, Omri, & Euchi, 2019).

Rectifying these epistemological deficits requires an evidence-based operationalization of the standards most critical to sustainability deliverables, determining their multivariate relationships and establishing empirically justified implementation priorities. Therefore, this research executes three distinct objectives. First, it systematically synthesizes the peer-reviewed literature detailing the impact of ISO standards on the tripartite economic, environmental, and social dimensions of sustainability. Second, it constructs a formalized taxonomy that segments ISO standards into six defined operational vectors. Third, it deploys an Importance-Performance Analysis (IPA) to quantify the discrepancy between the perceived strategic utility of these standards and their actualized implementation rates, isolating high-yield frameworks that remain statistically marginalized in practice (Aichouni et al., 2023).

Fulfilling these parameters provides measurable contributions to organizational theory and industrial practice. The resulting data conceptualizes ISO standards not as isolated checklists for compliance, but rather as interdependent variables within a National Quality Infrastructure (NQI), enabling organizational capabilities building for sustainability-driven market competitiveness. For strategic decision-makers, this study yields a deterministic matrix to synchronize ISO framework adoption with the UN SDGs and the precise operational excellence targets sought through ambitious national transformation visions.

II. RESEARCH METHODS

This research operationalizes a Systematic Literature Review (SLR) integrated with an Importance-Performance Analysis (IPA) as its primary methodological architecture. The SLR protocol functions as a highly structured, replicable mechanism designed to aggregate, critically appraise, and synthesize empirical data. Within the disciplines of quality engineering and macro-sustainability, narrative literature reviews are highly susceptible to selection bias and idiographic limitations. The SLR methodology neutralizes these threats, generating high construct validity and producing reproducible epistemological baselines (Kush et al., 2021; Merli et al., 2018).

The execution strictly adheres to the five-stage analytical heuristic formalized by Denyer and Tranfield (2009): (1) formulation of precise research inquiries, (2) systematic isolation of relevant empirical data, (3) application of rigorous quality assessment criteria, (4) standardized data extraction, and (5) structural synthesis. To eliminate subjective variance during study selection, the screening parameters were governed by a multidisciplinary panel of subject matter experts in Quality Management, Systems Engineering, and Organizational Excellence. The panel utilized a Delphi-based consensus mechanism to establish the inclusion/exclusion criteria and applied a double-blind review process for data extraction, achieving a high inter-rater reliability threshold necessary for scientifically valid syntheses.

Subsequent data analysis utilized a deterministic qualitative content analysis, replicating the integrated framework modeling procedures established by Senivongse, Bennet, and Mariano (2017). This analytical strategy facilitated the isolation of recurring operational variables, causal mechanisms, and theoretical constructs, generating an evidence-based taxonomy of how ISO standards actively drive sustainable development.

A. The SLR Protocol and Search Strategy

The review architecture was dictated by three fundamental research inquiries. First, identifying the specific ISO frameworks that demonstrate statistically significant, peer-reviewed correlations with sustainability performance metrics between 2000 and 2025. Second, constructing a formalized taxonomy that categorizes these standards based on their distinct functional properties. Third, measuring the variance between the theoretical utility of these standards and their actualized implementation rates across industries.

Data acquisition relied upon a highly structured Boolean search syntax executed across four primary scientific databases: Web of Science, Scopus, Semantic Scholar, and Google Scholar. The temporal boundary of 2000–2025 was enforced to capture the modern evolutionary cycle of standardized sustainability compliance, correlating with the global institutionalization of the UN Sustainable Development Goals. Table 1 details the exact search parameters and the resulting gross record retrieval.

TABLE I. DATABASES, ADVANCED BOOLEAN SEARCH SYNTAX, AND INITIAL YIELD

Database	Algorithmic Search Strings / Boolean Syntax	Time Frame	Gross Retrieval
Web of Science	TS=("ISO standard*" AND ("sustainability" OR "sustainable development" OR "SDG*"))	2000–2025	523
Scopus	TITLE-ABS-KEY(("ISO 9001" OR "ISO 14001" OR "management system*") AND "sustainable development")	2000–2025	614
Google Scholar	allintitle: ("ISO" OR "International Organization for Standardization") AND "sustainability"	2000–2025	472
Semantic Scholar	("ISO standards") AND ("environmental performance" OR "social responsibility")	2000–2025	392
Total			**2,001**

The algorithmic search generated a gross retrieval of 2,001 records. To ensure appropriate evidence-informed synthesis and replicability, the data filtration process is mapped according to the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standard.

Following automated algorithmic deduplication, the baseline corpus was established at 1,034 unique manuscripts. The first screening phase eliminated 481 records that lacked English translations, were non-peer-reviewed (e.g., editorial prefaces), or contained incomplete abstracts. A subsequent eligibility assessment removed 97 studies due to severe methodological weaknesses or failure to meet the expert panel's quality threshold. The final inclusion phase required full-text extraction to confirm a direct, empirically validated linkage between an ISO standard and a quantifiable sustainability outcome. This rigorous parameter excluded 406 tangential papers, isolating a core subset of 50 definitive studies for the final structural synthesis. The mathematical progression of this protocol is detailed in Table 2.

TABLE II. PRISMA 2020 DATA FILTRATION AND RETENTION MATRIX

PRISMA Stage	Operational Description	Input Records	Excluded	Retained	Cumulative Retention Rate
Identification	Gross retrieval across four databases	-	-	2,001	100.0%
Deduplication	Algorithmic removal of duplicate entries	2,001	967	1,034	51.7%
Screening	Title/Abstract evaluation (Language/Format constraints)	1,034	481	553	27.6%
Eligibility	Methodological quality and relevance assessment	553	97	456	22.8%
Inclusion	Full-text extraction confirming empirical linkages	456	406	50	2.5%

B. The Importance-Performance Analysis (IPA) Protocol

To operationalize the qualitative data extracted during the SLR phase, the isolated ISO frameworks were subsequently subjected to an Importance-Performance Analysis (IPA). Developed as a deterministic diagnostic mechanism, the IPA methodology provides a quantitative evaluation of organizational resource allocation by measuring the variance between the theoretical utility of a variable and its empirical execution. Within this study, the IPA plots the identified ISO standards on a two-dimensional Cartesian coordinate system. The vertical Y-axis quantifies the perceived strategic *Importance* of a specific standard in achieving sustainability targets, while the horizontal X-axis measures its *Performance*, defined here as the actualized degree of implementation and maturity within organizational structures. By calculating the mean coordinates for each standard, the frameworks are distributed across a four-quadrant matrix. This spatial distribution isolates high-yield standards that suffer from implementation deficits, translating theoretical literature into actionable operational intelligence (Aichouni et al., 2023). To ensure precise interpretation of the data, the coordinate geometry of the four quadrants is strictly defined in Table 3.

TABLE III. IMPORTANCE-PERFORMANCE ANALYSIS (IPA) MATRIX

Quadrant	Coordinate Mapping	Diagnostic Classification	Strategic Operational Directive
Quadrant I	High Importance / High Performance	Institutionalized Excellence ("Keep up the good work")	Maintain current resource allocation; utilize as anchor systems for further integration.
Quadrant II	High Importance / Low Performance	Critical Implementation Deficit ("Concentrate here")	Immediate priority for strategic intervention. These frameworks yield high sustainability returns but remain statistically underutilized.
Quadrant III	Low Importance / Low Performance	Marginal Utility ("Low priority")	Withhold resource deployment. These variables offer minimal structural impact on macro-sustainability targets.
Quadrant IV	Low Importance / High Performance	Resource Misallocation ("Possible over-investment")	Divest or redirect operational focus. Current implementation rates exceed the standard's functional value to the system.

The fusion of these two methodologies, the construct validity established by the PRISMA-compliant SLR, coupled with the quantitative diagnostic capability of the IPA matrix, produces a suitable for exploratory model development. This dual-method architecture prevents analytical reliance on isolated case studies and ensures that the final taxonomy functions as a mathematically grounded, strategic capability roadmap. Consequently, organizations operating under highly regulated national frameworks, such as the UN Sustainable Development Goals and Transformation visions, can deploy these findings to systematically eliminate operational variance and drive sustained value creation.

III. RESULTS

The outcomes of the systematic literature review (SLR) and qualitative content analysis quantify the direct causal relationships between ISO standard execution and macro-sustainability performance. The synthesized data is partitioned into six structural categories and isolated during the review process. This section evaluates the degree of empirical support for each categorization, maps the cross-functional dependencies between standards and sustainability outcomes, and exposes the structural limitations reported across the literature.

A. Systematic Literature Review (SLR) Results

The final SLR corpus isolated 50 high-validity primary studies, published between 2000 and 2025, that establish empirical linkages between ISO frameworks and sustainability metrics. The longitudinal distribution of the literature indicates a statistically significant increase in publication volume post-2015, directly correlating with the global institutionalization of the United Nations Sustainable Development Goals (SDGs). The retained studies utilize diverse epistemological designs—including large quantitative surveys, structural equation modeling, bibliometric mapping, and deterministic case analyses—deployed

across highly heterogeneous industrial matrices. Table 4 isolates five foundational manuscripts that exert dominant theoretical influence over the synthesized corpus.

TABLE IV. FOUNDATIONAL LITERATURE DRIVING THE EMPIRICAL SYNTHESIS

No.	Authors / Journal	Target Domain	Methodological Design	Synthesized Outcome
1	Zhao et al. (2020), Sustainability	SDG Linkages: Food & Agriculture	Systematic Literature Review	ISO standards provide measurable support for SDG 2 targets; however, adoption variance among small-scale producers restricts macro-impact.
2	Bravi et al. (2020), Corp. Soc. Resp. & Env. Mgmt.	ISO 14001:2015 Implementation	Quantitative Survey (n = 284)	Environmental Management System (EMS) execution optimizes risk prevention and ecological metrics but introduces significant administrative complexity thresholds.
3	Ronalter et al. (2023), The TQM Journal	Management System Standards & SDGs	Bibliometric Analysis	Research saturation exists within core standards; systemic integration of peripheral standards is required to scale corporate sustainability outcomes.
4	Hahn (2013), Business Strategy and the Environment	ISO 26000 & Strategic Management	Critical Analytical Review	ISO 26000 functions as an effective baseline for Corporate Social Responsibility (CSR) initiation but lacks operational mechanisms for advanced strategic integration.
5	Nunhes et al. (2021), Frontiers in Sustainability	Annex SL Architecture & Sustainability	Systematic Literature Review	Utilizing the Annex SL High-Level Structure to integrate ISO 9001, 14001, and 45001 forces structural coherence across corporate sustainability vectors.

B. Core Management Systems and Quality Tools

Core Management System architecture specifically ISO 9001, ISO 14001, and ISO 45001, constitute the most heavily analyzed frameworks within the empirical corpus. The data dictates that deploying these standards yields statistically significant improvements in economic, environmental, and social performance matrices, provided they are executed comprehensively rather than symbolically (Ronalter et al., 2023; Bravi et al., 2020; Zimon et al., 2020). ISO 14001 correlates strongly with structural waste reduction, optimized resource utilization, and quantifiable ecological gains (Zeng et al., 2011). Concurrently, ISO 9001 reinforces economic resilience by minimizing operational variance and enforcing strict statistical process control (Macdonald, 2005; Zutshi & Sohal, 2004). However, the literature confirms that the efficacy of these architectures is highly dependent on internal control variables; organizational maturity, cross-departmental integration, and executive sponsorship dictate the ultimate yield of the system (Karapetrovic & Willborn, 1998). Furthermore, integrating multiple architectures via the Annex SL High-Level Structure effectively neutralizes operational redundancy, cementing these standards as interconnected operational variables rather than isolated compliance checklists.

C. People, Culture, and Social Responsibility

Frameworks targeting human capital and organizational ethics, primarily ISO 26000 (Social Responsibility) and ISO 45001 (Occupational Health and Safety), operationalize the sociological dimension of sustainable development. These standards mandate stakeholder integration, verifiable safety controls, and ethical corporate governance (Herrera-Sánchez et al., 2024; Hahn, 2013; Silva et al., 2022). ISO 26000 functions as the baseline architecture for establishing formal CSR programs; however, the literature notes that its non-certifiable structure limits its utility as a strategic differentiator for mature organizations (Hahn, 2013; Lozano & von Haartman, 2018). The empirical data confirms that fusing social responsibility protocols with safety management systems generates high employee engagement and cultural alignment. Yet, the industrial penetration of these social frameworks remains significantly lower than that of their environmental and quality counterparts (Ronalter et al., 2023). The synthesis dictates an operational necessity to scale people-centric standards to correct the heavy ecological bias present in traditional ISO implementations.

D. Governance, Compliance, and Innovation

The literature demonstrates a rapid functional evolution in governance and innovation frameworks. Standards such as ISO 31000 (Risk Management), ISO 37001 (Anti-Bribery Management), and ISO 37301 (Compliance Management), alongside ISO 56002 (Innovation Management) and ISO 50001 (Energy Management), establish the structural mechanisms connecting accountability with sustainable innovation (Herrera-Sánchez et al., 2024; Fonseca et al., 2022; Škrinjar et al., 2023). These architectures operationalize transparency protocols, allowing organizations to manage asymmetric risks associated with climate volatility and digital transformation. However, the data reveals severe adoption variance across industrial sectors. The integration of these advanced governance modules into existing Core Management Systems is frequently restricted by capital constraints and a lack of executive technical competence (Nawrocka & Parker, 2009; Zutshi & Sohal, 2004). Consequently, the literature mandates embedding these advanced architectures directly within primary operational frameworks to prevent the formation of isolated compliance silos.

E. Integration, Limitations, and Critiques

The empirical corpus contains severe critiques regarding the mechanical implementation of ISO standards. The data isolates a recurring failure mode: superficial procedural compliance without structural cultural alignment leads to "symbolic certification"—generating audit approval without executing tangible sustainability improvements (Boiral, 2011; Lozano & von Haartman, 2018). Furthermore, implementing disparate standards without a centralized integration mechanism produces severe administrative bloat and structural fragmentation (Rebelo et al., 2015). The literature also identifies a principle of diminishing marginal returns; highly mature organizations experience a minimal performance delta when adding tertiary certifications to an already optimized system (Bernardo et al., 2009). These empirical critiques confirm that ISO frameworks do not function as autonomous solutions. Their capacity to drive macro-sustainability is entirely contingent upon strategic integration, continuous innovation cycles, and mature organizational governance.

F. Importance-Performance Analysis (IPA) RESULTS

The Importance-Performance Analysis (IPA) generated deterministic coordinate mappings for the isolated ISO standards, quantifying the systemic variance between strategic utility Y-axis) and actualized institutional performance (X-axis). The spatial distribution across the mathematical matrix reveals acute implementation disparities among the six structural categories (Table 5). Core Management Systems—specifically ISO 14001, ISO 9001, and ISO 45001—exhibit the highest spatial coordinates in both variables, confirming absolute market saturation and mature execution methodologies. In stark contrast, frameworks governing social accountability, ethical compliance, and digital innovation register high theoretical importance coordinates but suffer from severe performance deficits. These findings suggest a systemic imbalance worth further empirical examination: organizations recognize the strategic necessity of advanced sustainability mechanisms but lack the operational capacity to deploy them effectively.

TABLE V. IMPORTANCE–PERFORMANCE MATRIX

ISO Standard	Category	Importance (Mean)	Performance (Mean)	Quadrant Classification	Interpretation
ISO 14001	Core Management Systems	4.8	4.6	I – Keep up the good work	Strong environmental results and maturity of implementation
ISO 9001	Core Management Systems	4.7	4.5	I – Keep up the good work	Proven driver of quality and competitiveness
ISO 45001	Core Management Systems	4.6	4.4	I – Keep up the good work	Effective in improving occupational safety culture
ISO 26000	Sustainability & Social Responsibility	4.7	3.5	II – Concentrate here	High potential for social impact; limited formal adoption
ISO 20400	Sustainability & Social Responsibility	4.6	3.6	II – Concentrate here	Key for sustainable procurement; needs stronger implementation
ISO 56002	Innovation & Digital Transformation	4.5	3.4	II – Concentrate here	Emerging relevance for innovation-driven sustainability
ISO 37301	Governance, Risk & Compliance	4.3	3.9	I – Keep up the good work	Strengthening compliance and ethical practices
ISO 37001	Governance, Risk & Compliance	4.2	3.7	II – Concentrate here	Vital for anti-corruption but uneven diffusion
ISO 10015	People, Competence & Culture	3.9	3.3	III – Low priority	Moderate impact; potential to reinforce human-capital sustainability
ISO/IEC 42001	Innovation & Digital Transformation	4.6	3.2	II – Concentrate here	Critical for AI governance; very early in adoption cycle
ISO 31000	Governance, Risk & Compliance	4.2	4.1	I – Keep up the good work	Well-institutionalized approach to risk and resilience
ISO 50001	Core Management Systems	4.5	3.8	II – Concentrate here	Significant contribution to energy transition, limited uptake

The empirical mapping translates into four strict operational directives defined by the spatial distribution of the frameworks within the coordinate matrix. Quadrant I designates standards classified as Institutionalized Excellence (ISO 9001, 14001, 45001, 31000, 37301), which function as mature structural anchors requiring maintained capital allocation to facilitate the integration of secondary frameworks. The literature corroborates this positioning, confirming that executing these core frameworks establishes the fundamental architecture necessary to reduce operational variance, unify compliance audits, and convert baseline sustainability metrics into structural competitive advantage (Bravi et al., 2020; Zimon & Madzík, 2020). Conversely, Quadrant II exposes Critical Implementation Deficits, isolating ISO 26000, 20400, 56002, 37001, 50001, and ISO/IEC 42001 as primary vectors for future sustainability scaling that possess massive strategic leverage yet fail to penetrate operational reality. Recent empirical assessments demonstrate this variance; despite offering quantifiable financial and ecological benefits, ISO 50001 implementation is severely restricted by technical complexity, achieving sub-15% penetration within advanced industrial sectors (Pagulayan & Calderon, 2026). Similarly, ISO 56002 remains a historically underutilized mechanism required to bypass early-stage innovation failures (BSI, 2026; Fonseca et al., 2022), while the ISO 26000 social responsibility framework suffers from extreme latency due to procedural ambiguity (Castka, 2018; Hahn, 2013), dictating immediate executive intervention and targeted competency development. In stark contrast, Quadrant III identifies frameworks such as ISO 10015 as Marginal Utility variables, mapping them to a low-priority sector where they exert insufficient mathematical impact on macro-sustainability targets to justify immediate capital deployment. The literature demonstrates a strict principle of diminishing returns in this sector, proving that isolating micro-standards without anchoring them to mature Quality Management Systems generates negligible systemic improvement (Aichouni et al., 2023; Bernardo et al., 2009). Finally, Quadrant IV, representing Resource Misallocation, returned null results, confirming that organizations are not over-investing in low-impact standards; because all high-performance frameworks strictly align with high strategic importance metrics, the data validates that organizational capital—while restricted in scope as seen in Quadrant II—is deployed efficiently wherever it is actively allocated (Nicoletti Junior & de Oliveira, 2019).

IV. DISCUSSION: THE PROPOSED ISO CAPABILITY MODEL

The empirical evidence synthesized through the systematic review confirms that ISO standards function as fundamental structural mechanisms for sustainable development, provided they are deployed within integrated management systems that unify quality, environmental, and social responsibility parameters (Bernardo et al., 2009; Bravi et al., 2020; Castka & Balzarova, 2008; Jørgensen et al., 2006; Zimon & Madzík, 2020; Zutshi & Sohal, 2004). The six-category architectural framework constructed in this study establishes a deterministic taxonomy, demonstrating how organizations systematically align specific standards to discrete sustainability objectives—ranging from variance reduction and ecological control to ethical

governance and innovation matrices (Fonseca et al., 2022; Hahn, 2013; Herrera-Sánchez et al., 2024; Karapetrovic & Willborn, 1998; Škrinjar et al., 2023; Zutshi & Sohal, 2004). However, the synthesis dictates that the functional yield of ISO standards remains severely context dependent. The mathematical impact of these frameworks fluctuates significantly based on external sector constraints, organizational scale, executive sponsorship, and the degree of cross-functional integration (Bravi et al., 2020; Castka & Balzarova, 2008; Jørgensen et al., 2006; Lozano & von Haartman, 2018; Ochieng & Awuor, 2020; Zutshi & Sohal, 2004).

Analyzing the distribution of empirical support across the taxonomy, environmental and quality management standards (ISO 14001, ISO 9001, and ISO 45001) maintain the highest verifiable correlations with optimized economic and ecological outcomes (Bernardo et al., 2009; Bravi et al., 2020; Castka & Balzarova, 2008; Jørgensen et al., 2006; Zeng et al., 2011; Zimon & Madzík, 2020; Zutshi & Sohal, 2004). The literature provides exhaustive statistical proof of measurable reductions in material waste, increased energy utilization ratios, and enhanced continuous process capability metrics. Conversely, standards governing social accountability and ethical compliance—specifically ISO 26000, ISO 37001, and ISO 37301—exhibit critical implementation deficits. Despite their established necessity for structural sustainability, these frameworks remain statistically marginalized (Hahn, 2013; Karapetrovic & Willborn, 1998; Lozano & von Haartman, 2018; Ochieng & Awuor, 2020; Rebelo et al., 2015; Silva et al., 2022). This severe adoption asymmetry indicates a systemic flaw: organizational capital is disproportionately allocated toward certifiable, audit-driven quality standards rather than frameworks mandating ethical or sociological performance. Reactivating this imbalance requires aggressive institutional incentivization and strict executive mandates.

The literature also isolates critical failure modes associated with standardized execution. Foremost, isolated procedural adoption frequently triggers "symbolic compliance"—satisfying audit requirements without generating substantive operational improvements (Bernardo et al., 2009; Boiral, 2011; Lozano & von Haartman, 2018; Ochieng & Awuor, 2020; Rebelo et al., 2015). Furthermore, deploying multiple, disconnected ISO frameworks induces severe administrative bloat and structural fragmentation (Boiral, 2011; Lozano & von Haartman, 2018; Ochieng & Awuor, 2020). The data proves a strict principle of diminishing marginal returns; resource-constrained organizations, or those operating with immature baseline systems, extract negligible value from secondary or tertiary certifications (Bernardo et al., 2009; Lozano & von Haartman, 2018; Ochieng & Awuor, 2020). These empirical constraints dictate that ISO frameworks do not function as autonomous instruments. They must be operationalized as integrated variables within a mature Quality Management System, structurally reliant on continuous innovation and robust corporate governance. Consequently, the macro-impact of these standards only materializes when organizations leverage them as mechanical drivers for continuous capability expansion rather than static regulatory checklists (Table 6).

TABLE VI. KEY RESEARCH OUTCOMES

Synthesized Claim	Empirical Validity	Functional Logic	References
ISO 14001 implementation optimizes environmental performance and accelerates sustainable development metrics.	High	Large-scale empirical sampling confirms measurable reductions in ecological waste, optimized resource allocation, and verifiable alignment with macro-environmental targets.	(Bernardo et al., 2009; Bravi et al., 2020; Castka & Balzarova, 2008; Jørgensen et al., 2006; Zeng et al., 2011; Zimon & Madzík, 2020; Zutshi & Sohal, 2004)
Integrating ISO 9001, 14001, and 45001 drives triple-bottom-line structural coherence.	High	Structural integration neutralizes redundancy and synchronizes economic, environmental, and social governance variables into a singular operational control mechanism.	(Bernardo et al., 2009; Castka & Balzarova, 2008; Jørgensen et al., 2006; Ronalter et al., 2023; Zimon & Madzík, 2020; Zutshi & Sohal, 2004)
ISO 26000 functions as a foundational CSR architecture but lacks advanced strategic mechanics.	Moderate	The framework successfully initiates baseline sociological compliance, but its non-certifiable structure restricts its utility as a competitive differentiator for highly mature organizations.	(Hahn, 2013; Lozano & von Haartman, 2018; Ochieng & Awuor, 2020; Rebelo et al., 2015)
Unintegrated procedural adoption generates symbolic compliance without substantive performance yield.	Moderate	Executing standards purely for audit validation, without achieving cultural or operational alignment, severely limits actualized sustainability improvements.	(Bernardo et al., 2009; Boiral, 2011; Lozano & von Haartman, 2018; Ochieng & Awuor, 2020; Rebelo et al., 2015)
Innovation (ISO 56002) and Energy (ISO 50001) standards exhibit high strategic utility but suffer from systemic implementation deficits.	Moderate	These frameworks demonstrate high theoretical capacity to drive digital and energy transitions, yet empirical penetration remains constrained by capital costs and technical complexity.	(Fonseca et al., 2022; Nawrocka & Parker, 2009; Škrinjar et al., 2023; Zutshi & Sohal, 2004)
ISO standard efficacy is strictly moderated by external sector parameters and internal organizational scale.	Moderate	Quantitative assessments prove that the functional yield of standard deployment fluctuates based on localized constraints, structural maturity, and regional regulatory pressures.	(Bravi et al., 2020; Jørgensen et al., 2006; Karapetrovic & Willborn, 1998; Zeng et al., 2011)

A. The ISO Standards for Sustainable Development

The culmination of this empirical synthesis is a definitive architectural taxonomy comprising forty-two international standards. These frameworks represent the verified mechanisms required to institutionalize organizational sustainability across economic, environmental, and sociological dimensions. The extracted standards have been rigorously validated and segmented into six structural vectors: Core Management Systems; Quality Support, Methods, and Tools; People, Competence, and Culture; Sustainability and Social Responsibility; Governance, Risk, and Compliance; and Innovation, Technology, and Digital Transformation. This structural organization maps directly to the United Nations Sustainable Development Goals (SDGs) and the macroeconomic imperatives of the Transformation visions framework. By defining standards not as isolated regulations but as integrated mechanisms, this taxonomy provides strategic planners with an exact

operational roadmap to execute the triple-bottom-line methodology—People, Prosperity, and Planet. The literature confirms that deploying this holistic architecture converts theoretical sustainability targets into organizational capabilities (Zairi, 2014; Mosgaard & Kristensen, 2023; Ronalter et al., 2023).

Figure (1) shows the proposed integrated ISO standards for organizational sustainability; It illustrates how 42 ISO standards, organized into six capability domains and integrated around the triple sustainability core (economic, environmental, and social). It is suggested that through this model the six capability domains interact as a chain of reactions to leverage ISO standards from compliance to performance excellence and achieving sustainability for humanity (People, Prosperity and Planet).

TABLE VII. THE ARCHITECTURAL TAXONOMY OF ISO STANDARDS (PROPOSED MODEL)

Structural Category	Standard Designation	Functional Nomenclature	Operational Relevance within the System
Core Management Systems	ISO 9001:2015	Quality Management Systems	Foundational architecture for process control and variance reduction.
	ISO 14001:2015	Environmental Management Systems	Primary control mechanism for ecological impact mitigation.
	ISO 45001:2018	Occupational Health & Safety Systems	Enforces verifiable safety controls and risk neutralization.
	ISO 50001:2018	Energy Management Systems	Dictates strict energy efficiency and carbon output optimization.
	ISO 30401:2018	Knowledge management systems	Secures reliability across highly regulated biological supply chains.
	ISO/IEC 27001:2022	Information Security Management Systems	Protects structural data integrity against digital volatility.
	ISO 21001:2018	Educational Organizations Mgmt. Systems	Institutionalizes quality control within learning ecosystems.
Quality Support, Methods & Tools	ISO 10002–10004:2018	Quality management - Customer Satisfaction Guidelines	Calibrates external stakeholder feedback to internal process metrics.
	ISO 10005:2018	Quality management - Quality Plans Guidlines	Establishes execution reliability for isolated operational initiatives.
	ISO 10006:2017	Quality management – Quality in Projects	Establish guidelines to ensure quality in projects
	ISO 10017:2021	Quality management — Guidance on statistical techniques for ISO 9001	Forces strict mathematical rigor into qualitative decision-making.
	ISO 10009:2024	Quality management — Guidance for quality tools and their application	Guidelines for selection of Quality tools and techniques for process improvement projects
	ISO 19011:2018	Auditing Management Systems	Governs the internal continuous improvement and verification cycle.
People, Competence & Culture	ISO 10010:2022	Quality Culture Guidance	Engineers behavioral alignment with quality objectives.
	ISO 10015:2019	Competence Management	Standardizes human capital acquisition and capability retention.
	ISO 10018:2020	People Engagement	Structurally mandates participative operational excellence.
	ISO 20700:2017	Management Consultancy Services	Governs external advisory quality and intervention reliability.
	ISO 18404:2015	Six Sigma & Lean Competence	Certifies high-level statistical process capability within personnel.
	ISO 29993:2017	Learning Services Outside Formal Ed.	Regulates non-traditional capability development vectors.
	ISO 30414:2018	Human resource management — Guidelines	Guidelines for internal and external human capital reporting
Sustainability & Social Responsibility	ISO 14004 / 14031	Environmental Implementation & Eval.	Installs deterministic performance measurement for ISO 14001.
	ISO 26000:2010	Social Responsibility	Defines the baseline architecture for ethical corporate action.
	ISO 20400:2017	Sustainable Procurement	Forces supply chain networks to adopt environmental controls.
	ISO 37101 / 37120	Sustainable Development in Communities	Expands standard application from enterprise to municipal levels.
	ISO 59004:2024	Circular economy — Framework, principles and guidance for implementation	Guidance for implementation of circular economy concepts
Governance, Risk & Compliance	ISO 37001 / 37301	Anti-Bribery & Compliance Systems	Constructs rigid legal barriers and ethical transparency mandates.
	ISO 22301:2019	Business Continuity Management	Engineers systemic resilience against severe external shocks.
	ISO 31000:2018	Risk Management Guidelines	Provides the universal mathematical framework for managing uncertainty.
	ISO 37000:2021	Governance of organizations — Guidance	Guidance on implementation of governance in organizations
	ISO 37004 / 44000	Governance Maturity & Collaboration	Secures inter-organizational trust through auditable parameters.
Innovation & Digital Transformation	ISO/IEC 42001:2023	AI Management Systems	Implements critical algorithmic governance and ethical AI controls.
	ISO/IEC 38507:2022	Governance implications of the use of artificial intelligence by organizations	Governance implications of the use of artificial intelligence by organizations
	ISO/IEC 38500 / 38507	Governance of IT and AI	Mandates board-level accountability for technological deployment.

	ISO 56000 / 56002 / 56005	Innovation & IP Management	Systematizes R&D capability to prevent innovation failure modes.

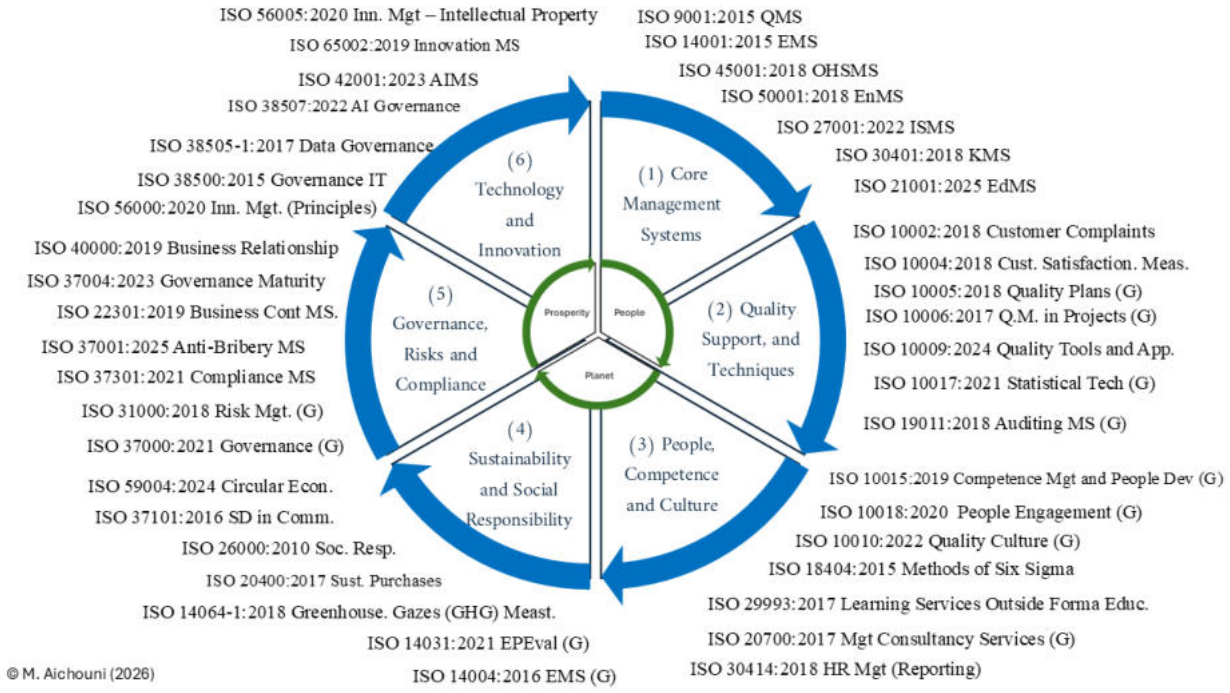


FIGURE I. INTEGRATED ISO STANDARDS CAPABILITY-TO-IMPACT MODEL FOR ORGANIZATIONAL SUSTAINABILITY

B. Structural and Managerial Implications

From a theoretical perspective, this research advances the conceptualization of ISO standards beyond passive regulatory checklists, identifying them instead as adaptive operational capabilities. The six-category architectural framework proves that cross-standard integration eliminates structural redundancy and accelerates macro-sustainability objectives. By linking this taxonomy to the UN SDGs and Transformation visions, the study provides robust literature demonstrating that standard-based governance functions as a highly scalable model for national economic transformation (Aichouni et al., 2020; Ronalter et al., 2023; Tasleem et al., 2019).

At the managerial level, the Importance-Performance Analysis (IPA) matrix delivers a deterministic roadmap for capital allocation. Executive leadership must abandon generic standard adoption and instead utilize the data to target Critical Implementation Deficits (Quadrant II). Prioritizing investment in underutilized, high-leverage frameworks—specifically ISO 26000 (Social Responsibility), ISO 37001 (Anti-Bribery), and ISO/IEC 42001 (Artificial Intelligence)—will generate severe competitive advantages. To avoid administrative failure, these secondary architectures must be hardwired directly into the mature Core Management Systems (ISO 9001, 14001) mapped in Quadrant I, ensuring balanced, triple-bottom-line performance (Zairi, 2014; Latan et al., 2020; Zimon & Madzík, 2020).

From a macroeconomic policy perspective, the empirical data exposes severe systemic vulnerabilities within the National Quality Infrastructure (NQI). Implementation latency is heavily concentrated within Small and Medium-sized Enterprises (SMEs) due to prohibitive capital costs and technical complexity barriers (Al-Tit, Omri, & Euchi, 2019; Lee & Kim, 2022). To correct this variance, policymakers must engineer targeted interventions. Establishing aggressive certification subsidization, deploying specialized competency training, and legally embedding advanced ISO standards into national procurement requirements will enforce mass compliance. Executing these policy adjustments will permanently align industrial operations with SDG parameters, securing Saudi Arabia's position as a regional authority in standard-driven economic excellence.

V. CONCLUSIONS

This research employed a dual-method Systematic Literature Review (SLR) and Importance–Performance Analysis (IPA) to examine the operational contribution of ISO standards to macro-sustainability objectives. The empirical synthesis resulted in a validated, six-domain architectural taxonomy comprising forty-two international ISO standards, demonstrating that these frameworks function not as isolated regulatory checklists but as integrated organizational capabilities for operationalizing economic, environmental, and social governance. The IPA mapping confirms that core environmental and quality management systems—most notably ISO 14001, ISO 9001, and ISO 45001—act as institutionalized anchors, delivering consistent process stability, risk reduction, and environmental performance improvements. At the same time, the analysis reveals critical implementation gaps in high-leverage domains related to social responsibility, innovation, and digital transformation, particularly ISO 26000, ISO 56002, and ISO/IEC 42001. Addressing these adoption asymmetries is essential for aligning organizational practices with the United Nations Sustainable Development Goals and national transformation visions. Overall, the findings reinforce the argument that long-term sustainability cannot be achieved through compliance-oriented

certification alone but requires a deliberate transition toward standard-driven capability integration and continuous organizational learning.

Building on these findings, several directions for future research are necessary and promising. First, empirical validation of the proposed model through large-scale surveys, structural equation modeling, or longitudinal case studies would allow quantitative testing of the causal relationships between integrated ISO capability domains and sustainability performance outcomes. Second, future studies could explore contextual differentiation, examining how organizational size, sector, maturity level, and national regulatory environments moderate the effectiveness of specific ISO standards within the model, with particular attention to small and medium-sized enterprises. Third, the dynamic interaction between emerging governance and digital standards, such as ISO/IEC 42001 for artificial intelligence management, and traditional management system standards warrants deeper investigation, especially in light of accelerating digital transformation and ethical risks. Finally, future research could operationalize the model as a staged or maturity-based implementation roadmap, enabling organizations to prioritize ISO adoption pathways aligned with resource availability and strategic objectives. Such research extensions would strengthen the empirical grounding of the proposed framework and support its evolution from a conceptual sustainability model into a robust decisions support tool for organizations and policymakers.

References

Aichouni, A. B. E., Kolsi, L. and Aichouni, M. (2020). 'The Engineering Students Innovation Club Project for Human Capital Development in the areas of Industry 4.0 – From the Design to Implementation', *2020 IEEE Industrial & Systems Engineering Conference (ISEC)*, pp. 1-4.

Aichouni, M., Touahmia, M., Alshammari, S., Said, M. A., Aichouni, A. B. E., Almudayries, M. and Aljohani, H. (2023). 'An Empirical Study of the Contribution of Total Quality Management to Occupational Safety and Health Performance in Saudi Organizations', *International Journal of Environmental Research and Public Health*, 20(2), p. 1495.

Al-Tit, A., Omri, A. and Euchi, J. (2019). 'Critical success factors of small and medium-sized enterprises in Saudi Arabia: Insights from sustainability perspective', *Administrative Sciences*, 9(2), p. 32.

Arana-Landín, G., Landeta-Manzano, B., Laskurain-Iturbe, I. and Priyadarshini, A. (2024). 'How Is ISO 14001:2015 Boosting the Spread of Product-Oriented Environmental Management Practices?', *Corporate Social Responsibility and Environmental Management*, 31(1), pp. 1-15.

Balzarova, M. A. and Castka, P. (2018). 'Social responsibility: Experts' viewpoints on adoption of the ISO 26000 standard', *Corporate Social Responsibility and Environmental Management*, 25(5), pp. 819-824.

Barni, A., Fontana, A., Menato, S., Sorlini, M. and Canetta, L. (2018). 'Exploiting the Digital Twin in the Assessment and Optimization of Sustainability Performances', *2018 IEEE 9th International Conference on Intelligent Systems (IS)*, pp. 706-713.

Bernardo, M., Casadesús, M., Karapetrovic, S. and Heras, I. (2009). 'How Integrated Are Environmental, Quality and Other Standardized Management Systems? An Empirical Study', *Journal of Cleaner Production*, 17(8), pp. 742-750.

Boiral, O. (2011). 'Managing with ISO Systems: Lessons from Practice', *Long Range Planning*, 44(3), pp. 197-220.

Bravi, L., Santos, G., Pagano, A. and Murmura, F. (2020). 'Environmental Management System According to ISO 14001:2015 as a Driver to Sustainable Development', *Corporate Social Responsibility and Environmental Management*, 27(6), pp. 2599-2614.

Castka, P. and Balzarova, M. A. (2008). 'ISO 26000 and supply chains—On the diffusion of the social responsibility standard', *International Journal of Production Economics*, 111(2), pp. 274-286.

Croft, N. H. (2014). 'ISO 9001 and sustainability', *ISO Focus*, 104(1), pp. 14-17.

Denicolai, S., Zucchella, A. and Magnani, G. (2021). 'Internationalization, Digitalization, and Sustainability: Are SMEs Ready? A Survey on Synergies and Substituting Effects Among Growth Paths', *Technological Forecasting and Social Change*, 166, p. 120650.

Denyer, D. and Tranfield, D. (2009). 'Producing a systematic review', in Buchanan, D. A. and Bryman, A. (eds.) *The Sage Handbook of Organizational Research Methods*. London: Sage Publications, pp. 671-689.

Fonseca, L. M., Domingues, J. P., Pereira, M. T., Martins, F. F. and Zimon, D. (2022). 'Assessment of Innovation Management Systems (ISO 56002) in Portuguese Organizations', *Sustainability*, 14(15), p. 9508.

Hahn, R. (2013). 'ISO 26000 and the Standardization of Strategic Management Processes for Sustainability and Corporate Social Responsibility', *Business Strategy and the Environment*, 22(7), pp. 442-455.

Herrera-Sánchez, G., Morán-Bravo, L. del C., Silva-Juárez, A. and Sampedro-Poblano, H. M. (2024). 'Sustainable Development Goals and ISO Standards in Industrial Maintenance', *Revista de Ingeniería Industrial*, 8(1), pp. 45-60.

Jørgensen, T. H., Remmen, A. and Mellado, M. D. (2006). 'Integrated Management Systems – Three Different Levels of Integration', *Journal of Cleaner Production*, 14(8), pp. 713-722.

Karapetrovic, S. and Willborn, W. (1998). 'Integration of quality and environmental management systems', *The TQM Magazine*, 10(3), pp. 204-213.

Kush, P., Cudney, E. A., Materla, T. and Antony, J. (2021). 'Systematic Literature Review of Quality Maturity Matrix', *Total Quality Management & Business Excellence*, 32(3-4), pp. 289-297.

Latan, H., Chiappetta Jabbour, C. J., Lopes de Sousa Jabbour, A. B., de Camargo Fiorini, P. and Foropon, C. (2020). 'Innovative Efforts of ISO 9001-Certified Manufacturing Firms: Evidence of Links Between Determinants of Innovation, Continuous Innovation and Firm Performance', *International Journal of Production Economics*, 223, p. 107526.

Lee, K. and Kim, D. (2022). 'The Integration of ISO 14001 and ISO 9001 for Sustainable Performance in SMEs', *Sustainability*, 14(10), p. 5896.

Louw, L. and Droomer, M. (2019). 'Development of a Low-Cost Machine Vision-Based Quality Control System for a Learning Factory', *Procedia Manufacturing*, 31, pp. 264-269.

Lozano, R. and von Haartman, R. (2018). 'Exploring the Integration of ISO 9001, ISO 14001, and ISO 45001 in Swedish Organizations', *Sustainability*, 10(6), p. 1959.

Macdonald, J. (2005). 'Strategic Sustainable Development Using the ISO 14001 Standard', *Journal of Cleaner Production*, 13(6), pp. 631-643.

Merli, R., Preziosi, M. and Acampora, A. (2018). 'How Do Scholars Approach the Circular Economy? A Systematic Literature Review', *Journal of Cleaner Production*, 178, pp. 703-722.

Missimer, M., Robèrt, K. and Broman, G. (2014). 'A Systems Perspective on ISO 26000', *Journal of Cleaner Production*, 75, pp. 16-24.

Moktadir, M. A., Dwivedi, A., Rahman, A., Chiappetta Jabbour, C. J., Paul, S. K., Sultana, R. and Madaan, J. (2020). 'An Investigation of Key Performance Indicators for Operational Excellence Towards Sustainability in the Leather Products Industry', *Business Strategy and the Environment*, 29(8), pp. 3331-3351.

Mosgaard, M. A. and Kristensen, H. (2023). 'From Certified Environmental Management to Certified SDG Management: New Sustainability Perceptions and Practices', *Sustainable Futures*, 5, p. 100108.

Nawrocka, D. and Parker, T. (2009). 'Finding the Connection: Environmental Management Systems and Environmental Performance', *Journal of Cleaner Production*, 17(6), pp. 601-607.

Nicoletti Junior, A. and de Oliveira, M. C. (2019). 'The Impact of the Implementation of the Quality Management System on Organisational Performance: An Action Research in a Brazilian Brewing Manufacture', *Total Quality Management & Business Excellence*, 30(11-12), pp. 1411-1427.

Nunhes, T. V., Bernardo, M. and Oliveira, O. J. (2021). 'Contributions of Annex SL to Corporate Sustainability', *Frontiers in Sustainability*, 2, p. 660851.

Ochieng, J. and Awuor, F. (2020). 'The Contribution of ISO 9001 to Sustainable Development and Organizational Performance', *The TQM Journal*, 32(4), pp. 695-711.

Rafique, M. Z., Ab Rahman, M. N., Saibani, N. and Arsad, N. (2019). 'A Systematic Review of Lean Implementation Approaches: A Proposed Technology Combined Lean Implementation Framework', *Total Quality Management & Business Excellence*, 30(3-4), pp. 386-421.

Rebelo, M. F., Santos, G. and Silva, R. (2015). 'A Generic Model for Integration of Quality, Environment and Safety Management Systems', *The TQM Journal*, 27(6), pp. 685-699.

Rezak, O., Hamadache, K., and Talahite, F. (2022). 'The impact of ISO 9001 certification on firm performance', *Valahian Journal of Economic Studies*, 13(2), pp. 105-114.

Ronalter, L. M., Bernardo, M. and Romaní, J. M. (2023). 'Management System Standards & the Sustainable Development Goals: A Bibliometric Analysis', *The TQM Journal*, 35(8), pp. 83-110.

Schwartz, B. and Tilling, K. (2009). 'ISO-lating Corporate Social Responsibility in the Organizational Context: A Dissenting Interpretation of ISO 26000', *Organization & Environment*, 22(3), pp. 289-299.

Senivongse, C., Bennet, A. and Mariano, S. (2017). 'Utilizing a systematic literature review to develop an integrated framework for knowledge management', *VINE Journal of Information and Knowledge Management Systems*, 47(2), pp. 250-264.

Shin, N., Park, S. H. and Lee, J. Y. (2018). 'The Effect of ISO 9001 Certification on Organizational Performance', *International Journal of Quality & Reliability Management*, 35(10), pp. 2029-2045.

Silva, R., Rebelo, M. and Santos, G. (2022). 'Management System Standards and Social Responsibility Integration in Portuguese Organizations', *Sustainability*, 14(9), p. 5262.

Škrinjar, R., Trivun, D. and Vukšić, V. (2023). 'Quality Management Standards as Enablers of Sustainable Organizational Learning', *Journal of Cleaner Production*, 390, p. 136154.

Tasleem, M., Khan, N. and Nisar, A. (2019). 'Impact of Technology Management on Corporate Sustainability Performance: The Mediating Role of TQM', *International Journal of Quality & Reliability Management*, 36(9), pp. 1574-1599.

Womack, J. P. and Jones, D. T. (1996). *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*. New York: Simon & Schuster.

Zairi, M. (2014). 'Building Organizational Capability', ECBPM, UK.

Zhao, X., Castka, P. and Searcy, C. (2020). 'ISO Standards: A Platform for Achieving Sustainable Development Goal 2', *Sustainability*, 12(22), p. 9504.

Zeng, S. X., Tam, C. M. and Tam, V. W. Y. (2011). 'Integrating Safety, Environmental and Quality Management Systems for Total Performance Improvement', *Journal of Cleaner Production*, 19(15), pp. 2106-2115.

Zimon, D. and Madzík, P. (2020). 'Standardized management systems and risk management in the supply chain', *International Journal of Quality & Reliability Management*, 37(2), pp. 305-327.

Zutshi, A. and Sohal, A. (2004). 'Adoption and Maintenance of Environmental Management Systems: Critical Success Factors', *Management of Environmental Quality: An International Journal*, 15(4), pp. 399-419.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Development of an operational excellence model for research centres organization and management

Aline Barros [1], Rui M. Sousa [2] and José Dinis-Carvalho [2]

[1] Centre of Biological Engineering, University of Minho, Braga, Portugal; alinebarros@ceb.uminho.pt
[2] Department of Production and Systems, ALGORITMI Centre, University of Minho, Guimarães, Portugal

***Abstract*— Operational Excellence (OpEx) models are well established in industrial and service settings, but their application to public Research Centres (RC) remains limited. These organisations operate in knowledge-intensive environments while facing rigid administrative systems, diffuse responsibilities, and strong public accountability, which make direct transfers of existing OpEx models inadequate. This paper introduces an OpEx model specifically developed to the organisation and management of RC. The model was developed using an Action Research approach, combining insights from literature, semi-structured interviews, and a questionnaire applied in a Portuguese RC. It is structured around seven guiding principles, two structural pillars, and a dual-cycle improvement logic linking tactical problem-solving with strategic development, all oriented towards a broader societal purpose. The results show that some principles can be readily adapted to the RC context, whereas others require reinterpretation. The model contributes by extending OpEx to a context that is still little explored and also by emphasizing the implementation of principles through concrete practical guidelines, instead of focusing only on the conceptual adaptation/adoption of these same principles.**

Keywords – Operational Excellence, Research Centres, Continuous Improvement, Organizational Performance

I. Introduction

In the pursuit of continuous improvement, organizations seek excellence by applying principles rooted in Operational Excellence (OpEx) models (Carvalho, 2021; Sony, 2019). Over recent years, a growing body of research has explored the implementation of such models across diverse organizational contexts (Liker and Morgan, 2006; Narayanamurthy, *et al.*,2017; Carvalho, 2021; Jengwa and Pellissier, 2022), including the public sector - police departments, hospitals, municipal agencies, and educational institutions (Caiado et al., 2020; Rahmati and Jalilvand, 2024; Shamsuzzaman et al., 2023). Systematic literature reviews confirm the relevance of these initiatives in such environments (Caiado et al., 2020; Lukrafka et al., 2020; Rodgers and Antony, 2019). Although OpEx models provide a structured means to enhance efficiency and performance (Caiado et al., 2020; Rodgers and Antony, 2019), their direct transferability to public contexts poses challenges. Originally developed for profit-oriented environments, these models often require adaptation to contexts characterized by complex bureaucracies and non-financial goals (Hildebrandt et al., 2022; Rahmati and Jalilvand, 2024; Rodgers and Antony, 2019). As Radnor and Walley (2008) highlight, public organizations may either adopt isolated Lean practices or pursue comprehensive transformations. However, many public sector improvement efforts remain superficial, often focusing on short-term cost reduction and process speed, without fostering a lasting culture of learning. According to Radnor's "House of Lean for Public Services" (Caiado *et al.* (2020), the goal is to achieve "[...] a whole process, value chain or system view, embedded improvement behaviours and stable robust processes".

Public Research Centres (RC) face a paradoxical condition. They operate in knowledge-intensive environments while facing rigid administrative and financial systems that limit autonomy. Their core activity depends on creativity and long-term collaboration, while their sustainability relies on compliance and continuous funding acquisition (Coccia, 2019a; Ryazanova and Jaskiene, 2022). Daily management is often tactical rather than strategic driven by the pressure to secure resources, report metrics, and satisfy overlapping evaluation systems. RC work within host institutions whose infrastructures and rules they do not fully control, creating tensions between academic freedom and operational constraints. These characteristics distinguish RC from other organisations where OpEx models are typically implemented. In the national context, these centres play a pivotal role in research and innovation systems. Often affiliated with universities, they conduct interdisciplinary Research and Development (R&D), foster knowledge transfer, and train highly qualified personnel. In Portugal, public RC account for a significant share of scientific output and contribute to strategic priorities such as digital transition, sustainability, and health ("FCT - Fundação para a Ciência e a Tecnologia," 2025). Yet these centres face operational pressures stemming from competitive funding and the need to demonstrate societal value ("CoARA," 2023; Coccia, 2019b, 2019a). These pressures are intensified by rigid structures and growing expectations for transparency and efficiency (Antony et al., 2017; "CoARA," 2023; Pekkanen and Pirttilä, 2022). Among the most recurrent operational challenges are inefficiencies in procurement (e.g., purchasing of equipment and materials), maintenance (e.g., equipment and facilities), and research management that directly affect institutional performance. Despite this, RC remain underexplored as complex operational systems.

To map the current state of knowledge on this matter, a systematic literature review (SLR) was conducted previously by the authors, guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses

(PRISMA) methodology (Mengist et al., 2020; Moher et al., 2010). No studies explicitly addressing OpEx models in RC were identified.

The prevailing perception that scientific rigor alone ensures excellence tends to obscure the importance of structured operational practices. Yet improving internal processes and eliminating inefficiencies are vital to guarantee performance, sustainability, and resilience. Developing and validating tailored OpEx models may thus strengthen institutional capacity and societal impact.

This paper presents the development of an OpEx model adapted to the specific characteristics of RC. Grounded in an Action Research (AR) approach, it integrates insights from literature, semi-structured interviews, and a questionnaire implemented at the Centre of Biological Engineering (CEB), from University of Minho, a Portuguese RC known for its scientific excellence. The resulting model is empirically grounded and sensitive to the context of publicly funded research institutions, contributing to both scholarly literature and practical management improvement.

II. Methodology

This study follows AR as the guiding methodology for developing an OpEx model adapted to RC. AR is particularly suited to contexts in which the goal is both to generate academic knowledge and to drive organizational change through direct involvement in practice (Coughlan and Coghlan, 2002). The overall process is summarized in Fig.1, which illustrates how the literature and the researcher embeddedness in the RC informed the empirical stages and how each phase contributed to the iterative development of the model.

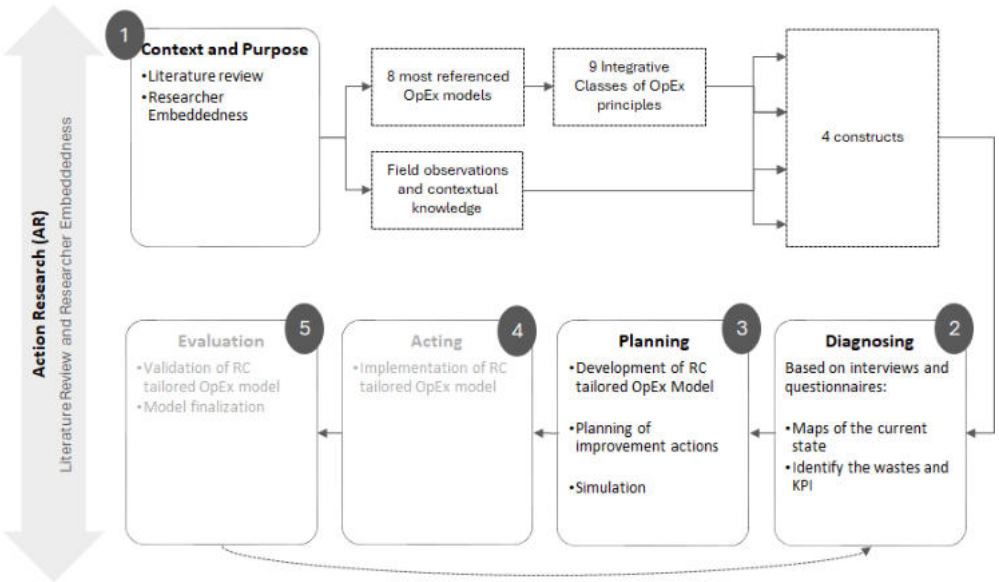


Fig. 1. Methodology.

III. The research was conducted in close collaboration with a Portuguese RC. Following an initial phase focused on understanding the context and purpose, the study progressed through four stages: diagnosis, planning, action, and evaluation. The first phase included a SLR that mapped existing OpEx models and principles, highlighting gaps in their application to public research organisations. Combined with researcher immersion, this stage enabled the identification of relevant constructs and contextual factors. Building on this, phase 2 involved a multi-method diagnostic, including semi-structured interviews with seven current and former directors of high-performing RC. Interview data were audio-recorded, transcribed, and analysed using a deductive thematic analysis approach grounded in the study constructs. Complementing these qualitative insights, a structured questionnaire was administered to collaborators across scientific, technical, and administrative roles at the RC under study, yielding 109 valid responses from a population of 260 participants. Quantitative data were analysed using descriptive statistical methods, and questionnaire reliability was also assessed prior to application. Both instruments were subjected to pre-testing using a think-aloud approach. Together, these methods connected strategic perspectives with daily realities. The instruments targeted four key constructs: OpEx Principles, Critical Success Factors, Problematic Processes, and Employee Engagement. Phase 3, which is the focus of this paper, concerns model development based on insights from the empirical data and existing model structures. The model was developed using an abductive approach, where theory and data were iteratively aligned (Åhlström et al., 2021). The study received approval from the Ethics Committee for Research in Social and Human Sciences of the University of Minho, and informed consent was obtained from all participants prior to data collection. Notably, the objective of the study was not reaching a generic outcome, but the development of a model grounded in the operational realities of research environments and capable of informing future applications and adaptations in similar contexts. The findings should therefore be interpreted within their specific organisational and institutional context. The remaining phases will address implementation and evaluation. Background

A. Conceptual Foundations

The conceptual foundations of this study were established during phase 1 of the AR cycle (Fig. 1), which combined a broad review of OpEx literature with researcher embeddedness in the daily life of the RC. This stage aimed to

consolidate existing knowledge and identify constructs and operational dynamics relevant to the RC context. Through this process, an integrative analysis of eight widely referenced OpEx models, namely, Lean Thinking, Toyota Way, Shingo, Deming, MBNQA, EFQM, Agile, and DevOps, led to the consolidation of nine classes of OpEx principles: Respect for Each Individual, Teamwork, Continuous Improvement, Transformational Leadership, Process Knowledge, Higher Purpose, Pull Flow, Quality at the Source, and Customer-Oriented Value(Barros et al., 2025a; Barros et al., 2025b).In parallel, four constructs capturing key dimensions of operational performance in RC were identified: OpEx Principles, Critical Success Factors, Problematic Processes, and Employee Engagement. These constructs emerged through an iterative process combining literature insights with observations from the researcher's embedded experience. Together, they provided the conceptual basis for the model and informed the design of the empirical instruments.

B. Empirical Evidence

Building on the conceptual foundation established in the first stage of the AR cycle, a multi-method diagnostic phase (phase 2, Fig. 1) was conducted to examine how the identified constructs manifest in the daily functioning of a public RC Semi-structured interviews were carried out with current and former directors of excellent RC, capturing their perspectives on organisational challenges, leadership practices, and operational priorities. Complementing these qualitative insights, a structured questionnaire was administered to collaborators across scientific, technical, and administrative roles at the RC under study. Together, these methods capture both strategic and operational perspectives across the organisation. Engagement, as one of the key constructs, was measured using items from the Gallup Q12® framework. Findings from both data sources highlight the central importance of Respect for Each Individual, Continuous Improvement, Process Knowledge, and Teamwork in the pursuit of excellence in RC. These principles were widely recognized across different organizational roles. Process Knowledge, emerged in the interviews as a key enabler of operational efficiency, particularly in contexts involving complex, interdisciplinary collaboration. Other principles showed more uneven recognition; Higher Purpose, for instance, was less prominent in the questionnaire but emerged strongly in interviews with directors, particularly regarding societal relevance. In contrast, Pull Flow and Quality at the Source were poorly understood and less frequently applied, pointing to a need for contextual adaptation. This reflects broader challenges in translating industrial concepts into research environments. The remaining constructs further contextualized these insights. Participants highlighted the importance of complementing bibliometric indicators with alternative metrics such as technology transfer, patents, partnerships, and altmetrics. Administrative support processes were frequently cited as a source of inefficiency, particularly due to bureaucracy and fragmented workflows, often linked to broader institutional constraints. Finally, the Employee Engagement construct revealed strong alignment with the centre's mission and high levels of commitment across teams. However, gaps in recognition and career development were identified. Overall, these findings inform the development of the OpEx model by identifying key areas requiring reinforcement and adaptation in RC.

IV. MODEL DEVELOPMENT

The conceptual foundations and empirical diagnosis presented clarified which elements of OpEx are already embedded in RC, which require adaptation, and which remain underdeveloped. Principles such as Respect for People, Teamwork, Continuous Improvement, and Process Knowledge, emerged as strongly aligned with daily practice, whereas others, including Higher Purpose, Pull Flow, and Quality at the Source, showed inconsistent understanding or limited applicability. Structural constraints, fragmented workflows, and leadership gaps further underscored the need for a model capable of translating OpEx principles into operational routines suited to research environments. Grounded in both literature and empirical fieldwork, the proposed model reflects the unique demands of scientific environments, namely, complex processes, interdisciplinary collaboration, and public accountability (Coccia, 2019a). The model integrates four key dimensions: OpEx principles, critical success factors, problematic processes, and employee engagement, which collectively served as the foundation for its structure. The model is composed of: (i) a visual framework, centred on two interconnected improvement cycles and supported by foundational principles classes; (ii) guiding principles, adapted to the RC context; and (iii) a set of implementation guidelines designed to translate these principles into concrete organisational practices.

A. Visual Framework of the Model

The proposed model is illustrated in Fig. 2.

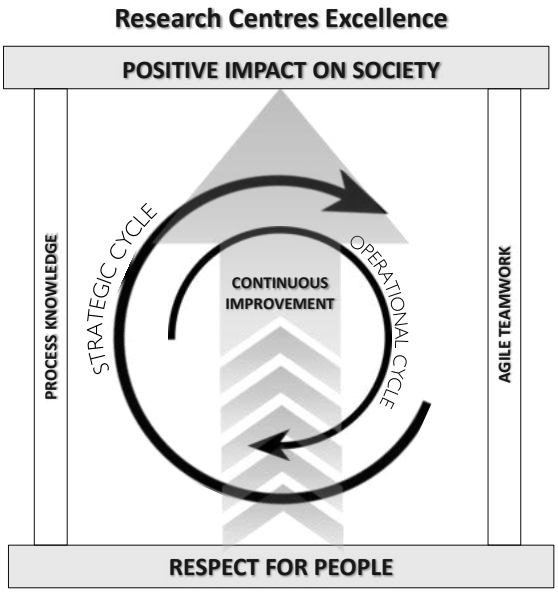


Fig. 2. Operational Excellence Model for Research Centres Organization and Management.

The model is structured around five main elements: a foundation, a central axis, two interconnected improvement cycles, two supporting pillars, and a guiding purpose at the top. At the base lies Respect for People, which emerged as a key driver of trust, engagement, and innovation across roles. In the RC context, this foundation also encompasses the way leadership is exercised, less through hierarchical authority and more through practices that enable participation, initiative, and collective alignment. This foundation supports the model's central axis, Continuous Improvement, which links individual contributions to organisational goals and fuels iterative progress.

The model incorporates a dual-cycle structure that mirrors the two interdependent dynamics of RC. The inner cycle, focused on problematic processes, operates at the tactical level, where continuous improvement drives change by addressing day-to-day obstacles, surfacing inefficiencies, resolving failures, and removing operational constraints such as bureaucracy, equipment failures, procurement delays, and workflow interruptions. In this cycle, the focus is on how to improve, identifying where breakdowns occur and correcting the errors that prevent success factors from being achieved. Surrounding it, the outer strategic cycle strengthens the success factors that support long-term scientific and organisational performance, such as leadership, collaboration, communication, and purpose alignment. Here, continuous improvement supports the development of practices that enable research performance, including the definition and monitoring of tailored indicators aligned with impact, quality, speed, and efficiency. These improvements help create the conditions for more effective research, as researchers perform better when not constrained by operational disruptions (Ryazanova and Jaskiene, 2022; Molinéro-Demilly *et al.*, 2025). This configuration is deliberate: strategic development defines what needs to be in place to enable excellence, while tactical problem-solving ensures that daily operations do not undermine those conditions.

At the centre of both cycles lies Continuous Improvement, which serves as the connecting axis between short-term correction and long-term capability building. Positioned vertically, it anchors the model by translating operational learning into strategic insight and translating strategic intentions into daily practice. Although key performance and behaviour indicators (KPI and KBI) are not represented in the visual framework, they play a supporting role in the functioning of the model. As part of the Continuous Improvement axis, measurement routines provide feedback to both cycles, helping to monitor progress, detect recurring issues, and align daily actions with long-term objectives. These cycles are supported by two structural pillars: Process Knowledge and Agile Teamwork. Together, they provide the understanding and collaboration needed to support improvement. *Process Knowledge* provides the analytical depth required to understand complex workflows, and interdependencies in RC and Agile *Teamwork* ensure coordination and shared accountability across roles and disciplines. At the top of the continuous improvement axis sits the Higher Purpose, represented as the Positive Impact on Society. Beyond acting as a mission statement, it provides direction and meaning, aligning daily actions to the broader societal and scientific value of the RC. In this sense, value is not defined in relation to a single customer, but to the wider public sphere that publicly funded research ultimately serves. The overall circular structure emphasizes iteration and adaptability (Romme, 1996), reinforcing the model as an evolving strategic tool that balances behavioural enablers with technical practices (Carvalho et al., 2019).

B. Principles and Implementation Guidelines

The model is underpinned by seven guiding principles, derived from the integrative OpEx classes of principles (section III.*A*) and adapted based on empirical evidence inherent to the specific context of RC. These guiding principles operate simultaneously as conceptual anchors and practical drivers for improvement and cultural alignment. Each principle encompasses a set of values and practices, offering a balanced view between behavioural and technical domains.

Principle 1 – *Promote a culture of dignity, inclusion, and individual initiative.*

In RC, promoting respect requires recalibration. In environments defined by advanced expertise and intellectual autonomy, respect is expressed through voice, initiative, and meaningful participation (Kallenberg, 2020). It involves recognising individuals as active contributors shaping the organisation, where intellectual recognition transcends hierarchy and values contributions across roles (Molinéro-Demilly et al., 2025). In such environment, encouraging initiative is not aspirational - it is fundamental for unlocking creativity and critical thinking. Human capital remains a stronger determinant of research performance. Conditions that enable focused work, transparent communication, and collaborative problem-solving reinforce dignity at work and strengthen engagement (Hines, 2022; Ryazanova and Jaskiene, 2022). Positioned at the base of the model as Respect for People, this principle forms the cultural foundation of both improvement cycles.

Examples of practices for implementing this principle include:

- Formalize inclusive decision-making structures: Include faculty, administrative, technical, research staff, and students in operational and improvement discussions. For example, establish cross-functional working groups and organise regular consultation sessions on process changes.
- Recognize intellectual contributions beyond formal roles: Value contributions to process improvement and collaborative culture alongside scientific output. Promote internal award initiatives (e.g., humorous or symbolic peer-voted prizes) to recognise contributions to team efficiency, mentoring, or collaborative problem-solving.

Principle 2 – Cultivate leadership that inspires, aligns, and enables long-term purpose.

In RC, leadership must be transformative, not merely directive (Ryazanova and Jaskiene, 2022). Traditional hierarchical leadership styles may fail to engage highly skilled professionals who value meaning, independence, and shared purpose. Transformational leadership connects personal motivation to organisational vision, inspires by example, and builds trust through transparency (Hines et al., 2022). Leaders of RC are often elected researchers without formal training in management, leadership, or organisational development. While this ensures scientific legitimacy, it can limit the adoption of transformational leadership behaviours that are essential to enable collective progress (Saeed et al., 2020). Evidence shows that investing in leadership development enhances operational effectiveness and team cohesion (Antony et al., 2023; Saeed et al., 2020). Rather than being positioned as a hierarchical layer within the model, transformational leadership is treated as embedded in the cultural foundation of Respect for People. In this view, leadership becomes the way respect is enacted in everyday practice, through transparency, active listening, shared decision-making, and long-term orientation. It is through these behaviours that cultural values are translated into coordinated action and alignment with the centre's mission. This view is consistent with perspectives that describe leadership as a relational and cultural process, rather than merely a structural function (Carvalho et al., 2017; Hines et al., 2022; Saeed et al., 2020)

While leadership practices must be tailored to each institutional setting, the following are examples of actions that support this principle:

- Formalise leadership development: Provide training in leadership, communication, and team management to directors, group leaders, and emerging leaders.
- Align leadership behaviour with institutional values: Promote regular coordination meetings, encourage regular informal meetings between research leaders (e.g., lean coffee sessions) to talk through ongoing challenges and find collaborative ways forward, such as pooled equipment purchases or shared lab budgets.

Principle 3 – Strengthening continuous improvement as the connecting axis of organizational progress.

Continuous improvement represents the core of sustainable OpEx and must be embedded in the organizational culture, supported by leadership and employee involvement (Sony, 2019). In RC, this process often emerges informally and in a fragmented way (Coccia, 2019a), relying on individual initiative and tacit knowledge within teams (Ryazanova and Jaskiene, 2022) rather than structured routines. The challenge is not its absence, but rather its lack of structure and alignment with the centre's strategic objectives. In the proposed model, continuous improvement plays a transversal role, acting as the central axis connecting the two improvement cycles (Fig. 1). The strategic cycle focuses on strengthening the conditions that support research excellence, while the operational cycle addresses day-to-day inefficiencies such as rework and process failures.

Making continuous improvement more visible and structured across the organisation is key to strengthening alignment and resilience (Bhasin and Found, 2020; Bradke et al., 2023). This requires methods and tools that translate the principle into concrete routines, supported by data, feedback cycles, and collective engagement.

Examples of how this principle can be implemented in RC are:

- Establish a cross-functional improvement team to define priorities, coordinate initiatives and communicate guidance to operational teams.
- Apply structured OpEx tools (operational cycle). For example, use PDCA cycles, A3 reports, Gemba Walks, process maps or root cause analysis to identify and resolve inefficiencies.
- Define and monitor performance indicators supporting long-term value creation (strategic cycle), such as lead times, response times, or internal satisfaction.

Principle 4 – Embed societal impact as the higher purpose of research activity.

In RC, excellence cannot be reduced to scientific output alone. A shared and visible sense of purpose, focused on delivering positive impact to society, is essential to guide collective efforts. Although often implicit, this purpose is rarely made explicit in institutional strategies or daily routines. Embedding higher purpose requires integrating societal relevance into how research is evaluated. This aligns with international calls for reform, such as the Coalition for Advancing Research Assessment (CoARA) (CoARA, 2023), which urges research organisations to go beyond bibliometrics and recognise diverse forms of impact. Alternative indicators, such as altmetrics, policy contributions, and partnerships, provide a broader view of value creation (Coccia, 2004; Kassab et al., 2020).

Positioned as the guiding purpose of the model, this principle links the cultural foundation of the RC to its broader societal mission. It provides direction to improvement efforts and clarifies why operational excellence matters in the first place. In this sense, the traditional OpEx notion of customer-oriented value is reframed as societal value creation. In publicly funded research, the "customer" is not a single actor but a broad set of stakeholders, including funders, scientific peers, industry partners, and society.

Examples of practices for implementing this principle include:

- Incorporate societal impact into recruitment and evaluation processes by including criteria such as contributions to policy, industry collaboration, technology transfer, or engagement with non-academic stakeholders.
- Encourage the use of impact narratives in CVs.
- Make the centre's purpose visible by communicating the societal impact of research (e.g., websites, outreach events).
- Use alternative metrics (e.g., policy citations, media uptake, engagement with non-academic audiences).

Principle 5 – Foster agile teamwork to enable adaptability and shared problem-solving

In RC, teamwork goes beyond collaboration within research groups; it represents a mechanism for adaptability, innovation, and resilience (Bradke et al., 2023; Coetzee et al., 2019). Teamwork has a positive effect on multiple performance dimensions, including employee attitudes, behaviour, operational efficiency, and financial outcomes, and its impact is amplified when combined with structural changes (Delarue et al., 2008). Agile frameworks reinforce this perspective by emphasising self-organised, cross-functional teams supported by continuous communication (Dingsøyr et al., 2012). This is particularly relevant in RC, where diverse roles, faculty, researchers, technicians, and administrative staff, must coordinate across disciplines. The value lies in the horizontal sharing of knowledge, where contributions from different roles accelerate problem identification and solution development (Coetzee et al., 2019). The organisational structure of RC, usually arranged into thematic research groups supported by transversal laboratories, reinforces this dynamic by enabling knowledge reuse and scaling across groups (Simonyte et al., 2022). Agile teamwork thus acts as a connector, enabling knowledge to flow across boundaries and supporting systemic growth. In the proposed model, agile teamwork is positioned as one of the two structural pillars, enabling coordination and supporting continuous improvement across organisational boundaries.

Examples of how this principle can be implemented in RC are:

- Establish cross-functional teams integrating scientific, technical, and administrative staff to manage key processes, with visible progress tracking (e.g., Kanban boards), aligned with the strategic improvement priorities defined at the organisational level.
- Promote collaboration in transversal laboratories. Share methods and models across groups through joint experiments, open training, or catalogues of validated techniques.
- Adopt agile routines: Use retrospectives and stand-up meetings in laboratories and offices to review progress (e.g., using burndown charts) and identify improvements.

Principle 6 – Enhance process knowledge to sustain effective research operations

Most OpEx reference models underline the importance of processes for achieving reliable results (Barros et al., 2025). Good intentions or individual effort cannot compensate for poorly designed processes (Deming, 1986). In practice, this means organisations need clear and resilient processes to support effective work (Bhasin and Found, 2020; Liker and Morgan, 2006). In RC, this principle is particularly relevant. RC activity depends on interconnected flows - scientific, management, administrative, and technical. When these flows are not aligned, inefficiencies emerge and research capacity is reduced. A process perspective goes beyond mapping tasks; it shows how resources are transformed into performance outcomes (Lee and Choi, 2003). Coccia's (2019) idea of an "organisational metabolism" reinforces this view: as in a biological system, if one part is blocked, the entire organism is affected. Similarly, value chain logic (Von Rosing et al., 2015; Zamora, 2016) highlights how weaknesses in support or management processes can compromise core research activities.

Empirical evidence from RC highlights excessive bureaucracy, fragmented workflows, rework, and unclear responsibilities as recurring barriers to organisational performance. Strengthening this dimension therefore involves mapping processes, clarifying responsibilities and ensuring that knowledge is shared rather than kept in silos. It also means recognising that support processes, often invisible until they fail, are as critical as core research tasks in enabling scientific output.

Within the model, process knowledge is placed as one of the two structural pillars, alongside agile teamwork. While teamwork provides adaptability, process knowledge offers stability and clarity. Together, sustain the link between daily operations and the broader societal purpose of research.

Examples of how this principle can be implemented in RC are:

- Map and make processes visible. Use mapping tools such as Business Process Model and Notation (BPMN), Value Stream Design in Indirect Areas (VSDia) or other.
- Implement standard work and simplify as much as possible. Create standard operation procedures to reduce misunderstandings and rework.
- Define responsibilities across process stages to avoid duplication and gaps. Create a competencies matrix.
- Review standards periodically to ensure relevance and compliance.

Principle 7 – Embed quality at the source to reduce rework and build reliability

In OpEx models, Quality at the Source represents the idea of "doing it right the first time", preventing errors from propagating through the system and avoiding costly late-stage correction (Deming, 1986). This principle highlights the need to embed quality upstream rather than relying on inspections or corrective actions. In RC, rework is a recurring issue, often linked to a lack of standardisation, informal practices, or reactive responses to problems (Hildebrandt et al., 2022). In RC, this implies designing and executing processes in ways that prevent errors at their origin, ensuring that outputs, whether experimental data, administrative procedures, or research deliverables, are correct from the outset. Embedding quality upstream builds trust in results and reduces waste of time and resources. It also strengthens institutional credibility.

Examples of how this principle can be implemented in RC are:

- Standardise core procedures. Use Standard Operating Procedures (SOP) for recurring activities (e.g., laboratory protocols, procurement steps, onboarding).
- Introduce error-proofing mechanisms (poka-yoke): Apply simple controls such as mandatory fields in digital forms, pre-experiment checklists. For example, double-checking the correct labelling of samples before freezing.
- Validate inputs before processing: Verify data, documentation, and materials at the beginning of processes to avoid downstream corrections. For example, in laboratory work, this includes checking that all reagents, materials, and equipment are available and functional before starting an experiment, as missing items or equipment failures may involve bureaucratic and time-consuming procedures.
- Foster a "stop and fix" culture. Encourage teams to pause workflows to resolve issues immediately, using simple tools such as the 5 Whys.

V. Discussion

This study addresses an important gap: while OpEx models are well established in industrial and service settings, no integrative model has been developed for RC. These organisations combine knowledge-intensive work, complex governance structures, and public accountability, making direct transfers from industry inadequate. The proposed model integrates guiding principles, structural pillars, and a dual-cycle improvement logic oriented towards societal impact (Fig. 2).

The foundations of most influential OpEx models, such as respect for people, continuous improvement, and leadership (Barros et al., 2025; Din et al., 2021; Dinis-Carvalho and Macedo, 2021; Found et al., 2018; Sony, 2019; Talwar, 2011) are present in the proposed RC model, confirming their relevance beyond industrial and service settings. However, RC operate under very different conditions: project cycles are often unpredictable, stakeholders are diffuse, bureaucracy is layered, and scientific, technical, and administrative activities are tightly interconnected. Empirical evidence from the diagnostic phase highlights operational constraints, including procurement delays and contract approvals, that reduce research capacity. Reflecting this context, OpEx principles were not incorporated uniformly in the RC model. Some required minor adjustments, while others demanded reinterpretation due to the specific conditions of RC. Respect for People and Transformational Leadership, reflected in the first and second guiding principles of the model, are closely interconnected. RC are composed of highly qualified professionals whose autonomy and expertise shape how coordination occurs in practice. The challenge is less about empowerment, as often discussed in industrial settings, and more about balancing individual autonomy with organisational coordination. Leadership roles are typically assumed by researchers whose authority derives from scientific merit rather than formal preparation in management. This reinforces academic legitimacy but can limit the adoption of structured leadership practices. In practice, this context highlights the need for mechanisms that support coordination and collective decision-making beyond hierarchical structures.

Continuous Improvement, reflected in the third guiding principle, plays a central role in the RC model. In RC, improvement activities are often present but are typically fragmented and informal, driven by necessity rather than structured routines. The model addresses this by positioning continuous improvement as the central axis through which OpEx principles become embedded at both tactical and strategic levels of RC organisation. This axis connects two interlinked cycles: a strategic cycle that consolidates the success factors sustaining research excellence, and an operational cycle that tackles systemic frictions. The direction of these improvement efforts is captured in principle four, which frames positive societal impact as the higher purpose guiding the model. In publicly funded research organisations, value creation is ultimately directed towards society, reflecting the public mandate of research and accountability to society (Bornmann, 2012; CoARA, 2023; Coccia, 2019a, 2008) rather than market value capture. To operationalise this improvement logic, the model is sustained by two structural pillars: Process Knowledge and Agile Teamwork. Teamwork, reflected in the fifth guiding principle, represents a key adaptation. It is elevated to the level of a structural pillar, reflecting its role in enabling coordination and addressing operational challenges. This interpretation resonates with Coetzee, van Dyk and van der Merwe (2019), who describe teamwork as a concrete expression of the Toyota "Respect for People" principle, fostering trust, knowledge sharing, and shared accountability. Likewise, Delarue et al. (2008) provide robust evidence that teamwork has a positive impact across attitudinal,

behavioural, operational, and financial outcomes. Agile frameworks further reinforce this perspective by emphasising self-organising and cross-functional teams supported by continuous communication (Dingsøyr et al., 2012; Moe et al., 2010). This is particularly relevant in RC, where transversal laboratories and multifunctional teams connect different disciplines and roles. In this context, teamwork acts as a mechanism for coordination and organisational resilience. Process knowledge, reflected in the sixth guiding principle, represents the second structural pillar of the model. It complements teamwork by providing visibility over the complex set of processes that sustain research activity. Empirical findings from the diagnostic phase highlight recurring issues such as bureaucratic delays, fragmented workflows, and unclear responsibilities. Strengthening process knowledge supports more coordinated and reliable operations.

The class Quality at the Source was retained but reframed for RC through the seventh guiding principle. In most referenced models, this principle is largely assumed (Barros et al., 2025), framed by Deming (1986) as "doing it right the first time", by Shingo (1986) as immediate defect detection through error-proofing mechanisms, and by Liker and Morgan (2006) as ensuring quality from the earliest stages of development. In RC, however, this logic does not manifest in the same way. Unlike production lines or software development, where defects are typically visible and contained, research tasks are less repetitive and often fragmented. Moreover, RC operate within institutional environments characterised by layered governance structures and strong bureaucratic procedures. Responsibilities tend to be diffuse, and these structures can delay or complicate error detection (Bradke et al., 2023; Coccia, 2019a). Coccia (2009) further argues that bureaucratization in public research settings undermines efficiency and performance. Under these conditions, mistakes in administrative procedures, procurement processes, or experimental planning may remain unnoticed for extended periods, generating rework and delays. Empirical findings from the diagnostic phase reflect similar patterns, highlighting the importance of "first-time quality" and proactive planning, while also pointing to rework, missing inputs, and administrative returns. These findings align with Hildebrandt et al. (2022), who show how such errors can cascade into costly delays. In such environments, quality cannot be taken for granted. This is why the RC model makes Quality at the Source explicit as a guiding principle rather than an implicit assumption.

Research activities are typically organised around projects, funding cycles, and exploratory processes that are difficult to synchronise through demand-driven flow mechanisms. For this reason, the integrative class Pull Flow (section III.A), was not embedded in the model, as pull logic does not align with these dynamics. The integrative class Customer-oriented value also did not translate into a distinct guiding principle. Instead, its underlying logic was reframed through the emphasis on societal impact, as defined in the fourth guiding principle. In RC, value creation is directed toward society as the ultimate beneficiary of research activity. In this context, the notion of "customer" remains ambiguous, extending beyond transactional stakeholders to include funders, scientific peers, industry partners, and society. This reflects the need for further conceptual clarification before these ideas can be operationalised in RC. This approach avoids forcing unsuitable concepts and keeps the model closer to the operational reality of RC.

To position the proposed model within the OpEx field, it can be related to three main streams: TPS-based models (Toyota Way, Shingo, Lean), TQM-based models (Deming, EFQM, MBNQA), and software-based models (Agile, DevOps). Table I summarises how the proposed RC model relates to these streams while introducing specific adaptations for research organisations. The model places strong emphasis on the operationalisation of excellence principles through practical guidelines that support their implementation in everyday organisational practices. This is particularly relevant for RC operating within larger institutional structures, where limited managerial autonomy can constrain organisational change. In such contexts, the model aims not only to identify what should be improved, but also to provide guidance on how excellence principles can be translated into actionable practices.

TABLE I. POSITIONING OF THE PROPOSED OPEX MODEL FOR RC IN RELATION TO MAJOR OPEX STREAMS

	TPS-based (Toyota, Shingo, Lean)	TQM-based (Deming, EFQM, MBNQA)	Software-based (Agile, DevOps)	RC-based (Proposed model)
Purpose	Deliver customer value with minimal waste	Organisational excellence and stakeholder satisfaction	Fast delivery and adaptability	**Societal impact and public value creation**
Customer	End user	Multiple stakeholders	End user / client	**Diffuse beneficiaries (funders, peers, industry, society)**
Improvement logic	*Kaizen* (continuous improvement)	PDCA cycle, self-assessment, benchmarking	Continuous feedback and short development cycles	**Dual-cycle improvement (strategical and operational)**
Role of teamwork	Implicit in "Respect for People"	Recognised, but implicit	Central organizing principle (self-organizing teams)	**Structural pilar: Agile teamwork**
Process perspective	Flow and waste elimination	Process standardisation and quality systems	Adaptive workflows and flexible processes	**Structural pillar: Process knowledge**
Quality at the source	*Jidoka* (stop to fix problems, make problems visible)	Prevention through system-wide quality management	Shift-left testing and shared responsibility	**Context-specific practices**
Operationalization of principles	Operational tools and production practices	Assessment frameworks and maturity models	Team routines and development practices	**Guiding principles translated into practical guidelines**

This emphasis on operationalisation is particularly important in research environments, where excellence cannot be achieved solely through the availability of resources or scientific recognition. In RC, efficiency cannot be attained through large grants or awards. It also depends on how effectively resources are mobilised and transformed into

knowledge with societal impact. Agostino et al. (2012) argue that RC should be assessed not only by scientific outputs, but also by their capacity to use resources efficiently and respond to multiple stakeholders. Coccia (2008) reinforces the point, showing that measuring performance requires considering both inputs and outputs to drive strategic change. By contrast, Roshani et al. (2021) show that funded research often achieves higher citation impact, yet this does not necessarily translate into broader societal value and may reinforce cumulative advantage effects. Taken together, these perspectives suggest that sustainable excellence in RC rests on more than recognition, funding, or output metrics; it depends on how operational practices convert resources into knowledge that generates societal value. Overall, the model advances OpEx in RC, placing the emphasis on implementing its principles in complex, knowledge-intensive environments through concrete, practical guidance, rather than focusing solely on the conceptual adaptation/adoption of those principles.

VI. Conclusion

This study addressed the limited applicability of existing OpEx models to RC, proposing a model adapted to the specific characteristics of these organisations. Unlike industrial and service settings, RC operate under conditions of project unpredictability, diffuse responsibilities, and public accountability, requiring the reinterpretation rather than direct transfer of OpEx principles.

The findings show that while some principles can be readily adapted to the RC context, others require reinterpretation. This reinforces the need to contextualise OpEx concepts when applied to knowledge-intensive and publicly funded environments. These observations highlight the broader contribution of the proposed model. It extends OpEx into the RC context, an area that remains largely unexplored, and shifts the focus from principles' adoption to their operational translation through structured improvement cycles and practical implementation guidelines.

From a practical perspective, the model provides guidance for RC leaders and staff by highlighting priorities such as teamwork, process clarity, and the reduction of rework, while linking operational practices to the broader societal mission of research. In this sense, the model emphasises that the efficiency of everyday organisational practices shapes the conditions under which research centres generate knowledge and deliver impact. More efficient use of resources and better-coordinated processes create the conditions for research activities to be carried out more effectively, aligning daily operations with the broader purpose of producing positive societal outcomes.

The model is empirically grounded, having been developed through an AR approach combining literature review, interviews, and a questionnaire applied in a Portuguese RC. While the empirical phase focused on a single case study, this reflects the original objective of analysing that specific organisational context in depth, with the resulting model expected to be transferable to other RC with similar characteristics.

Future work will focus on the implementation and evaluation phases of the AR cycle (phases 4 and 5, Fig. 1), where the model will be tested and refined through practical application. Broader validation across different RC and contexts will be essential to assess its transferability, as well as to further explore less developed principles, namely Pull Flow and Customer-Oriented Value.

Overall, the proposed model offers a structured yet flexible framework for supporting OpEx in RC, contributing to both academic knowledge and practical improvement.

Acknowledgment

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Unit Project Scope UID/00319/2025 - Centro ALGORITMI (ALGORITMI/UM) https://doi.org/10.54499/UID/00319/2025.

References

Agostino, D., Arena, M., Azzone, G., Dal, M. and Masella, C. (2012), "Developing a performance measurement system for public research centres", International Journal of Business Science and Applied Management, Vol. 7 No.1, pp. 43–60.

Åhlström, P. et al., (2021), "Is lean a theory? Viewpoints and outlook", International Journal of Operations and Production Management, Vol. 41No.12, pp. 1852–1878. Available at: https://doi.org/10.1108/IJOPM-06-2021-0408.

Antony, J. et al., (2023), "Critical success factors for operational excellence initiatives in manufacturing: a meta-analysis", Total Quality Management and Business Excellence, Vol. 34 Nos 9–10, pp. 1152–1172. Available at: https://doi.org/10.1080/14783363.2022.2157714.

Antony, J., Rodgers, B. and Cudney, E. (2017), "Lean Six Sigma for public sector organizations: is it a myth or reality?", International Journal of Quality and Reliability Management, Vol. 34 No. 9, pp. 1402–1411. Available at: https://doi.org/10.1108/IJQRM-08-2016-0127.

Barros, A., Sousa, R.M. and Dinis-Carvalho, J. (2025a), "A Study on the Adoption of Operational Excellence Principles in the Organization and Management of University Research Centres", in O. McDermott et al. (eds). Galway: Springer Nature Switzerland, pp. 82–94. Available at: https://doi.org/http://dx.doi.org/10.1007/978-3-031-84816-2_7.

Barros, A., Sousa, R.M. and Dinis-Carvalho, J. (2025b), "Towards Structured Integration of Operational Excellence Principles: A Cross Model Analysis", unpublished.

Bhasin, S. and Found, P. (2020), "Sustaining the lean ideology", Management Decision, Vol. 59 No. 3, pp. 568–585. Available at: https://doi.org/10.1108/MD-09-2019-1254.

Bornmann, L. (2012), "Measuring the societal impact of research. Research is less and less assessed on scientific impact alone-we should aim to quantify the increasingly important contributions of science to society", EMBO Reports, Vol. 13 No. 8, pp. 673–676. Available at: https://doi.org/10.1038/embor.2012.99.

Bradke, F., Maartens, A. and Teichmann, S.A. (2023), "Key attributes of successful research institutes", PLoS Biology, Vol. 21 No. 9, pp. 1–15. Available at: https://doi.org/10.1371/journal.pbio.3002267.

Caiado, R.G.G., Carocha, D.M., Goulart, A.K. and Tortorella, G.L. (2020), "Critical success factors-based taxonomy for Lean Public Management: a systematic review", Production, Vol. 30, pp. 1–15. Available at: https://doi.org/10.1590/0103-6513.20200030

Carvalho, A.M., Sampaio, P., Rebentisch, E. and Saraiva, P. (2017), "Operational excellence as a means to achieve an enduring capacity to change – revision and evolution of a conceptual model", Procedia Manufacturing, Vol. 13, pp. 1328–1335. Available at: https://doi.org/10.1016/j.promfg.2017.09.109.

Carvalho, A.M., Sampaio, P., Rebentisch, E. and Saraiva, P. (2019), "Operational excellence, organisational culture and agility: the missing link?", Total Quality Management and Business Excellence, Vol. 30 Nos. 13–14, pp. 1495–1514. Available at: https://doi.org/10.1080/14783363.2017.1374833.

Carvalho, J.D. (2021), Melhoria Contínua nas Organizações, Lidel, Lisboa.

CoARA (2023), "Coalition for Advancing Research Assessment (CoARA)". Available at: https://coara.eu/.

Coccia, M. (2004), "New models for measuring the R&D performance and identifying the productivity of public research institutes", R and D Management, Vol. 34 No. 3, pp. 267–280. Available at: https://doi.org/10.1111/j.1467-9310.2004.00338.x.

Coccia, M. (2008), "Measuring scientific performance of public research units for strategic change", Journal of Informetrics, Vol. 2 No. 3, pp. 183–194. Available at: https://doi.org/10.1016/j.joi.2008.04.001.

Coccia, M. (2009), "Research performance and bureaucracy within public research labs", Scientometrics, Vol. 79 No. 1, pp. 93–107. Available at: https://doi.org/10.1007/s11192-009-0406-2.

Coccia, M. (2019a), "Metabolism of Public Research Organizations: How Do Laboratories Consume State Subsidies?", Public Organization Review, Vol. 19 No. 4, pp. 473–491. Available at: https://doi.org/10.1007/s11115-018-0421-y.

Coccia, M. (2019b), "Why do nations produce science advances and new technology?", Technology in Society, Vol. 59(March), p. 101124. Available at: https://doi.org/10.1016/j.techsoc.2019.03.007.

Coetzee, R., van Dyk, L. and van der Merwe, K.R. (2019), "Towards addressing respect for people during lean implementation", International Journal of Lean Six Sigma, Vol. 10 No. 3, pp. 830–854. Available at: https://doi.org/10.1108/IJLSS-07-2017-0081.

Coughlan, P. and Coghlan, D. (2002), "Action research for operations management", International Journal of Operations and Production Management, Vol. 22 No. 2, pp. 220–240. Available at: https://doi.org/10.1108/01443570210417515.

Delarue, A., Van, G.H., Procter, S. and Burridge, M. (2008), "Teamworking and organizational performance: A review of survey-based research", International Journal of Management Reviews, Vol. 10 No. 2, pp. 127–148. Available at: https://doi.org/10.1111/j.1468-2370.2007.00227.x.

Deming, W.E. (1986), Out of the Crisis. Cambridge, MA: MIT Press.

Din, M., Asif, M., Awan, U. and Thomas, G. (2021), "What makes excellence models excellent: a comparison of the American, European and Japanese models", TQM Journal, Vol. 33 No. 6, pp. 1143–1162. Available at: https://doi.org/10.1108/TQM-06-2020-0124.

Dingsøyr, T., Nerur, S., Balijepally, V. and Moe, N.B. (2012), "A decade of agile methodologies: Towards explaining agile software development", Journal of Systems and Software, Vol. 85 No. 6, pp. 1213–1221. Available at: https://doi.org/10.1016/j.jss.2012.02.033.

Dinis-Carvalho, J. and Macedo, H. (2021), "Toyota Inspired Excellence Models", IFIP Advances in Information and Communication Technology, Vol. 610, pp. 235–246. Available at: https://doi.org/10.1007/978-3-030-92934-3_24.

FCT - Fundação para a Ciência e a Tecnologia (2025). Available at: https://www.fct.pt/sobre/documentos-de-gestao/ (Accessed: 26 January 2025).

Found, P., Griffiths, G., Harrison, R. and McDermott, O. (2018), "Towards a theory of operational excellence", Total Quality Management and Business Excellence, Vol. 29 Nos. 9–10, pp. 1012–1024. Available at: https://doi.org/10.1080/14783363.2018.1486544.

Hildebrandt, M.G. et al., (2022), "How to increase value and reduce waste in research: initial experiences of applying Lean thinking and visual management in research leadership", BMJ Open, Vol. 12 No. 6, pp. 1–8. Available at: https://doi.org/10.1136/bmjopen-2021-058179.

Hines, P. et al. (2022), "Turning the lean world upside down", International Journal of Lean Six Sigma, Vol. 13 No. 5, pp. 989–1024. Available at: https://doi.org/10.1108/IJLSS-09-2021-0166.

Jengwa, E. and Pellissier, R. (2022), "An operational excellence strategy implementation model for growth in a volatile, uncertain, complex, and ambiguous environment", Acta Commercii, Vol. 22 No. 1, pp. 1–11. Available at: https://doi.org/10.4102/ac.v22i1.960.

Kallenberg, T. (2020), "Differences in influence: different types of university employees compared", Tertiary Education and Management, Vol. 26 No. 4, pp. 363–380. Available at: https://doi.org/10.1007/s11233-020-09058-w.

Kassab, O., Bornmann, L. and Haunschild, R. (2020), "Can altmetrics reflect societal impact considerations?: Exploring the potential of altmetrics in the context of a sustainability science research center", Quantitative Science Studies, Vol. 1 No. 2, pp. 792–809. Available at: https://doi.org/10.1162/qss_a_00032.

Lee, H. and Choi, B. (2003), "Knowledge management enablers, processes, and organizational performance: An integrative view and empirical examination", Journal of Management Information Systems, Vol. 20 No. 1, pp. 179–228. Available at: https://doi.org/10.1080/07421222.2003.11045756.

Liker, J.K. and Morgan, J.M. (2006), "The toyota way in services: The case of lean product development", Academy of Management Perspectives, Vol. 20 No. 2, pp. 5–20. Available at: https://doi.org/10.5465/AMP.2006.20591002.

Lukrafka, T.O., Silva, D.S. and Echeveste, M. (2020), "A geographic picture of Lean adoption in the public sector: Cases, approaches, and a refreshed agenda", European Management Journal, Vol. 38 No. 3, pp. 506–517. Available at: https://doi.org/10.1016/j.emj.2020.02.004.

Mengist, W., Soromessa, T. and Legese, G. (2020), "Method for conducting systematic literature review and meta-analysis for environmental science research", MethodsX, Vol. 7, p. 11. Available at: https://doi.org/10.1016/j.mex.2019.100777.

Moe, N.B., Dingsøyr, T. and Dybå, T. (2010), "A teamwork model for understanding an agile team: A case study of a Scrum project", Information and Software Technology, Vol. 52 No. 5, pp. 480–491. Available at: https://doi.org/10.1016/j.infsof.2009.11.004.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. (2010), "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement", International Journal of Surgery, Vol. 8 No. 5, pp. 336–341. Available at: https://doi.org/10.1016/j.ijsu.2010.02.007.

Molinéro-Demilly, V. et al. (2025), "Effects of quality management on well-being at work and performance in a research laboratory: an exploratory study", International Journal of Quality and Reliability Management, Vol. 42 No. 5, pp. 1317–1333. Available at: https://doi.org/10.1108/IJQRM-12-2023-0379.

Narayanamurthy, G., Gurumurthy, A. and Chockalingam, R. (2017), "Applying lean thinking in an educational institute – an action research", International Journal of Productivity and Performance Management, Vol. 66 No. 5, pp. 598–629. Available at: https://doi.org/10.1108/IJPPM-07-2016-0144.

Pekkanen, P. and Pirttilä, T. (2022), "Output measurement in professional public organizations: insights from European justice systems", International Journal of Productivity and Performance Management, Vol. 73 No. 11, pp. 1–19. Available at: https://doi.org/10.1108/IJPPM-05-2021-0279.

Radnor, Z. and Walley, P. (2008), "Learning to walk before we try to run: adapting lean for the public sector", Public Money and Management, Vol. 28 No. 1, pp. 13–20. Available at: https://doi.org/10.1111/j.1467-9302.2008.00613.x.

Rahmati, M.H. and Jalilvand, M.R. (2024a), "An optimal organizational excellence model for the public sector", International Journal of Quality and Reliability Management, Vol. 41 No. 3, pp. 944–963. Available at: https://doi.org/10.1108/IJQRM-03-2023-0098.

Rodgers, B. and Antony, J. (2019), "Lean and Six Sigma practices in the public sector: a review", International Journal of Quality and Reliability Management, Vol. 36 No. 3, pp. 437–455. Available at: https://doi.org/10.1108/IJQRM-02-2018-0057.

Romme, G. (1996), "Making organizational learning work: Consent and double linking between circles", European Management Journal, Vol. 14 No. 1, pp. 69–75. Available at: https://doi.org/10.1016/0263-2373(95)00048-8.

Roshani, S., Bagherylooieh, Mohammad-Reza, Mosleh, M. and Coccia, M. (2021), "What is the relationship between research funding and citation-based performance? A comparative analysis between critical disciplines", Scientometrics, Vol. 126 No. 9, pp. 7859–7874. Available at: https://doi.org/10.1007/s11192-021-04077-9.

Von Rosing, M. et al. (2015), "Phase 3: Process concept evolution, The Complete Business Process Handbook: Body of Knowledge from Process Modeling to BPM" Elsevier Inc. Available at: https://doi.org/10.1016/B978-0-12-799959-3.00003-3.

Ryazanova, O. and Jaskiene, J. (2022), "Managing individual research productivity in academic organizations: A review of the evidence and a path forward", Research Policy, Vol. 51 No. 2, p. 104448. Available at: https://doi.org/10.1016/j.respol.2021.104448.

Saeed, B., Tasmin R., Ayyaz M. and Aamer H. (2020), "Exploring the impact of transformational leadership and human resource practices on operational excellence mediated by knowledge sharing: A conceptual framework", International Journal of Scientific and Technology Research, Vol. 9 No. 2, pp. 4458–4468.

Shamsuzzaman, M. et al. (2023), "Improving the admission process in a higher education institute using lean six sigma: a case study", International Journal of Lean Six Sigma, Vol. 14 No. 2, pp. 525–554. Available at: https://doi.org/10.1108/IJLSS-05-2021-0098.

Shingo, S. (1986), "Zero Quality Control", Cambridge, Massachusetts: Productivity Press.

Simonyte, S., Adomaitiene, R. and Ruzele, D. (2022), "Experience of lean application in higher education institutions", International Journal of Lean Six Sigma, Vol. 13 No. 2, pp. 408–427. Available at: https://doi.org/10.1108/IJLSS-11-2020-0208.

Sony, M. (2019), "Implementing sustainable operational excellence in organizations: an integrative viewpoint", Production and Manufacturing Research, Vol. 7 No. 1, pp. 67–87. Available at: https://doi.org/10.1080/21693277.2019.1581674.

Talwar, B. (2011), "Business excellence models and the path ahead", TQM Journal, Vol. 23 No. 1, pp. 21–35. Available at: https://doi.org/10.1108/17542731111097461.

Zamora, E.A. (2016), "Value Chain Analysis: A Brief Review", Asian Journal of Innovation and Policy, Vol. 5 No. 2, pp. 116–128. Available at: https://doi.org/10.7545/ajip.2016.5.2.116.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Towards Zero Defect and Zero Waste Food Systems: Extending ZDM to Controlled Environment Agriculture

Victor Azamfirei Ionita[1, 2, 3] and Foivos Psarommatis[4, 5]

1 DEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, FCT NOVA, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal; v.ionita@fct.unl.pt
2 UNIDEMI, Research and Development unit for Mechanical and Industrial Engineering, FCT NOVA, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
3 Laboratório Associado de Sistemas Inteligentes, LASI, 4800-058 Guimarães, Portugal
4 University of Oslo, SIRIUS, Centre for Scalable Data Access, Gaustadalleen 23B, 0373, Oslo, Norway
5 Zerofect GmbH, Albisblick 49, Baar, Switzerland

Abstract: This paper extends the Zero-Defect Manufacturing (ZDM) paradigm to Controlled Environment Agriculture (CEA), introducing the concepts of Zero-Defect Crops and Zero-Waste CEA. A qualitative theory-building approach based on the Gioia methodology is employed, supported by a Systematic Literature Mapping (SLM) of CEA and smart agriculture studies. The analysis identifies key first order concepts and second order themes, structured under the ZDM dimensions of Detect, Predict, Prevent, and Repair. The study develops a structured conceptual framework that maps CEA capabilities onto the four ZDM dimensions of Detect, Predict, Prevent, and Repair, and uses this structure to define the constructs of Zero-Defect Crops and Zero-Waste CEA. The findings suggest that CEA can be interpreted as a cyber-physical biological production system in which crop quality assurance and resource efficiency are jointly governed through closed-loop sensing, prediction, and adaptive control. As a conceptual study grounded in literature, the framework now requires empirical validation across crops, facility types, and operational settings.

Keywords: *Zero-Defect Manufacturing (ZDM), Controlled Environment Agriculture (CEA), Zero Waste, Smart Agriculture, Cyber-Physical Systems*

I. Introduction

In recent decades, the global food system has been confronted with a set of converging pressures that challenge the assumptions underlying conventional agricultural production. Rising world population, accelerating climate variability, land degradation, and the spatial concentration of demand in rapidly urbanising regions have combined to expose the structural fragility of food supply chains reliant on open-field cultivation and seasonally constrained growing cycles. Controlled Environment Agriculture (CEA) has emerged as a technological response to these pressures, enabling the year-round, location-independent production of crops within enclosed, instrumentally managed systems, including vertical farms, plant factories, and smart greenhouses. By decoupling production from climatic and geographic constraints, CEA systems offer a pathway to food supply resilience that is particularly significant in the context of climate adaptation. Leiding et al. (2026) frame CEA as a central component of future resilient food production, while Dutta et al. (2025) position soilless precision farming as a strategic opportunity for global food security. The convergence of affordable sensor hardware, cloud computing, and machine learning has further accelerated CEA adoption, enabling levels of environmental control and crop monitoring that were impractical at commercial scale only a decade ago.

However, despite these technological advances, CEA systems continue to exhibit persistent inefficiencies that limit both their economic viability and their environmental credentials. Energy consumption remains one of the most critical constraints: artificial lighting, climate regulation, and fertigation systems collectively account for a substantial fraction of operating costs, and their sub-optimal management generates avoidable resource losses across the production cycle. Wang et al. (2025) demonstrate that even lighting system design, typically treated as a capital rather than operational concern, has measurable implications for energy efficiency in plant factories. while Hadj Abdelkader et al. (2023) show that IoT-based automated light control architectures can substantially reduce energy expenditure in indoor farms while maintaining optimal growth conditions, with scalable results confirmed across large-scale production configurations. Nutrient and water management present analogous challenges: Patel et al. (2024) show that conventional fertigation schedules routinely result in nutrient wastage in hydroponic systems, while Costantino et al. (2021) document the energy performance penalties associated with poorly controlled greenhouse climate systems. Beyond resource flows, crop quality variability represents a further unresolved inefficiency: the same biological stochasticity that makes living production systems adaptive also makes them prone to yield inconsistency, morphological deviation, and unpredictable harvest timing. These losses are not merely operational inconveniences; at scale, they undermine the economic case for CEA investment and erode its sustainability credentials. The paradox of CEA is therefore that the very precision it promises has not yet been systematically converted into guaranteed outcomes.

Zero-Defect Manufacturing (ZDM) offers a conceptual framework for resolving this paradox. Originating in industrial quality engineering and formalised as a comprehensive production paradigm by Psarommatis et al. (2020), ZDM organises quality assurance around four operational dimensions, i.e., Detect, Predict, Prevent, and Repair, structured to enable production systems to identify deviations, anticipate failures, preclude their occurrence, and recover from disruptions when prevention is insufficient. In advanced manufacturing, this framework has been operationalised through a combination of real-time sensing,

data-driven modelling, and closed-loop control, enabling production systems to progressively narrow their variability envelope through continuous learning. The functional parallels between this architecture and the instrumented, sensor-rich environment of modern CEA are evident: CEA systems already deploy many of the constituent technologies that ZDM requires, including computer vision, IoT-based monitoring, and adaptive control. Yet these capabilities have developed in a fragmented and largely atheoretical manner, without the integrative logic that would allow them to function as a coordinated quality and resource assurance system.

This study addresses this gap directly. While ZDM has been extensively studied and applied in manufacturing contexts, no systematic framework exists for its application to agricultural production systems, and particularly not to biological production environments where variability is dynamic, adaptive, and governed by genotype-environment interaction rather than mechanical tolerance. This paper bridges that gap by extending the ZDM paradigm to CEA through a qualitative theory-building approach, introducing two novel constructs, namely Zero-Defect Crops, defined as agricultural outputs produced in controlled environments that consistently meet predefined quality, yield, and safety standards through adaptive detection and control of deviations throughout the crop lifecycle, and Zero-Waste CEA, defined as a production paradigm in which resource flows are continuously monitored, optimised, and recirculated to eliminate inefficiencies and ensure closed-loop resource utilisation. Together, these constructs constitute a unified production paradigm that repositions CEA from a technologically advanced but conceptually fragmented practice into a prescriptive, system-level quality and efficiency framework grounded in established manufacturing theory.

Despite the rapid development of sensing, automation, optimisation, and AI-based control in CEA, the literature remains fragmented across technical sub-problems such as growth monitoring, climate control, resource optimisation, and predictive modelling. What is still lacking is an integrative production-systems framework that can organise these capabilities within a coherent quality and resource assurance logic. This gap is significant because, without such a framework, CEA technologies risk remaining a collection of isolated smart functions rather than evolving into truly closed-loop production systems. In response, this paper asks whether the established ZDM logic of Detect, Predict, Prevent, and Repair can be systematically translated from industrial manufacturing to CEA. Using a Gioia-based theory-building approach supported by systematic literature mapping, the paper develops a conceptual framework for Zero-Defect Crops and Zero-Waste CEA and discusses its theoretical and operational implications for future agricultural cyber-physical systems.

The remainder of the paper is structured as follows. Section II describes the research design, detailing the Systematic Literature Mapping process used for data collection and the Gioia methodology applied for theory building. Section III presents the analysis and theoretical interpretation of the identified aggregate dimensions within the ZDM-CEA framework. Section IV develops the Discussion, examining the cross-dimensional couplings and broader system-level implications. Section V presents the Conclusions, summarising the theoretical and practical contributions of the study and outlining directions for future research.

II. Methods and Materials

This study adopts a qualitative theory-building design to develop a conceptual extension of the Zero-Defect Manufacturing (ZDM) paradigm into Controlled Environment Agriculture (CEA). The methodological approach combines Systematic Literature Mapping (SLM) for structured evidence collection with Gioia's methodology for concept development and theoretical abstraction. This combination enables the study to move from a fragmented body of literature on smart agriculture and CEA toward a coherent framework structured around the ZDM dimensions of Detect, Predict, Prevent, and Repair. On this basis, the paper introduces the concepts of Zero-Defect Crops and Zero-Waste CEA as analytically linked dimensions of a closed-loop biological production paradigm.

A. Data collection through Systematic Literature Mapping (SLM)

The data collection process was conducted using a SLM approach to identify and structure relevant research in CEA, smart agriculture, and resource efficient food production systems. SLM extends traditional systematic literature review methods by enabling a structured and visual representation of research domains, supporting the identification of research gaps and emerging themes. The SLM process follows established principles of systematic reviews, including the definition of research objectives, conceptual boundaries, and inclusion criteria (Cahyo, 2021). Particular attention was given to studies addressing automation not only in terms of physical operations (e.g. harvesting), but also in decision-support systems, optimisation strategies, and intelligent control mechanisms. The aim was to gain insights on the considerations. In this state-of-the-art study, we have gathered bibliometric data from the most widely referred search engines for academic engineering papers, i.e. Scopus. The search fields for each paper were the title, abstract, and keywords specified by the author/s in the papers, with variations depending on the database, i.e. with Scopus, the tag combination was TITLE-ABS-KEY. The search has been performed on 30th of March 2026 and the search string was compounded by "controlled environment agriculture" AND ("automat*" OR "smart" OR "optimisation"). The search was limited to articles from all years, written in English, and within all fields of engineering.

To ensure methodological transparency, the search results were screened in multiple stages. First, duplicate and clearly irrelevant records were removed based on title and abstract review. Second, the remaining papers were assessed for conceptual relevance to at least one of the four focal capabilities later aligned with the ZDM framework: real-time detection, predictive modelling, preventive control, or corrective/adaptive response in CEA or closely related smart agriculture contexts. Studies focused exclusively on conventional open-field agriculture, purely biological crop science without digital or control relevance, or physical automation unrelated to monitoring, decision-making, or closed-loop regulation were excluded. The final corpus comprised [insert number] papers, which formed the evidentiary basis for first-order concept extraction and second-order thematic abstraction. Although the purpose of the review was theory building rather than bibliometric exhaustiveness, this structured screening process improved the traceability and reproducibility of the conceptual synthesis.

Scopus was selected as the primary database because of its broad coverage of engineering, applied technology, and interdisciplinary agricultural systems research, which aligns with the cross-domain scope of this study.

B. Theory Building using Gioia's approach

For the data analysis and conceptual synthesis, this study adopts the qualitative methodology proposed by Gioia et al. (2013), which enables the development of rigorously grounded theoretical frameworks through a structured, multi-stage coding process. Following this approach, first-order concepts were extracted directly from the literature as informant-centric terms, preserving the original meaning of the reviewed studies, and subsequently grouped into second-order themes through iterative abstraction and researcher-centric interpretation. These themes were then consolidated into aggregate dimensions aligned with the established ZDM framework (Psarommatis et al., 2020), i.e., Detect, Predict, Prevent, and Repair, thereby enabling a hybrid inductive-deductive theory-building process; see **Table 1**. Within this framework, *Zero-Defect Crops* are conceptualised as "agricultural outputs produced in controlled environments that consistently meet predefined quality, yield, and safety standards through continuous sensing/monitoring, predictive analytics, and adaptive control mechanisms that detect, predict, prevent, and correct deviations (minimise variability) throughout the crop lifecycle". In parallel, *Zero-Waste CEA* is defined as a "production paradigm in which water, nutrients, energy, and biomass flows are continuously monitored, optimised, and recirculated to minimise losses, eliminate inefficiencies and ensure closed-loop resource utilisation across the entire agricultural system". This approach ensures both methodological rigour and theoretical coherence, enabling the systematic translation of ZDM principles from industrial manufacturing to biological production systems. Unlike a purely inductive Gioia design in which aggregate dimensions emerge entirely from the data, this study uses a theory-anchored abductive logic. Specifically, the aggregate dimensions were predefined through the validated ZDM structure of Detect, Predict, Prevent, and Repair, while the first-order concepts and second-order themes were derived iteratively from the reviewed literature. This preserves interpretive openness at the level of empirical coding while maintaining theoretical coherence at the framework level.

III. Analysis and Theoretical Interpretations

This section presents the coded evidence and theoretical interpretation of the literature through the four ZDM dimensions of Detect, Predict, Prevent, and Repair. The resulting data structure comprises 25 first-order concepts extracted from the literature, 13 second-order themes generated through cross-domain abstraction, and four aggregate dimensions aligned with the ZDM paradigm i.e., Detect, Predict, Prevent, and Repair. Table 1 presents the Gioia data structure developed in this study. It shows how literature-derived first-order concepts were grouped into second-order themes and then aligned with the four aggregate dimensions of the ZDM framework. The table serves as the analytical bridge between the reviewed CEA literature and the conceptual constructs developed in this paper.

TABLE I. Gioia method

First order Concepts	Second-order Themes	Aggregated Dimensions of ZDM
"real-time autonomous horticulture monitoring system… viewing greenhouse crop data and camera feed of plants" "non-destructive and continuous plant weight measurement system" Aji et al. (2020)	**Real-time plant state sensing & monitoring**	**Detect**
"early detection of nutrient deficiencies and plant stress through sensor data"	**Early anomaly identification**	
"environmental parameters such as temperature, humidity, and CO_2 are continuously tracked"	**System-wide condition tracking**	
"machine learning models are used to predict crop yield and performance" Patel et al. (2024) "noninvasive estimation of leaf pigments" Cammarisano et al. (2022) "scalable, data-driven method to predict yield in CEA" Rahman et al. (2025)	**AI/ML Predictive crop analytics**	**Predict**
"forecasting plant growth and optimal harvest time using time-series models" "hybrid model systematically captures environmental variability, simulates plant growth" Kharraz & Szabó (2025)	**Growth forecasting and Harvest prediction**	
"data-driven models simulate crop-environment interactions" "model simulates the operation of the systems and the energy performance" Costantino et al. (2021) "two-stage MLP-lookup table model… decouples environmental predictions from equipment-specific performance mapping" (Choi et al., 2026) "model for predicting wind speed and direction" Kurafeeva et al. (2024)	**Digital crop modelling / Physics-informed and hybrid system modelling**	
"adaptive lighting systems dynamically adjust based on plant requirements"	**Multi-objective environmental parameter regulation**	**Prevent**
"precision fertigation enables optimal nutrient delivery to crops" "real-time adjustments...minimizes human intervention and reduces nutrient wastage" Patel et al. (2024)	**Resource optimisation and waste reduction**	
"model predictive control is applied to regulate greenhouse climate conditions"	**Preventive control systems**	
"closed-loop systems allow water and nutrient recirculation"	**Resource circularity**	
"treatment strategies...must be modified to account for these differences" Kurafeeva et al. (2024) "corrective actions are applied when deviations in plant conditions are detected"	**Corrective post-detection interventions**	**Repair**
"stress response...influenced by precision irrigation" Khan et al. (2026)	**Adaptive response to plant physiological stress**	
"nutrient levels are adjusted after imbalance is identified"	**System recovery mechanisms**	

A. ZDM-Detect

ZDM-Detect refers to the systematic capacity of a CEA system to continuously observe, identify, and record deviations from desired crop and environmental states, forming the essential epistemic foundation upon which all downstream quality and resource assurance actions depend. In CEA, detection is not merely a passive monitoring function; it constitutes the first boundary condition between variability and control, between reactive management and proactive optimisation. Without a robust detection layer, the aspiration of "Zero-Defect Crops" remains structurally inaccessible. Studies across the reviewed literature converge on the idea that effective detection in CEA demands non-destructive, multi-modal sensing operating in real time. At the plant level, the development of systems capable of "non-destructive and continuous plant weight measurement" (Aji et al., 2020) marks a critical transition from episodic, manual sampling, which is inherently limited in temporal resolution, to continuous physiological surveillance. This shift is not merely technological but epistemological: it transforms the crop from an object periodically assessed to a dynamic system whose internal state is perpetually legible. Cammarisano et al. (2022) extend this logic to the spectral domain, demonstrating the "non-invasive estimation of leaf pigments," thus enabling the detection of biochemical perturbations, e.g., early-stage chlorophyll degradation or anthocyanin accumulation, that precede visible morphological symptoms by days. This temporal advantage is decisive for Zero-Defect Crops: the earlier a deviation is identified, the smaller the loss envelope. The environmental dimension of detection is equally critical. Studies consistently underscore the importance of tracking "environmental parameters such as temperature, humidity, and CO_2 continuously," reflecting the understanding that crop physiology is inseparable from its aerial and root-zone microenvironment. What remains underdeveloped in the literature, however, is the integration of plant-level and environment-level signals into unified state representations. While Kim et al. (2025) demonstrates the feasibility of "non-destructive, real-time monitoring using only low-cost fixed cameras," reducing instrumentation cost barriers, most detection architectures reviewed operate either at the plant or environmental layer, and rarely simultaneously across both with semantic fusion. A notable exception is presented by Lee et al. (2025a), who develop a "near-field remote sensing framework that fuses RGB image features" with environmental variables, including CO_2 concentration, temperature, and light conditions, to predict CO_2 uptake and respiration dynamics across five leafy vegetable species. This fusion architecture begins to close the semantic gap between plant morphological state and environmental state monitoring, demonstrating that image-derived crop features carry predictive information about CO_2 dynamics that environmental sensors alone cannot resolve. This fragmentation represents a structural limitation: in biological systems, a CO_2 anomaly may manifest as a photosynthetic efficiency change before it registers as a growth deviation; yet systems that monitor these channels in isolation will fail to recognise the causal chain. In CEA terms, this implies that realising true Zero-Defect Crop detection requires transitioning from distributed, single-variable sensing toward integrated, multi-parameter physiological-environmental state estimation, i.e., one where "morphological traits such as area, major, and minor axes" (Rahman et al., 2025) are interpreted alongside spectral and microclimate data in a coherent, real-time diagnostic frame.

B. ZDM-Predict

ZDM-Predict in CEA denotes the analytical capability to anticipate future system states, i.e., whether at the scale of crop yield trajectories, growth-stage transitions, or energy consumption dynamics, based on current and historical process data. Where detection establishes what is, prediction establishes what will be, enabling the system to shift from reactive correction to proactive intervention. In the context of *Zero-Defect Crops*, predictive capacity is not a supplementary feature but a structural necessity: biological production systems operate under inherent temporal lags, whereby the consequences of environmental mismanagement may not manifest in measurable crop outcomes for hours or days. A system that can only detect is always already behind; a system that can predict can intervene before the damage is done. The reviewed literature reveals a clear trajectory toward data-driven predictive architectures, though with significant variation in modelling philosophy and scope. At one end, Rahman et al. (2025) proposes a "scalable, data-driven method to predict yield in CEA," leveraging time-series vision models to forecast growth dynamics from morphological observations, i.e., a purely empirical approach that requires abundant observational data but remains agnostic to the underlying biological mechanisms. Patel et al. (2024) similarly demonstrates that "machine learning models are used to predict crop yield and performance" in IoT-instrumented hydroponic systems, linking sensor streams to production outcomes. These approaches excel at capturing complex, nonlinear relationships within the data envelope on which they are trained, but their generalisability across cultivars, growth stages, and environmental regimes remains a recognised limitation. Critically, a more theoretically grounded strand of the literature challenges the sufficiency of data-only models for CEA prediction. Kharraz and Szabó (2025) argue for a hybrid architecture in which a "model systematically captures environmental variability, simulates plant growth" by combining stochastic, empirical, and optimisation methods with machine learning, a convergence that substantially improves extrapolation capacity and interpretability. This tension between purely data-driven and physics-informed modelling approaches is one of the defining intellectual conflicts in the ZDM-Predict domain and carries real system implications: in a commercial CEA context, a yield prediction model that fails under novel spectral regimes or novel cultivars is a liability, not an asset. At the system level, predictive capability must extend beyond crop biology to encompass facility-level resource dynamics. Costantino et al. (2021) show that simulation models can capture both system operation and energy performance in mechanically ventilated greenhouses, highlighting the importance of integrating environmental control with energy optimisation. Extending this perspective, Choi et al. (2026) develop a two-stage MLP–lookup table model that first predicts greenhouse temperature and humidity and subsequently estimates heat pump energy consumption through manufacturer performance mapping. This architecture demonstrates robust generalisation across varying setpoint temperatures, facility locations, and control strategies, and critically, across hardware replacement scenarios where end-to-end black-box models tend to fail. Complementing this, Lee et al. (2025a) show that fusing image-based crop monitoring with environmental sensing enables accurate prediction of

CO_2 dynamics at the crop level, with scenario analyses indicating potential reductions in water demand under optimised CO_2 enrichment conditions, directly supporting resource efficiency objectives in Zero-Waste CEA. Meanwhile, Kurafeeva et al. (2024) demonstrate that even localised physical phenomena, such as predicting wind speed and direction within screened citrus systems, have practical implications for disease risk and environmental uniformity, underscoring the need for multi-scale predictive coverage. Together, these studies reveal that prediction in CEA is not a single capability but a nested hierarchy of forecasting functions, ranging from leaf-level physiological responses to facility-level energy demand, whose integration is essential to achieving both Zero-Defect Crops and Zero-Waste CEA.

C. ZDM-Prevent

ZDM-Prevent in CEA refers to the proactive, control-oriented mechanisms by which the system acts on detected states and predictive outputs to preclude deviations from materialising as crop defects or resource inefficiencies. Prevention operationalises the intelligence gathered by the Detect and Predict dimensions, transforming data into regulatory action. It is the dimension through which variability is not merely observed or anticipated but actively suppressed and, crucially, through which the *Zero-Waste CEA* objective is most directly pursued, since resource losses in CEA are predominantly a function of suboptimal environmental regulation rather than catastrophic failure events. The literature reveals prevention operating across two interlocking domains: precision input management and multi-objective environmental control. At the crop-resource interface, precision fertigation emerges as a canonical preventive mechanism. Patel et al. (2024) demonstrates that through IoT-driven automation, "real-time adjustments minimise human intervention and reduces nutrient wastage," collapsing the gap between crop demand and nutrient supply. This point is theoretically significant: conventional batch-scheduled fertigation applies inputs according to predetermined schedules, accumulating deviation between supply and actual plant need over time. In contrast, continuous-feedback fertigation maintains nutrient concentrations within tight tolerances at each growth stage, directly reducing the probability of deficiency- or toxicity-induced crop defects. Leiding et al. (2026) extend this logic to the systems level, demonstrating that "closed-loop water and nutrient cycles with waste-to-food pathways" can be constructed, transforming what are traditionally treated as terminal waste streams into recirculated productive inputs, i.e., a direct architectural expression of Zero-Waste CEA. At the facility level, multi-objective climate regulation presents a more complex challenge, as the simultaneous optimisation of temperature, humidity, CO_2, and light intensity involves trade-offs that single-variable control loops cannot resolve. Chen et al. (2022) demonstrates that "model predictive control applied to regulate greenhouse climate conditions" offers a principled approach to this challenge, enabling the system to anticipate future states and select control actions that jointly satisfy multiple performance objectives, e.g., energy consumption, thermal comfort, and crop developmental targets, rather than optimising each in isolation. Niu et al. (2024) highlight that "improvements in the light environment of greenhouse plants are essential," framing spectral management as a preventive instrument: well-designed lighting not only drives photosynthesis but prevents the photoinhibition and developmental dysregulation that arise from poorly calibrated irradiance regimes. Lee et al. (2025b) extends this argument at the spectral specificity level, demonstrating that "white LED treatments" can simultaneously sustain "favourable photochemical efficiency" in Korean thistle while maintaining yield quality across a range of spectral conditions, a finding that highlights the dual preventive function of targeted spectral selection, precluding photochemical stress without sacrificing biomass accumulation. Abdelkader et al. (2023) demonstrate that IoT-enabled automated light control systems, using low-computational-cost algorithms to regulate light intensity only when deviations from target conditions are detected, can achieve effective energy reduction across both small- and large-scale plant factory configurations, reinforcing the principle that precision input management is simultaneously a quality assurance and a resource efficiency mechanism. However, a critical tension emerges across the literature, Kurafeeva et al. (2024) warn that "treatment strategies must be modified to account for these differences" when applying control protocols developed in one environmental context to structurally distinct systems, a reminder that preventive control architectures, however sophisticated, remain locally situated and require careful validation before generalisation. This suggests that the path to universal Zero-Waste CEA is not a single control strategy but a portfolio of validated preventive mechanisms, instantiated and calibrated to the specific physiology-environment coupling of each production system.

D. ZDM-Repair

ZDM-Repair in Controlled Environment Agriculture refers to the capacity of the system to recognise, respond to, and recover from deviations that have already manifested, i.e., whether at the level of plant physiological stress, nutrient imbalance, environmental disturbance, or model prediction error, in a manner that minimises cumulative performance loss and restores the system to its intended operational state. While Detect and Predict concern the informational architecture of quality assurance, and Prevent concerns its proactive regulatory architecture, Repair constitutes the residual corrective layer that operates when preventive mechanisms are insufficient, as they inevitably will be, given the inherent stochasticity of biological production. The aspiration of Zero-Defect Crops does not presuppose the elimination of deviation; it presupposes that any deviation that does occur is contained, corrected, and learned from rapidly enough to preclude material quality loss. The reviewed literature reveals repair operating at multiple system levels, with varying degrees of sophistication and closure. At the plant physiological level, Khan et al. (2026) demonstrates that "stress response is influenced by precision irrigation" in medicinal plant cultivation under smart farming regimes, a finding that reframes irrigation not merely as a resource delivery mechanism but as a therapeutic instrument, capable of reversing or attenuating stress phenotypes once detected. Similarly, the principle that "nutrient levels are adjusted after imbalance is identified" reflects a corrective logic in which chemical monitoring triggers targeted reformulation of the fertigation solution, i.e., a reactive but highly effective intervention when detection latency is sufficiently low. What the literature does not yet adequately address is the quantification of this latency threshold: below some critical delay, correction can reverse the defect trajectory; above it, the physiological damage becomes irreversible and the crop is lost. This gap represents a significant frontier for Zero-Defect Crop research. At the system and model level, Fernández et al. (2025) introduce the concept of "feedback-based model refinement" within a digital twin architecture for hydroponic greenhouse horticulture, i.e., a repair mechanism operating not on the crop but on the predictive and control models themselves. This is a conceptually important distinction: as CEA systems accumulate operational experience across seasons,

cultivars, and environmental conditions, the capacity to update and recalibrate internal models constitutes a form of systemic learning that progressively reduces the frequency and severity of future deviations. Leiding et al. (2026) extend the repair logic to the material cycle level, proposing "closed-loop water and nutrient cycles with waste-to-food pathways" that not only prevent resource losses but actively recover value from process byproducts, i.e., a circularity mechanism that closes the metabolic loop of the agricultural system. Kurafeeva et al. (2024) further note that effective repair requires contextual specificity: correction protocols "must be modified to account for these differences" across system configurations, rejecting the notion of universal repair recipes. Taken together, these contributions suggest that Repair in CEA is not a single corrective act but a layered adaptive process, e.g., spanning physiological recovery, environmental re-regulation, model recalibration, and resource recirculation, whose coherent integration constitutes the final pillar of a truly closed-loop, Zero-Defect and Zero-Waste agricultural production system.

IV. Discussion

The application of ZDM principles to CEA represents more than a cross-domain technology transfer; it constitutes a fundamental reconceptualisation of what agricultural production systems are and what they can become. In conventional manufacturing, variability is predominantly mechanical in origin and therefore, in principle, deterministically reducible through tighter process control, better tooling, and improved materials. In CEA, by contrast, variability is biological, which is to say it is dynamic, adaptive, and only partially reducible through external regulation. Plants are not passive workpieces: they respond to environmental signals, adjust their physiology across developmental stages, and express genotype-environment interactions that no static control protocol can fully anticipate. The central theoretical claim of this study is that this distinction, rather than invalidating the ZDM analogy, actually enriches it. The increasing integration of real-time sensing, data-driven prediction, and closed-loop control in CEA systems does not make crops behave like manufactured parts; it makes the production system capable of tracking, responding to, and learning from biological variability in ways that progressively constrain its consequences. CEA, in this framing, is best understood not as a farm but as a cyber-physical production system operating on living material, and it is precisely this framing that makes ZDM a productive analytical lens.

Within this reconceptualised production context, the concepts of Zero-Defect Crops and Zero-Waste CEA emerge as two analytically distinct but operationally inseparable dimensions of a unified production paradigm. Zero-Defect Crops address the output dimension of quality, targeting the minimisation of biological variability across the crop lifecycle such that each harvested unit consistently meets predefined physiological, morphological, and safety standards. Zero-Waste CEA addresses the input-throughput dimension of efficiency, targeting the continuous optimisation and recirculation of water, nutrients, energy, and biomass flows such that the system approaches full resource closure. The interdependence between these two dimensions is not incidental but structural: a system that detects and corrects plant stress in real time necessarily consumes fewer corrective inputs than one that allows deviations to accumulate; conversely, a system that maintains precise resource delivery reduces the probability of the physiological stress that generates crop defects in the first place. Together, they reflect a paradigm shift from yield-oriented agriculture, in which the primary performance metric is total biomass produced per unit area, toward system-optimised agriculture, in which the performance envelope encompasses output quality, resource utilisation efficiency, and adaptive system resilience simultaneously.

From an implementation perspective, the framework can also be read as a capability architecture for CEA system design and assessment. Detect corresponds to the sensing and data acquisition layer; Predict corresponds to the modelling and forecasting layer; Prevent corresponds to the closed-loop optimisation and control layer; and Repair corresponds to the corrective intervention and learning layer. Framed in this way, the ZDM-derived structure can support the staged development of CEA facilities, enabling operators and technology providers to identify capability gaps, prioritise investments, and evaluate whether an apparently "smart" agricultural system is in fact capable of operating as an integrated quality and resource assurance system.

Beneath the visible technological layer of CEA, however, lies a set of structural challenges that the reviewed literature collectively surfaces but rarely addresses in integrated terms. Much of the discourse on smart agriculture and precision CEA tends to foreground individual capabilities, e.g., computer vision for growth monitoring, machine learning for yield prediction, or model predictive control for climate regulation, as discrete solutions to discrete problems. Yet the deeper challenge is not the availability of these capabilities in isolation but their coherent integration into a system that can close the loop between sensing, reasoning, and action across the full biological and operational complexity of a living production environment. This challenge operates at multiple layers simultaneously. At the data layer, signals from heterogeneous sensors, e.g., spectral cameras, CO_2 monitors, weighing platforms, and flow meters, must be semantically fused into a coherent state representation rather than monitored in parallel silos. At the model layer, the tension between purely empirical, data-driven models and physics-informed hybrid architectures reflects a deeper uncertainty about which aspects of crop-environment interaction are sufficiently stable to be modelled mechanistically and which require continuous empirical recalibration. At the control layer, the multi-objective nature of CEA regulation means that optimising for one performance dimension, i.e., maximising photosynthetic efficiency through elevated CO_2, may conflict with another, i.e., minimising energy consumption. Recognising these layered integration challenges is not an argument against the ZDM framework; it is an argument for why the framework is needed. A system that lacks explicit Detect, Predict, Prevent, and Repair functions will manage these tensions reactively and inconsistently; a system that instantiates all four functions coherently will manage them systematically.

The Detect-Prevent coupling constitutes the most direct operational expression of Zero-Defect Crop logic in CEA. Real-time sensing of plant physiological state, including the continuous tracking of morphological indicators such as "morphological traits such as area, major, and minor axes" (Rahman et al., 2025) and biochemical signals such as "chlorophyll fluorescence as a valuable tool for monitoring light use efficiency" (Van Iersel et al., 2016), generates a high-resolution dynamic picture of crop condition that can trigger immediate environmental adjustments before stress expressions consolidate into irreversible quality losses. The interpretive significance of this coupling lies in the compression of the deviation-response cycle: in a system without

real-time detection, the interval between the onset of a physiological deviation and its discovery through periodic inspection may span hours or days, during which the deviation propagates. In a system with continuous sensing, this interval is reduced to minutes or seconds, and the corresponding window for effective preventive intervention remains open. From a Zero-Waste CEA perspective, the same coupling delivers a secondary benefit: targeted, demand-driven resource delivery, of the kind demonstrated by Patel et al. (2024) in which "real-time adjustments minimise human intervention and reduces nutrient wastage," eliminates the systematic oversupply that characterises schedule-based input management, thereby reducing both the cost and the environmental burden of production. This resource efficiency dimension extends to atmospheric management as well. Lee et al. (2025a) shows that continuous image-sensor fusion monitoring of CO_2 dynamics can inform enrichment strategies that reduce modelled water demand by approximately 11%, a finding that illustrates how detection-informed environmental adjustment generates cascading resource gains across multiple input streams simultaneously. The Detect-Prevent linkage, in this light, is not merely a quality assurance mechanism but a resource governance mechanism operating in real time.

The Predict-Prevent coupling extends this logic into the temporal domain, enabling the system to act on anticipated states rather than observed ones. Where the Detect-Prevent loop is reactive at the timescale of minutes, the Predict-Prevent loop is proactive at the timescale of days and growing seasons. Rahman et al. (2025) demonstrates that a "vision-based framework to forecast plant growth dynamics" can generate reliable trajectories of crop development that enable preventive adjustment of environmental setpoints, harvest scheduling, and resource provisioning long before any deviation is detectable in sensor data. Kharraz and Szabó (2025), operating at the model architecture level, show that a "hybrid model systematically captures environmental variability, simulates plant growth" through the integration of stochastic and mechanistic components, enabling the system to anticipate not just the most likely future state but the range of plausible futures given current environmental uncertainty. The implication for Zero-Waste CEA is particularly significant: a system that can predict energy demand trajectories, as Costantino et al. (2021) demonstrate through a "model simulates the operation of the systems and the energy performance," can schedule resource-intensive operations, i.e., heating, artificial lighting, CO_2 enrichment, to coincide with periods of low energy cost or high renewable availability, thereby reducing total energy expenditure without compromising crop developmental targets. The Predict-Prevent coupling, therefore, is the mechanism through which CEA systems transition from responsive to anticipatory, and from locally optimised to globally efficient.

The Detect-Repair coupling serves a distinct but complementary function: it ensures that deviations which escape the preventive layer are identified and corrected before they propagate to irreversible crop damage or systemic resource loss. The theoretical importance of this coupling lies in its acknowledgment that no preventive architecture, however sophisticated, can eliminate all deviations in a biological production system. Biological stochasticity, model uncertainty, sensor failure, and exogenous disturbances, e.g., pathogen incursion or supply chain disruption, will periodically breach the preventive envelope. What determines whether such breaches result in significant crop or resource losses is the speed and precision of the corrective response. Khan et al. (2026) demonstrates that "stress response is influenced by precision irrigation," illustrating that targeted corrective input can reverse physiological stress trajectories once detected, provided the intervention is applied within a biologically meaningful window. Fernández et al. (2025) introduce a higher-order form of repair through "feedback-based model refinement," in which detection of prediction errors triggers updates to the system's internal models rather than, or in addition to, the crop environment itself. This model-level repair is conceptually significant because it represents a form of systemic learning: the system does not merely correct the immediate deviation but improves its capacity to prevent analogous deviations in the future. From a Zero-Defect Crops perspective, the Detect-Repair coupling thus closes the quality loop not only in the operational cycle but across production cycles, progressively narrowing the variance envelope through accumulated corrective intelligence.

Taken together, these three couplings, i.e., Detect-Prevent, Predict-Prevent, and Detect-Repair, constitute the operational architecture of a closed-loop, self-optimising CEA system. The functional logic is not sequential but recursive: detection feeds both prevention and repair; prediction amplifies the reach and precision of prevention; repair generates feedback that refines detection sensitivity and predictive accuracy; and the entire system progressively converges toward a tighter correspondence between intended and actual crop and resource states across successive production cycles. This architecture is structurally consonant with the principles of Industry 4.0 and its evolution toward Industry 5.0, in which cyber-physical systems are not merely data-generating or data-consuming but genuinely self-aware and self-correcting. The difference in CEA is that the cyber-physical system is coupled to a living organism rather than a mechanical process, which introduces irreducible biological dynamics that no digital architecture can fully internalise. The productive tension between the determinism of the control system and the stochasticity of the biological system is, in this sense, the defining engineering challenge of advanced CEA, and the ZDM framework provides the conceptual vocabulary to address it systematically rather than ad hoc.

The broader implications of this work extend beyond the technical domain of CEA system design. The framing of agricultural production as a defect- and waste-minimising cyber-physical system connects to some of the most pressing imperatives of contemporary food system governance: the need to produce more food with fewer resources, to decouple agricultural productivity from environmental degradation, and to build resilience into food supply chains increasingly vulnerable to climate volatility. The convergence of Zero-Defect Crops and Zero-Waste CEA within a unified ZDM-derived framework suggests that the apparent trade-off between food quality and resource efficiency is, in a well-designed CEA system, a false dichotomy: the same real-time data infrastructure that enables quality assurance also enables resource optimisation, because both objectives are ultimately expressions of the same underlying imperative, i.e., to minimise the gap between the biological potential of the crop and the actual performance of the production system. In this light, the future of controlled environment food production may be understood as a form of biological manufacturing, not in the reductive sense of treating organisms as machines, but in the precise sense that production variability is systematically characterised, modelled, controlled, and learned from with the same rigour that advanced manufacturing applies to mechanical tolerance management. Whether that aspiration is

realisable at the scale and cost required for mainstream food production remains an open and consequential question, one that this framework is designed to help answer.

This study should be interpreted in light of several limitations. First, the paper is conceptual and theory-building in nature; it does not empirically validate the proposed framework in a real CEA facility. Second, the reviewed literature is heterogeneous in scope, ranging from plant monitoring and environmental control to predictive modelling and smart agriculture architectures, which means that the resulting framework necessarily synthesises evidence across partially different operational contexts. Third, while the ZDM dimensions provide a powerful integrative structure, the transfer from industrial manufacturing to biological production is not exact, because plant systems remain adaptive, developmental, and partly irreducible to deterministic control logic. These limitations do not weaken the value of the framework, but they define the conditions under which it should be interpreted.

Future research should therefore move in three directions. First, the framework should be empirically tested in commercial or pilot CEA facilities across different crops and production regimes. Second, quantitative maturity indicators should be developed for each ZDM dimension in order to assess how close a facility is to operating as a closed-loop zero-defect and zero-waste system. Third, future work should explore how digital twins, physics-informed AI, and multi-objective optimisation can operationalise the framework in real-time control environments where crop quality, yield stability, energy use, water recirculation, and nutrient efficiency must be managed simultaneously.

V. Conclusions

This study demonstrates that the ZDM paradigm can be systematically extended to CEA, establishing a theoretically grounded and operationally coherent framework for defect-free, resource-efficient biological production. By applying the Gioia methodology to a systematic literature mapping of CEA and smart agriculture research and using the four established ZDM aggregate dimensions of Detect, Predict, Prevent, and Repair, structured in alignment with established ZDM theory, the study identifies and translates them into the novel constructs of Zero-Defect Crops and Zero-Waste CEA. This constitutes the first systematic application of ZDM principles to a biological production context and advances a conceptual framework that positions CEA not as a form of managed agriculture but as a cyber-physical production system capable of operating with industrial-grade quality and resource assurance.

This work contributes to the theoretical literature in two respects. First, it extends the conceptual boundary of ZDM from mechanical systems, where variability is predominantly deterministic, to biological systems, where variability is dynamic, adaptive, and genotype-environment dependent, demonstrating that the ZDM logic remains analytically productive despite this fundamental ontological difference. Second, it advances the Gioia methodology by introducing a hybrid inductive-deductive design in which aggregate dimensions are not derived purely from the literature but anchored a priori in a validated theoretical framework. This disciplined abduction enables structured cross-domain theory building and offers a replicable methodological model for researchers seeking to translate industrial production paradigms into emerging biological, ecological, or service-oriented domains where fragmented empirical literatures resist conventional systematic review.

From a practical perspective, the framework proposed in this study offers CEA system designers, facility operators, and agricultural technology developers a structured vocabulary for diagnosing gaps in their production systems and prioritising capability investments. A CEA facility that monitors crop growth but lacks predictive models cannot anticipate deviations before they manifest; one that predicts yield trajectories, but lacks closed-loop control cannot translate foresight into prevention. The Detect-Predict-Prevent-Repair architecture makes these functional dependencies explicit, enabling operators to evaluate their systems not as collections of individual technologies but as integrated quality assurance systems with identifiable failure modes. Beyond the facility level, the Zero-Waste CEA construct, with its emphasis on full resource circularity across water, nutrient, energy, and biomass flows, connects directly to sustainability objectives in agri-food policy, offering a production-systems rationale for regulatory incentives targeting resource efficiency and environmental decoupling in urban and peri-urban food production.

These findings suggest that the trajectory of CEA development points toward what may be characterised as biological manufacturing, i.e., a production paradigm in which crop variability is not merely tolerated or reactively corrected but systematically characterised, modelled, controlled, and learned from across production cycles. Realising this trajectory at operational scale will require advances in multi-modal sensor fusion, hybrid physics-informed predictive modelling, and empirical validation of the ZDM-CEA framework across diverse crops, facility configurations, and production regimes. These represent the natural next steps for a research programme that this study is designed to initiate.

Acknowledgment

Victor Azamfirei-Ionita acknowledge Fundação para a Ciência e a Tecnologia (FCT-MCTES) for its financial support via the project UID/00667:Unidade de Investigação e Desenvolvimento em Engenharia Mecânica e Industrial. The presented work was partially supported by the projects CIRCMAN5.0, EU H2020 projects under grant agreements No 10117833.

ORCID

Victor Azamfirei Ionita: https://orcid.org/0000-0001-5159-5276

Foivos Psarommatis: https://orcid.org/0000-0002-2731-8727

References

Aji, G. K., Hatou, K., & Morimoto, T. (2020, December). Optimum strategies for the control of root-zone temperature to promote early-stage growth of chili pepper in soilless culture using an intelligent approach. In III Asian Horticultural Congress-AHC2020 1312 (pp. 173-180). https://doi.org/10.17660/ActaHortic.2021.1312.26

Cahyo, W. N. (2021, February). Finding novelty of research with Systematic Literature Mapping (SLM). In Journal of Physics: Conference Series (Vol. 1764, No. 1, p. 012186). IOP Publishing. https://doi.org/10.1088/1742-6596/1764/1/012186

Cammarisano, L., Graefe, J., & Körner, O. (2022). Using leaf spectroscopy and pigment estimation to monitor indoor grown lettuce dynamic response to spectral light intensity. Frontiers in Plant Science, 13, 1044976. https://doi.org/10.3389/fpls.2022.1044976

Chen, W. H., Mattson, N. S., & You, F. (2022). Intelligent control and energy optimization in controlled environment agriculture via nonlinear model predictive control of semi-closed greenhouse. Applied Energy, 320, 119334. https://doi.org/10.1016/j.apenergy.2022.119334

Choi, E. J., Lee, D., Lee, S. M., & Lim, S. (2026). Two-stage MLP-lookup table model for predicting heat pump power in greenhouses. Energy and Buildings, 117027. https://doi.org/10.1016/j.enbuild.2026.117027

Costantino, A., Comba, L., Sicardi, G., Bariani, M., & Fabrizio, E. (2021). Energy performance and climate control in mechanically ventilated greenhouses: A dynamic modelling-based assessment and investigation. Applied Energy, 288, 116583. https://doi.org/10.1016/j.apenergy.2021.116583

Dutta, M., Gupta, D., Tharewal, S., Goyal, D., Sandhu, J. K., Kaur, M., ... & Alanazi, J. M. (2025). Internet of things-based smart precision farming in soilless agriculture: Opportunities and challenges for global food security. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3540317

Fernández, R., Navas, E., Rodríguez-Nieto, D., Rodríguez-González, A. A., & Emmi, L. (2025). DigiHortiRobot: an AI-driven digital twin architecture for hydroponic greenhouse horticulture with dual-arm robotic automation. Future Internet, 17(8), 347. https://doi.org/10.3390/fi17080347

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. Organizational research methods, 16(1), 15-31. https://doi.org/10.1177/1094428112452151

Hadj Abdelkader, O., Bouzebiba, H., Pena, D., & Aguiar, A. P. (2023). Energy-efficient IoT-based light control system in smart indoor agriculture. Sensors, 23(18), 7670. https://doi.org/10.3390/s23187670

Khan, S., Pathania, N., Kumar, P., Kumar, R., Kumar, J., Kumar, N., & Sharma, A. (2026). Smart farming approaches in medicinal plant cultivation: a review of techniques, benefits, and sustainability. Planta, 263(4), 92. https://doi.org/10.1007/s00425-026-04960-w

Kharraz, N., & Szabó, I. (2025). Hybrid Plant Growth: Integrating Stochastic, Empirical, and Optimization Models with Machine Learning for Controlled Environment Agriculture. Agronomy, 15(1), 189. https://doi.org/10.3390/agronomy15010189

Kim, J. S. G., Shin, S. H., & Chung, S. (2025). Instance-level phenotype-based growth stage classification of basil in multi-plant environments. Frontiers in Plant Science, 16, 1707985. https://doi.org/10.3389/fpls.2025.1707985

Kurafeeva, L., Wolski, R., Krintz, C., & Smyth, T. (2024). Airflow modeling for citrus under protective screens. Sensors, 24(19), 6200. https://doi.org/10.3390/s24196200

Lee, J. H., Kim, E. A., Sunwoo, Y., Cha, E. B., Joo, S., Lee, S., ... & Nam, S. Y. (2025b). Light Spectrum Strategies to Optimize Yield and Physiological Performance of Korean Thistle (Cirsium setidens (Dunn) Nakai) in a Closed-type Plant Production System. 인간식물환경학회지 (JPPE), 28(6), 799-823. https://doi.org/10.11628/ksppe.2025.28.6.799

Lee, S., Kim, B., Cheon, S. G., & Lee, J. W. (2025a). Proximal Monitoring of CO2 Dynamics in Indoor Smart Farming: A Deep Learning and Image-Sensor Fusion Approach. Sustainability, 17(23), 10838. https://doi.org/10.3390/su172310838

Leiding, B., Paulus, S., & Lehmann, D. J. (2026) The Future of Resilient Food Production-Current Challenges and Future Opportunities. Frontiers in Sustainable Food Systems, 10, 1673667. https://doi.org/10.3389/fsufs.2026.1673667

Niu, Y., Han, Y., Li, Y., Zhang, M., & Li, H. (2024). Low–carbon regulation method for greenhouse light environment based on multi–objective optimization. Expert Systems with Applications, 252, 124228. https://doi.org/10.1016/j.eswa.2024.124228

Patel, J., Bhatt, T., & Joshi, A. (2024, August). IoT-driven enhancement of hydroponic fertilization efficiency through machine learning: a data-centric strategy. In 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI) (pp. 298-302). IEEE. https://doi.org/10.1109/ICoICI62503.2024.10696575

Psarommatis, F., May, G., Dreyfus, P. A., & Kiritsis, D. (2020). Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research. International journal of production research, 58(1), 1-17. https://doi.org/10.1080/00207543.2019.1605228

Rahman, M. H., Ayipio, E., Lukwesa, D., Zheng, J., Wells, D. E., Syed, H. H., & Rehman, T. U. (2025). Forecasting growth dynamics of hydroponic kale in controlled environment agriculture through vision-based phenotyping and time-series modeling. Smart Agricultural Technology, 11, 101039. https://doi.org/10.1016/j.atech.2025.101039

Van Iersel, M. W., Mattos, E., Weaver, G., Ferrarezi, R. S., Martin, M. T., & Haidekker, M. (2016, May). Using chlorophyll fluorescence to control lighting in controlled environment agriculture. In VIII International Symposium on Light in Horticulture 1134 (pp. 427-434). https://doi.org/10.17660/ActaHortic.2016.1134.54

Wang, Z., Huang, H., Shi, M., Xiong, Y., Wang, J., Wang, Y., & Zou, J. (2025). Optimized spectral and spatial design of high-uniformity and energy-efficient LED lighting for Italian lettuce cultivation in miniature plant factories. Horticulturae, 11(7), 779. https://doi.org/10.3390/horticulturae11070779

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Acceptance sampling by variables for double specification limits: save space, and compute on the Web

Miguel Casquilho [1,3], Francisco Rolo [2], Pedro P. Santos [2] and Rui Galhano [1,3]

1 Department of Chemical Engineering, Instituto Superior T´ecnico (IST), University of Lisbon (ULisboa), Lisbon,, Portugal; IST, Ave. Rovisco Pais, IST; 1049-001 Lisboa, Portugal; MC[1], RG (mcasquilho, rui.galhano)@tecnico.ulisboa.pt
2 Department of Computer Science and Engineering, Instituto Superior T´ecnico (IST), University of Lisbon (ULisboa), Lisbon, Portugal; FR, PPS (franciscorolo, pedro.filipe.santos)@tecnico.ulisboa.pt
3 CERENA, *Centro de Recursos Naturais e Ambiente* (Centre for Natural Resources and the Environment), IST, ULisboa

***Abstract*— In this paper, we discuss computations in Acceptance Sampling "by variables", the classical complement to Statistical Process Control. We focus on the case of double (lower and upper) specification limits (Gaussian variable), and relate the cases of single and double specification limits. After briefly reviewing the topic, we expose a detail in its core mathematical basis, relating the quality index to the estimated percent defective, in contrast to the verbosity in literature, and provide a web page for the direct computation. There, the interested reader, needing no additional software or tables, can verify our computations in a clear, trustworthy manner, rare in the literature. This methodology also points to an easier collaboration between Academia and Industry.**

Keywords – Quality Control; acceptance sampling by variables; Web computing; "Form 2"; international standards

I. 1 Background and purpose

Quality, if we couple it to Statistics, has been around already since one century and is everywhere, namely in manufacture and in business in general. A sketch of the history of Quality is in Fig. 1, noticing the leap in the "Early 20.th century", from the American Society for Quality (ASQ) website[2], where pioneers, such as Shewhart, Deming, Juran, Ishikawa, are naturally mentioned.

In this paper, in a simplified manner, we separate what occurs inside an organization and what happens across its frontiers. In terms familiar to us, Quality partisans, these would be Statistical Process Control (SPC) and Acceptance Sampling (AS), respectively. We address the latter, to (*i*) recollect some of its statistical basis, (*ii*) bring it to the open light of execution on a constructed web page, shortening the scientific path, and (*iii*) criticizing simultaneous verbosity and gaps in essential support documents.

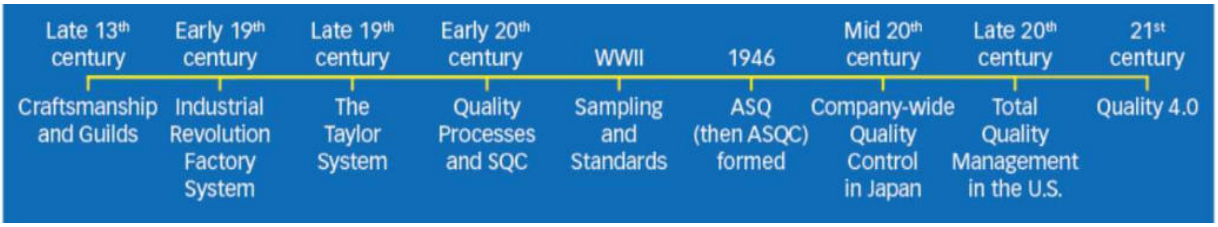


Fig. 1. ASQ: "The History of Quality".

Acceptance sampling by variables to inspect a lot of similar items (detailed in III.A) is, briefly:

- From the lot, draw a sample and calculate its average, $\overline{X}$, and standard deviation, s.
- Is $\overline{X}$ "safely" inside (*L*, *U*), i.e., at a good distance of them, in terms of s ?
- If Yes, Accept the lot; if No, Reject the lot.

We also review basic aspects of Quality, often diluted in management concepts, as once remarked by (Gunther 1998). The fact is that the international standards underlying our subject keep being updated and maintained (e.g., (ISO/TC 69/SC 5 2022b)). In the areas of Engineering, these authors' field of interest, such as Chemical or Computer Science Engineering, Acceptance Sampling becomes more purposeful when, simultaneously, the continuous variable inspected is subject to the two specification limits, lower and upper[3]. In the Quality terminology, this is "AS by variables" under "Form 2" (ISO/TC 69/SC 5

[1] Corresponding author M. Casquilho: mcasquilho@tecnico.ulisboa.pt
[2] URL: https://asq.org/quality-resources/history-of-quality
[3] Not to be confused with the lower and upper control limits, LCL and UCL, in SPC.

2022a), whereas Form 1 is meant for single specification limits. This becomes computationally accessible, as will be seen, and is advantageous, namely, to non-specialized decision makers.

In the following parts: Section II, "Approach to bibliographical reproach", comments some unsatisfactory literature features; Section III, "Inspection procedure and mathematical transformation", describes the sampling procedure according to the applicable standards and the said transformation; Section IV, "Results and discussion", shows and explains the web-based functionality, and settles our results; and Section v, "Conclusions", sums up the usefulness of the study.

II. Approach to bibliographical reproach

As a first thought, we note that, despite the presence of Informatics[4] in most domains (Engineering, Medicine, Linguistics, etc.), we (these non-Informatics Engineering authors) regret the often-weak connection from the informatic engineers, leading to poor results in the said domains. Second thought, in the scientific literature, is the lack of links to web pages for the verification of results by the reader, a fact about which we are not neutral. We always provide them, but found no literature examples to cite, apart from our own, such as (Casquilho 2013), (Carolino et al. 2016), (Barros and Casquilho 2019), (Casquilho and Buescu 2022), (Casquilho et al. 2025), (Casquilho and Pacheco 2023), (Casquilho 2025), as verifiable, e.g., in (Kuntsche et al. 2011)[5]. One reason for this absence may be the editors' worry (reasonably, as seen in next paragraph) that the targets disappear, which leads to the broader issue of the permanence of the publications themselves, nowadays as 'pdf' documents on the Web, no longer in paper.

A remarkable case of web computing (not typical articles) is Ponce's (Ponce 1999)[6] website (Hydraulics) with hundreds of dynamic problems. Another example of web computing, also academic, is Arsham's (Arsham 1994) website, a large repository on Operations Research and Statistics. As concrete article, see, e.g., (Kuntsche et al. 2011), on Chemical Engineering, reporting a website for process simulation for which one of the two links is "not found" (well, since 2011) and showing no clickable illustration. (We resort to open the 'pdf's with the free version of (PDF Xchange 2026), usually able to detect links.)

These authors will be thankful for any examples of (open access or not) technological literature articles that include links for the verification by the interested readers. We could not find them even with the help from several AI (artificial intelligence) engines.

III. Inspection procedure and mathematical transformation

We now detail the inspection procedure for practical application, as prescribed in the standard, and show the mathematical transformation of the acceptance constant, *k*, into the maximum estimated percent defective, *M*.

A. Inspection procedure

The use of AS by variables, as defined in the applicable international standards, assumes that the material to be inspected is constituted by a series, the "lot", of individual, measurable, similar parts (such as screws, resistors, bags of cereal, bottles of chemicals with prescribed composition, etc.). The inspection of incoherent material, the so-called "bulk sampling" (ISO/TC 69/SC 5 2022a), such as powders, slurry, liquids, is not in our scope.

The inspection procedure observes the following steps, beginning with the establishment of the size of the random sample to be drawn from the whole material, i.e., the lot, of, say, size $N = 100$ (to match the example used further below):

The inspection procedure begins with the establishment of the size of the random sample to be drawn from the whole material, i.e., the lot, of, say, size N = 100 (to match the example used further below):

a) Identify the size of the lot, *N* , to be "sentenced" (accept or reject).

b) Stipulate the two specification limits (the "specs"), lower spec, *L*, and upper spec, *U*.

c) Stipulate the AQL (Acceptable Quality Limit) for each spec, AQL_L and AQL_U (percentage).

d) Find the size, *n*, of the sample to be randomly drawn from the lot.

e) Follow the decision procedure.

In the above list, step 'd)' (finding the adequate sample size) is implemented by the sampling plan, given in a table in the standard as follows:

a) Supply the lot size, *N*.

b) Select the "sampling level", which we admit Normal, in the present context.

c) Obtain the sample size letter from 'Table A-2', in Fig. 2 (assuming the neutral "inspection level" IV). The letter (F) will become $n = 10$ in 'Table B-3', in Fig. 3.

We illustrate the procedure with the example in (Schilling and Neubauer 2009), 2.nd ed. 288.th page), which reappears in (Schilling and Neubauer 2017), 3.rd ed. (p. 318), itself taken from the standard. For a lot of $N = 100$ (from the standard, 'Table B-3', in Fig. 3), it is $n = 10$ items, with values such as the following:

[4] Or "Information Science", "Computer Science" (in the USA), etc..

[5] No visible link: https://mosaic-modeling.de/?page id=5490 , accessed 2026-Feb-01.

[6] V. M. Ponce, OnlineCalc, San Diego State University, https://ponce.sdsu.edu/online calc.php, accessed 2026-Feb-01.

$\mathbf{x}$ = 643 651 619 627 668 670 673 642 638 650

These values are supposed to be a sample of measured electrical resistance, in ohm, with stipulated specification limits of (650 ± 30) Ω and AQL's

$$L = 620,\ AQL_L = 1\%,\ U = 680,\ AQL_U = 2.5\% .$$

The numerical procedure to take a decision ("sentence" the lot) is detailed in Table 1. A lot is accepted if (only if) the three criteria are Ac ("accept"); and a lot is rejected if any criterion is Re ("reject"). So, in this example, the lot would be rejected.

Contrary to Form 1, which just concludes Ac or Re, Form 2 yields a real, more intuitive notion of how far (above or below, respectively) the inferred quality is from the limiting percent defective. (The words percent and defective are, respectively, an adjective and a noun, i.e., a substantive or nominalized adjective.) So, Form 2 is generally recommended for communication with the deciders.

In this study, one of our main points is that 'Table B-3', in Fig. 3, is a mathematical transformation (and a simple one) from 'Table B-1', in Fig. 4, as will be seen.

MIL-STD-414 Table A-2—Sample Size Code Letters[a]

	Inspection Levels				
Lot Size	I	II	III	IV	V
3–8	B	B	B	B	C
9–15	B	B	B	B	D
16–25	B	B	B	C	E
26–40	B	B	B	D	F
41–65	B	B	C	E	G
66–110	B	B	D	F	H
111–180	B	C	E	G	I
181–300	B	D	F	H	J
301–500	C	E	G	I	K
501–800	D	F	H	J	L
801–1,300	E	G	I	K	L
1,301–3,200	F	H	J	L	M
3,201–8,000	G	I	L	M	N
8,001–22,000	H	J	M	N	O
22,001–110,000	I	K	N	O	P
110,001–550,000	I	K	O	P	Q
550,001 and over	I	K	P	Q	Q

Source: United States Department of Defense, *Military Standard, Sampling Procedures and Tables for Inspection by Variables for Percent Defective*, MIL-STD-414, U.S. Government Printing Office, Washington, DC, 1957, p. 4.

[a] Sample size code letters given in body of table are applicable when the indicated inspection levels are to be used.

Fig. 2. Table A-2 (MIL-STD-414), giving the sample size, *n*, from the lot size, *N*.

B. Transformation of k into M

Let us return to the example problem in Table 1 and take AQL_U = 2.5%. In this case, for the said n = 10, Table B-1' (in Fig. 4) gives k = 1.41. (In this article, we are not analyzing this table, also called "the k s-table", so its underlying consumer and producer's risks are assumed.)

We now cite the books by (Sarkadi and Vincze 1974) and by (DeVor et al. 2007). The first book (415 pp), after 229 pages of mathematical theory, presents its Table 27 (pp 326–329), "Table for estimating the lot fraction defective" (from Table C-5 of MIL-STD-414)". Thus, surprisingly: a heavily theoretical book reproduces a table from a standard where that table was calculated from theory; and the discourse suddenly switches to the "range method", the alternative measure of variability. (The expected table would be Table B-5, not C-5.) No mention was found there to the incomplete beta distribution.

The second book (809 pp) just reads (p 392) that "the objective of acceptance sampling is to arrive at a decision of either accepting or rejecting a fixed and generally large lot of products in the basis of a small sample of observations of that lot", so no mention to the sought underlying theory.

Our last attempt was the fundamental book by (Duncan 1986), 1145 pp, certainly upstream from the two previously cited. Notice that Duncan co-authored another book in 1944 and probably made the original edition in 1952, according to the vast Japanese international repository CiNii [7]. After describing Form 2 (which he also calls Procedure 2), the conversion of Q into

[7] CiNii Research, NII (National Institute of Informatics, Tokyo, Japan), Acheson Duncan (search "1952"), https://cir.nii.ac.jp/books?q=Acheson+Duncan, accessed 2026-02-28.

$\hat{p}$ (our *p*) is passed to a "chart" (a nomogram), a crafty way of saving many pages in a table but mentioning nothing of the theory. We resort to our own longtime personal knowledge (Alves 1970), theoretically corroborated now by AI.

$$M = \mathrm{B}(x;a,b) = \\ = \mathrm{B}\left(\frac{1}{2} - \frac{1}{2}k\frac{\sqrt{n}}{n-1}; \frac{n}{2} - 1; \frac{n}{2} - 1\right) \quad (1)$$

where: **B**(*x*; *a*. *b*) is the cumulative incomplete beta function (cumulative beta probability function) of equal parameters *a* and *b*; *n* is the sample size; and *k* is the acceptance constant for which *M* is the correspondent maximum allowable fraction defective.

If we use Excel to calculate (1), we get

M = BETADIST(x, α, β, cumulative) = BETADIST(0.2522882, 4, 4, True) = 7.269% (2)

which does reproduce the value of 7.27% in 'Table B-3' (Fig. 3) for the pair (*n*, AQL) = (10, 2.5%).

If we use Python (Python.scipy.special.betainc, n.d.),

```
import numpy as np
import scipy.special as sc
n = 10; k = 1.41
x = 0.5 - 0.5 * k * np.sqrt(n) / (n - 1); a = 0.5 * n - 1
print ('x, a,', x, a)
print ('M,', sc.betainc(a, a, x))
```

we also get *M*, 0.072687.

The above segment derogates the fact that a fundamental work, such as (Schilling and Neubauer 2009) (2.nd ed.) and (Schilling and Neubauer 2017) (idem, 3.rd ed.), spends, respectively, 12 pages (626.th–637.th [8]) and 14 pages (pp 735–748) giving the "estimation of the lot percent defective" from values of Q_L or Q_U and sample sizes. These pages are necessarily limited to a range of quality indices, *Q*, and several sample sizes, *n*, whereas the simple formulas above might replace or complement those long tables (Lieberman and Resnikoff 1957).

[8] This numbering is from a 'pdf' reader, since in this case our copy (pdf) has no page numbers.

MIL-STD-414 Table B-3—Master Table for Normal and Tightened Inspection for Plans Based on Variability Unknown: Standard Deviation Method (Double Specification Limit and Form 2, Single-Specification Limit)

Sample Size Code Letter	Sample Size	Acceptable Quality Levels (Normal Inspection)													
		.040	.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00
		M	M	M	M	M	M	M	M	M	M	M	M	M	M
B	3	⇩	⇩	⇩	⇩	⇩	⇩	⇩	⇩	⇩	7.59	18.86	26.94	33.69	40.47
C	4	⇩	⇩	⇩	⇩	⇩	⇩	⇩	1.53	5.50	10.92	16.45	22.86	29.45	36.90
D	5	⇩	⇩	⇩	⇩	⇩	⇩	1.33	3.32	5.83	9.80	14.39	20.19	26.56	33.99
E	7	⇩	⇩	⇩	⇩	0.422	1.06	2.14	3.55	5.35	8.40	12.20	17.35	23.29	30.50
F	10	⇩	⇩	⇩	0.349	0.716	1.30	2.17	3.26	4.77	7.29	10.54	15.17	20.74	27.57
G	15	0.099	0.186	0.312	0.503	0.818	1.31	2.11	3.05	4.31	6.56	9.46	13.71	18.94	25.61
H	20	0.135	0.228	0.365	0.544	0.846	1.29	2.05	2.95	4.09	6.17	8.92	12.99	18.03	24.53
I	25	0.155	0.250	0.380	0.551	0.877	1.29	2.00	2.86	3.97	5.97	8.63	12.57	17.51	23.97
J	30	0.179	0.280	0.413	0.581	0.879	1.29	1.98	2.83	3.91	5.86	8.47	12.36	17.24	23.58
K	35	0.170	0.264	0.388	0.535	0.847	1.23	1.87	2.68	3.70	5.57	8.10	11.87	16.65	22.91
L	40	0.179	0.275	0.401	0.566	0.873	1.26	1.88	2.71	3.72	5.58	8.09	11.85	16.61	22.86
M	50	0.163	0.250	0.363	0.503	0.789	1.17	1.71	2.49	3.45	5.20	7.61	11.23	15.87	22.00
N	75	0.147	0.228	0.330	0.467	0.720	1.07	1.60	2.29	3.20	4.87	7.15	10.63	15.13	21.11
O	100	0.145	0.220	0.317	0.447	0.689	1.02	1.53	2.20	3.07	4.69	6.91	10.32	14.75	20.66
P	150	0.134	0.203	0.293	0.413	0.638	0.949	1.43	2.05	2.89	4.43	6.57	9.88	14.20	20.02
Q	200	0.135	0.204	0.294	0.414	0.637	0.945	1.42	2.04	2.87	4.40	6.53	9.81	14.12	19.92
		.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00	
		Acceptability Quality Levels (tightened inspection)													

Source: United States Department of Defense, *Military Standard, Sampling Procedures and Tables for Inspection by Variables for Percent Defective*, MIL-STD-414, U.S. Government Printing Office, Washington, DC, 1957, p. 45.

All AQL and table values are in percent defective.

⇩ Use first sampling plan below arrow, that is, both sample size as well as *M* value. When sample size equals or exceeds lot size, every item in the lot must be inspected.

Fig. 3. Table B-3, giving *M* , maximum allowable percent defective..

Table 1: Decision procedure to accept (Ac) or reject (Re)

Item	Value	Source
$\bar{X}, S$	647.0, 17.2	—
$Q_L = \frac{\bar{X} - L}{S}$	1.57	—
$Q_U = \frac{U - \bar{X}}{S}$	1.92	—
$Q_L \Rightarrow p_L$	4.92%	(From Table B-5)
$Q_U \Rightarrow p_U$	1.62%	(From Table B-5)
M_L	3.26%	(From Table B-3)
M_U	7.29%	(From Table B-3)
$p_L < M_L$?	No (Re)	—
$p_U < M_U$?	Yes (Ac)	—
$p_L + p_U$	6.54%	—
$p_L + p_U < \max(M_L, M_U)$?	Yes	(See below)

MID-STD-414 Table B-1—Master Table for Normal and Tightened Inspection for Plans Based on Variability Unknown: Standard Deviation Method (Single-Specification Limit, Form 1)

Sample Size Code Letter	Sample Size	Acceptable Quality Levels (Normal Inspection)													
		.04	.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00
		k	k	k	k	k	k	k	k	k	k	k	k	k	k
B	3	⇩	⇩	⇩	⇩	⇩	⇩	⇩	⇩	⇩	1.12	.958	.765	.566	.341
C	4	⇩	⇩	⇩	⇩	⇩	⇩	⇩	1.45	1.34	1.17	1.01	.814	.617	.393
D	5	⇩	⇩	⇩	⇩	⇩	⇩	1.65	1.53	1.40	1.24	1.07	.874	.675	.455
E	7	⇩	⇩	⇩	⇩	2.00	1.88	1.75	1.62	1.50	1.33	1.15	.955	.755	.536
F	10	⇩	⇩	⇩	2.24	2.11	1.98	1.84	1.72	1.58	1.41	1.23	1.03	.828	.611
G	15	2.64	2.53	2.42	2.32	2.20	2.06	1.91	1.79	1.65	1.47	1.30	1.09	.886	.664
H	20	2.69	2.58	2.47	2.36	2.24	2.11	1.96	1.82	1.69	1.51	1.33	1.12	.917	.695
I	25	2.72	2.61	2.50	2.40	2.26	2.14	1.98	1.85	1.72	1.53	1.35	1.14	.936	.712
J	30	2.73	2.61	2.51	2.41	2.28	2.15	2.00	1.86	1.73	1.55	1.36	1.15	.946	.723
K	35	2.77	2.65	2.54	2.45	2.31	2.18	2.03	1.89	1.76	1.57	1.39	1.18	.969	.745
L	40	2.77	2.66	2.55	2.44	2.31	2.18	2.03	1.89	1.76	1.58	1.39	1.18	.971	.746
M	50	2.83	2.71	2.60	2.50	2.35	2.22	2.08	1.93	1.80	1.61	1.42	1.21	1.00	.774
N	75	2.90	2.77	2.66	2.55	2.41	2.27	2.12	1.98	1.84	1.65	1.46	1.24	1.03	.804
O	100	2.92	2.80	2.69	2.58	2.43	2.29	2.14	2.00	1.86	1.67	1.48	1.26	1.05	.819
P	150	2.96	2.84	2.73	2.61	2.47	2.33	2.18	2.03	1.89	1.70	1.51	1.29	1.07	.841
Q	200	2.97	2.85	2.73	2.62	2.47	2.33	2.18	2.04	1.89	1.70	1.51	1.29	1.07	.845
		.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00	
		Acceptability Quality Levels (tightened inspection)													

Source: United States Department of Defense, *Military Standard, Sampling Procedures and Tables for Inspection by Variables for Percent Defective*, MIL-STD-414, U.S. Government Printing Office, Washington, DC, 1957, p. 39.
All AQL values are in percent defective.
⇩ Use first sampling plan below arrow, that is, both sample size as well as k value. When sample size equals or exceeds lot size, every item in the lot must be inspected.

Fig. 4. Table B-1, giving *k*, acceptance constant.

IV. Results and Discussion

We now present the web page where a user can test a sample, by automatically following the decision procedure defined in the standard, and verify some values of 'Table B-3' (as in Fig. 3) in order to discuss whether it is rendered superfluous, and more so the extensive, verbose 'Table B-5' (as in Fig. 5).

TABLE T12.7: MIL-STD-414 Table B-5—Table for estimating the lot percent defective using standard deviation method.[a]

Q_U or Q_L	Sample Size															
	3	4	5	7	10	15	20	25	30	35	40	50	75	100	150	200
0	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
.10	47.24	46.67	46.44	46.26	46.16	46.10	46.08	46.06	46.05	46.05	46.04	46.04	46.03	46.03	46.02	46.02
.20	44.46	43.33	42.90	42.54	42.35	42.24	42.19	42.16	42.15	42.13	42.13	42.11	42.10	42.09	42.08	42.08
.30	41.63	40.00	39.37	38.87	38.60	38.44	38.37	38.33	38.31	38.29	38.28	38.27	38.25	38.24	38.22	38.22
.31	41.35	39.67	39.02	38.50	38.23	38.06	37.99	37.95	37.93	37.91	37.90	37.89	37.87	37.86	37.84	37.84
.32	41.06	39.33	38.67	38.14	37.86	37.69	37.62	37.58	37.55	37.54	37.52	37.51	37.49	37.48	37.46	37.46
.33	40.77	39.00	38.32	37.78	37.49	37.31	37.24	37.20	37.18	37.16	37.15	37.13	37.11	37.10	37.09	37.08
.34	40.49	38.67	37.97	37.42	37.12	36.94	36.87	36.83	36.80	36.78	36.77	36.75	36.73	36.72	36.71	36.71

(a) Table B-5 (title).

1.40	0.00	3.33	5.88	6.98	7.44	7.69	7.80	7.86	7.90	7.92	7.94	7.97	8.01	8.02	8.04	8.05
1.41	0.00	3.00	5.66	6.80	7.27	7.53	7.64	7.70	7.74	7.77	7.79	7.82	7.86	7.87	7.89	7.90
1.42	0.00	2.67	5.44	6.62	7.10	7.37	7.49	7.55	7.59	7.62	7.64	7.67	7.71	7.73	7.74	7.75
1.43	0.00	2.33	5.23	6.45	6.94	7.22	7.34	7.40	7.44	7.47	7.50	7.52	7.56	7.58	7.60	7.61
1.44	0.00	2.00	5.01	6.27	6.78	7.07	7.19	7.26	7.30	7.33	7.35	7.38	7.42	7.44	7.46	7.47
1.45	0.00	1.67	4.81	6.10	6.63	6.92	7.04	7.11	7.15	7.18	7.21	7.24	7.28	7.30	7.31	7.33
1.46	0.00	1.33	4.60	5.93	6.47	6.77	6.90	6.97	7.01	7.04	7.07	7.10	7.14	7.16	7.18	7.19
1.47	0.00	1.00	4.39	5.77	6.32	6.63	6.75	6.83	6.87	6.90	6.93	6.96	7.00	7.02	7.04	7.05
1.48	0.00	.67	4.19	5.60	6.17	6.48	6.61	6.69	6.73	6.77	6.79	6.82	6.86	6.88	6.90	6.91
1.49	0.00	.33	3.99	5.44	6.02	6.34	6.48	6.55	6.60	6.63	6.65	6.69	6.73	6.75	6.77	6.78

(b) Table B.5 (continued), sector for $1.40 \leqslant Q \leqslant 1.49$.

1.70	0.00	0.00	0.66	2.62	3.41	3.84	4.02	4.12	4.18	4.22	4.25	4.30	4.35	4.38	4.41	4.42
1.71	0.00	0.00	0.55	2.51	3.31	3.75	3.93	4.02	4.09	4.13	4.16	4.20	4.26	4.29	4.31	4.32
1.72	0.00	0.00	0.45	2.41	3.21	3.65	3.83	3.93	3.99	4.04	4.07	4.11	4.17	4.19	4.22	4.23
1.73	0.00	0.00	0.36	2.30	3.11	3.56	3.74	3.84	3.90	3.94	3.98	4.02	4.08	4.10	4.13	4.14
1.74	0.00	0.00	0.27	2.20	3.02	3.46	3.65	3.75	3.81	3.85	3.89	3.93	3.99	4.01	4.04	4.05
1.75	0.00	0.00	0.19	2.11	2.93	3.37	3.56	3.66	3.72	3.77	3.80	3.84	3.90	3.93	3.95	3.97
1.76	0.00	0.00	0.12	2.01	2.83	3.28	3.47	3.57	3.63	3.68	3.71	3.76	3.81	3.84	3.87	3.88
1.77	0.00	0.00	0.06	1.92	2.74	3.20	3.38	3.48	3.55	3.59	3.63	3.67	3.73	3.76	3.78	3.80
1.78	0.00	0.00	0.02	1.83	2.66	3.11	3.30	3.40	3.47	3.51	3.54	3.59	3.64	3.67	3.70	3.71
1.79	0.00	0.00	0.00	1.74	2.57	3.03	3.21	3.32	3.38	3.43	3.46	3.51	3.56	3.59	3.63	3.63

(c) Table B-5 (continued), sector for $1.70 \leqslant Q \leqslant 1.79$.

Fig. 5. Table B-5 (10 pages, 628 – 637, parts), with $0 \leqslant Q \leqslant 3.90$.

A. Test your sample

The adopted example, in III.A (Table 1), or others supplied by the interested reader can be run at the web page "Test your sample" in List 1, where web pages related to the subject are mentioned. Web page 'b)' applies directly to this subsection,

List 1. Web pages for computations

a) "Transforming Q into ω in AS by variables"

→ http://web.tecnico.ulisboa.pt/ mcasquilho/compute/qc/Fx-ASvarQtoLotFD.php

b) "Test your sample"

→ http://web.tecnico.ulisboa.pt/mcasquilho/compute/icqem2026/accByVars/P.testsample.php

The example in Table 1 embodies the default data in the web page, as depicted in Fig. 6. With these default values, the user just clicks the 'Execute' button to obtain the results in a new, temporary web page that reproduces all the computed results (equal to those already shown, omitted here for shortness) and shows a plot of the sample data, as shown in Fig. 7. In this particular case, it happens that one point, $x_3 = 619$, lies below the lower spec, $L = 620$. This would be a defective if our AS were by attributes, but, strangely to the diffident layman, that criterion is extraneous to inspection by variables.

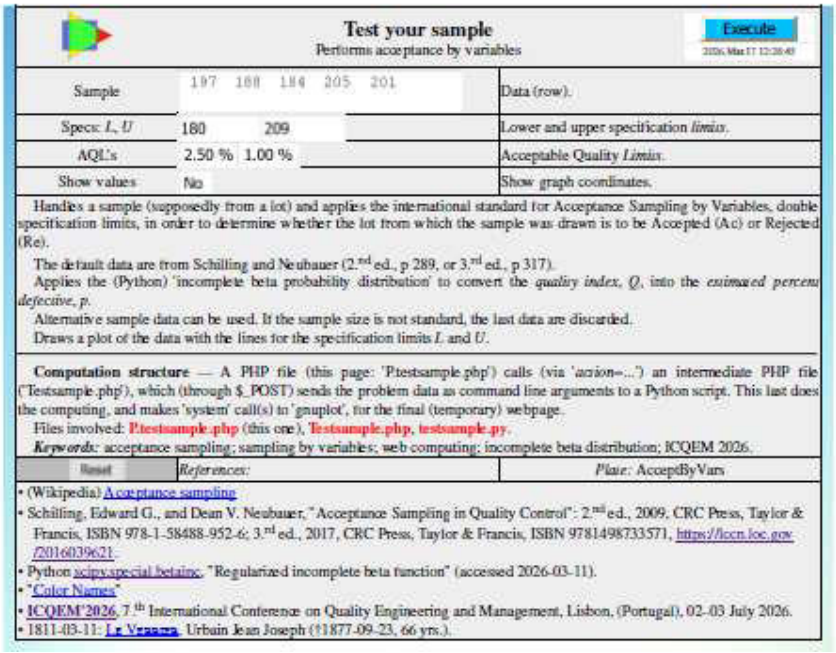


Fig. 6. Web page to test a sample in AS by variables (Form 2).

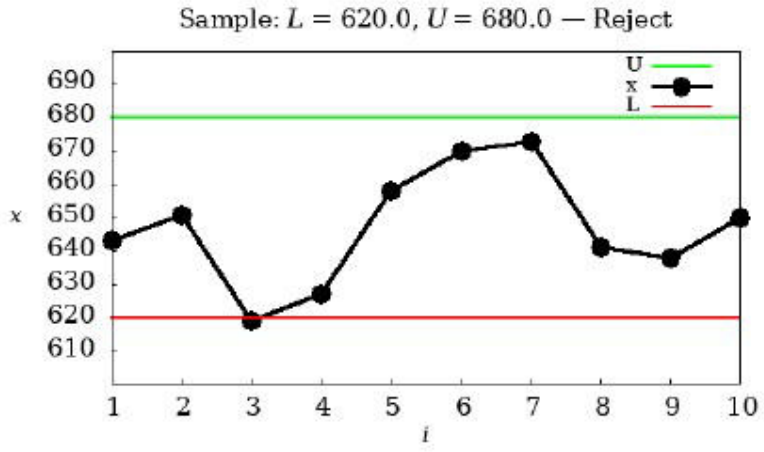


Fig. 7. Plot of the example sample, with given specification limits, *L* and *U*.

B. Quasi-superfluous tables

In the expression "quasi-superfluous tables". we intermix two cases: tables that make sense for the easy use of a standard; and extensive tables that simple mathematical expressions can replace or, at least, accompany.

To illustrate the idea above, we recover the example given in Table 1. To confirm our objective about the nonnecessity of certain tables, either superfluous or quasi-superfluous, we will use 'a)' and 'b)' to obtain 'c)':

a) *k* in 'Table B-1' (Fig. 4)
b) Eq. (1) or, equivalently, (2) or the 'scipy.special.betainc' in the code in III.B
c) *M* in 'Table B-3' (Fig. 3)

Thus, from the example, let us take sample size letter F (*n* = 10) and AQL = 1%. It is *k* = 1.72 (if rounded from 1.715) from 'Table B-1' (Fig. 4). Using, e.g., Excel, as in (2), we get *M* = 3.261%, which rounds to 3.26%, reproducing the value in Table 3. For the other *k* = 1.41 (if rounded from 1.409), we get *M* = 7.285%, rounding to 7.29%, reproducing again the value in the table. Now, by using the conversion in the verbose 'Table B-5' (Fig. 5), we read, respectively, 3.21% and 7.27%, which (before rounding) coincide with the above.

In sum, we did the verification as above for a region of the tables of the *k* values (Table B-1', Fig. 4), and of the *M* values (Table B-3', Fig. 3). The result is on Fig. 8. With one exception, *n* = 5, AQL = 1%, where the *k* corresponding to *M* does not round correctly (1.52 to replace 1.53), all the values of *M* are, indeed, the direct transformation we mentioned in (1) or (2), according to our affirmations in the article.

	k					*M*				
n \ AQL	0,65%	1%	1,50%	2,50%	4%	0,65%	1%	1,50%	2,50%	4%
5	1,65	1,53	1,40	1,24	1,07					
	1,64630	1,52494	1,40230	1,23660	1,06800	1,33%	3,33%	5,83%	9,80%	14,39%
7	1,75	1,62	1,50	1,33	1,15					
	1,7470	1,6210	1,4956	1,3258	1,1510	2,14%	3,55%	5,35%	8,40%	12,25%
10	1,84	1,72	1,58	1,41	1,23					
	1,8400	1,7150	1,5810	1,4090	1,2335	2,17%	3,26%	4,77%	7,29%	10,54%
15	1,91	1,79	1,65	1,47	1,3					
	1,9148	1,7870	1,6550	1,4745	1,2965	2,11%	3,05%	4,31%	6,56%	9,46%
20	1,96	1,82	1,69	1,51	1,33					
	1,9570	1,8230	1,6930	1,5125	1,3315	2,05%	2,95%	4,09%	6,17%	8,92%
25	1,98	1,85	1,72	1,53	1,35					
	1,9840	1,8495	1,7162	1,5344	1,3515	2,00%	2,86%	3,97%	5,97%	8,64%
30	2	1,86	1,73	1,55	1,36					
	2,0000	1,8630	1,7292	1,5467	1,3640	1,98%	2,83%	3,91%	5,86%	8,47%
35	2,03	1,89	1,76	1,57	1,39					
	2,0300	1,8905	1,7580	1,5740	1,3885	1,87%	2,68%	3,70%	5,57%	8,10%
40	2,03	1,89	1,76	1,57	1,39					
	2,03480	1,89100	1,75950	1,57497	1,39050	1,88%	2,71%	3,72%	5,58%	8,09%
50	2,08	1,93	1,80	1,61	1,42					
	2,0800	1,9330	1,7970	1,6117	1,4240	1,71%	2,49%	3,45%	5,20%	7,61%

Fig. 8. Verification of values of M from (original) k.

V. Conclusions

In this study in Quality, we addressed Acceptance Sampling (AS), the complement to Statistical Process Control, focusing on its component of AS "by variables" (vs. "by attributes") in the case of double specification limits (lower and upper). We started by the basic operations in AS by variables as stated in the applicable international standards and related the cases of single and double specification limits, the so-called Form 1 and Form 2, respectively, this latter being clearer and more intuitive, namely in the context of organizations' management.

We discussed the conversion of the *acceptance constants*, "*k*", into the *maximum allowable percent defective*, "*M*", which are associated to an extensive, verbose table (around 15 pages) in the applicable standard. We showed that at least this latter table is superfluous and its computation is direct and easy, a simple use of the incomplete beta distribution. Yet, surprisingly, the information about this function is silently absent in fundamental sources, while overloading these documents with many pages reducible to a mere mathematical expression.

In the study, besides its content, its form is to be emphasized, both through the available web page and by the concomitant insertion of hyperlinks for the verification of our computations. In the technological publications, such links, again surprisingly not found, would permit the interested reader to verify and try with other data. This methodology makes the information more useful and scientifically trustworthy, and promotes mutual advantages for Academia and Industry.

VI. Acknowledgments

We thank (M. Casquilho and R. Galhano) the Department of Chemical Engineering, *Instituto Superior Técnico* (IST), University of Lisbon (ULisboa), Portugal, and CERENA, *Centro de Recursos Naturais e Ambiente* (Centre for Natural Resources and Environment), doi:10.54499/UID/04028/2025; and (F. Rolo and P. Pimpão dos Santos) the Department of Computer Science and Engineering, IST, ULisboa. We also thank CIIST, *Centro de Informática do IST* (Computing Centre of IST), ULisboa, which supports the public system.

ORCID

MC	0000-0003-2062-370X	FR	0009-0005-4370-1813
PPS	0009-0002-6750-9007	RG	0000-0002-6738-0672

References

Alves, Luiz Augusto de Almeida. 1970. "Conversion of the Quality Index into Estimated Percent Defective."
Arsham, H. 1994. "Bio, Research, Web Textbooks, Services." http://home.ubalt.edu/ntsbarsh/Business-stat/home.html.
Barros, M., and M. Casquilho. 2019. "Linear Programming with CPLEX: An Illustrative Application over the Internet." *14.ª Iberian Conf. on Information Systems and Technologies* (Coimbra, Portugal), 1.
Carolino, E., M. Casquilho, M. Rosario Ramos, and I. Barão. 2016. "Applied Scientific Computing over the Web: Robust Methods in Acceptance Sampling for Weibull Variables." *ISBIS 2016 Meeting on Statistics in Business and Industry* (Barcelona, Spain), 1.
Casquilho, M. 2013. "Computação Científica, Internet, Indústria (Scientific Computing, Internet, Industry)." *1st Portuguese Meeting on Mathematics for Industry* (Porto, Portugal), 1.
Casquilho, M. 2025. "Assignment Problem with Random." http://web.tecnico.ulisboa.pt/ mcasquilho/compute/CISTI_2025/assignMC/P-assignMC.php.
Casquilho, M., and J. Buescu. 2022. "From Unequal Size Sample Sums to Estimating the Standard Deviation." *International Conference on Mathematical Analysis and Applications in Science and Engineering* (Porto, Portugal).
Casquilho, M., R. Galhano, and J. Buescu. 2025. "Effect of Random Variation of the Initial Values of Ordinary Differential Equations and Its Web-Based Application to Chemical Kinetics." *Romanian Journal of Information Science and Technology* 28 (3).
Casquilho, M., and P. Pacheco. 2023. "Three Languages Simulate Polygons and Perform on the Web." *icSoftComp2023, International Conference on Soft Computing and Its Engineering Applications* (Changa, Anand, Gujarat, India).
DeVor, Richard E., Tsong-how Chang, and John W. Sutherland. 2007. *Statistical Quality Design and Control: Contemporary Concepts and Methods*. 2.nd. Pearson/Prentice Hall.
Duncan, Acheson J. 1986. *Quality Control and Industrial Statistics*. 5.th. Irwin (1123 pp).
Gunther, Bert. 1998. "Farewell Fusillade." *Quality Progress* 31 (4): 111–14.
ISO/TC 69/SC 5. 2022a. *Acceptance Sampling Plans and Procedures for the Inspection of Bulk Materials*. International Standard. ISO.
ISO/TC 69/SC 5. 2022b. *Sampling Procedures for Inspection by Variables*. International Standard. ISO.
Kuntsche, S., T. Barz, R. Kraus, H. Arellano-Garcia, and G. Wozny. 2011. "MOSAIC a Web-Based Modeling Environment for Code Generation." *Computers & Chemical Engineering* 35 (11): 2257–73. https://doi.org/10.1016/j.compchemeng.2011.03.022.
Lieberman, G. J., and G. J. Resnikoff. 1957. *Tables of the Non-Central t-Distribution*. Stanford University Press.
PDF Xchange. 2026. https://www.pdf-xchange.com/product/pdf-xchange-editor.
Ponce, V. M. 1999. "Online Calculations." https://ponce.sdsu.edu/.
Python.scipy.special.betainc. n.d. "Incomplete Beta Function." https://docs.scipy.org/doc/scipy/reference/generated/scipy.special.betainc.html##scipy.
Sarkadi, K. (Károly), and I. (István) Vincze. 1974. *Mathematical Methods of Statistical Quality Control*. Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic Press.
Schilling, Edward G., and Dean V. Neubauer. 2009. *Acceptance Sampling in Quality Control*. 2.nd. Statistics : Textbooks and Monographs. CRC Press. https://ci.nii.ac.jp/ncid/BA90544597.
Schilling, Edward G., and Dean V. Neubauer. 2017. *Acceptance Sampling in Quality Control*. 3.rd. Statistics: Textbooks and Monographs. CRC Press.

■

-

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Industry 4.0 Maturity Models: A Systematic Analysis

Larissa Freire Lekakis[1], Fabio Neves Puglieri [2] and Camila Fabrício Poltronieri [1,3]

1 Federal University of Goiás; larissa.freire@discente.ufg.br
2 Federal Technological University of Paraná; puglieri@utfpr.edu.br
3 University of São Paulo; camilafp@usp.br

Abstract*—This study aims to analyze Industry 4.0 (I4.0) maturity models applied across different organizational contexts, focusing on identifying their key dimensions, benefits, and limitations. It seeks to understand how these models support companies in assessing and advancing their digital transformation processes. The theoretical foundation is based on concepts of digital transformation, organizational maturity, and technological innovation, linking the evolution of I4.0 to strategic management theories and capability development. The research adopts a Systematic Literature Review (SLR) to identify, evaluate, and synthesize existing studies on I4.0 maturity models. A total of sixteen relevant studies were analyzed, considering their evaluation criteria, structural dimensions, and practical applications. The findings reveal significant variation among the models in terms of structure, scope, and applicability. While they provide useful frameworks for assessing digital readiness, most models lack proper integration between technical, organizational, and human dimensions. The study highlights the need for more adaptable and empirically validated models that can be applied across diverse industrial contexts. From a theoretical perspective, it contributes by mapping existing models and identifying conceptual gaps. Practically, it supports managers in selecting models that better align with their organizational needs and promotes greater efficiency in digital transformation initiatives. Its originality lies in consolidating fragmented knowledge on I4.0 maturity models and offering a structured perspective that bridges theory and practical application.

Keywords: Industry 4.0, Maturity Models, Dimensions, Systematic Literature Review

I. Introduction

Industry 4.0 (I4.0) represents a new phase of industrial transformation driven by digitalized processes, reconfigured business models, and intelligent technologies (Forschungsbeirat Industrie 4.0, 2025; Vaidya et al., 2018). It enables product customization, efficiency gains, competitiveness, and organizational resilience (Shabur, 2024). However, its implementation still faces challenges such as unclear economic returns, workforce skill gaps, technical and legal barriers, high costs, and limited resources (Alsheyadi, 2024; Moura et al., 2020). These obstacles affect organizations of all sizes, differing mainly in technological demands and workforce qualification (Pech and Vaněček, 2022). To address these issues, governments and private institutions have promoted strategies and policies to support I4.0 adoption due to its strategic importance (Ghobakhloo, 2020). I4.0 adoption varies significantly across countries according to technological, economic, social, and regulatory conditions, as well as challenges related to cybersecurity, interoperability, and workforce training (Ramos et al., 2023; Sadeghi-Niaraki, 2020). Consequently, several maturity models have been proposed to align organizational strategies with technological demands (Puchan et al., 2018), incorporating structural, technical, cultural, and behavioral dimensions (Bititci et al., 2015). Nevertheless, these models differ considerably in scope, detail, and sectoral applicability (Elibal and Özceylan, 2024). Previous studies also report limitations such as restricted validation beyond manufacturing (Castelo-Branco et al., 2022), insufficient consideration of interdependencies among dimensions (Mohanta and Mahanty, 2024), lack of multi-phase roadmaps, and limited empirical assessment (Wagire et al., 2021). Despite the growing literature, there remains a lack of comparative studies evaluating how maturity model dimensions, gaps, and benefits differ across sectors and regional contexts. This gap limits the selection and adaptation of suitable models for distinct organizational realities. Accordingly, this study investigates the following research question: *Which dimensions, gaps, and benefits emerge in Industry 4.0 maturity models applied across sectors and regions?* By addressing this question, the study aims to support more flexible and context-oriented applications, while strengthening theoretical understanding and practical decision-making.

II. Theoretical Framework

Di Bona et al. (2021) highlight that I4.0 is transforming industrial maintenance through the integration of digital and intelligent resources, replacing traditional processes with more efficient and autonomous solutions. This shift is driven by enabling technologies that enhance automation, connectivity, and process integration, resulting in more flexible and intelligent production systems (Calabrese et al., 2022). Key technologies include the Internet of Things, big data, cloud computing, autonomous robotics, additive manufacturing, systems integration, simulation, augmented reality, and cybersecurity (Asadollahi-Yazdi et al., 2020; Gorkhali, 2022; Silva et al., 2024). However, adopting these technologies alone does not ensure successful implementation. I4.0 requires a broader transformation of value chains, increasing flexibility, efficiency, and mass customization (Balkan and Akyuz, 2023). In this context, organizations must assess their maturity and capacity to implement digital practices effectively. Maturity models serve as essential tools to diagnose current conditions and guide technological and organizational development (Ünlü et al., 2024). They help identify gaps, prioritize investments, and monitor digital transformation across different business contexts, including small and medium-sized enterprises (Häring, 2023). Thus, they support not only the assessment of technology adoption but also the structured development of digital capabilities (Alcácer et al., 2022; Colli et al., 2019; Schumacher et al., 2016). The

literature also distinguishes between readiness and maturity. Readiness refers to the initial capacity to begin digital transformation, while maturity reflects a more advanced stage in which technologies are fully integrated into organizational processes (Schumacher et al., 2016). This distinction is crucial for understanding the depth of transformation. Various maturity models have been proposed, ranging from specific to comprehensive approaches. Prim et al. (2022) presents a model with eight dimensions, including leadership, strategy, and knowledge sharing, allowing organizations to assess maturity on a scale of 1 to 5 and define target stages. More comprehensive models, such as Schumacher et al. (2016), are widely used to evaluate readiness by translating abstract concepts into measurable parameters (Ávila-Bohórquez and Gil-Herrera, 2022). Similarly, the IMPULS Industrie 4.0 Readiness model (Lichtblau et al., 2015) defines six dimensions and five maturity levels, supporting integrated assessment and decision-making in digital transformation.

III. METHOD

This study conducts a systematic literature review (SLR), guided by the methodological stages of Denyer and Tranfield (2009) and structured according to the PRISMA protocol (Page et al., 2021). The steps comprise: (i) formulation of the research questions, (ii) identification of potentially relevant studies, (iii) screening and careful evaluation of the selected publications, (iv) analysis and synthesis of the findings, and (v) presentation and interpretation of the results. For the search and selection stage of the bibliographic material, the Scopus and Web of Science databases were chosen for their broad coverage of international journals and their frequent updates, making them reliable and relevant sources for academic research (Singh et al., 2021). Data collection was carried out in February 2025. For the definition of the search strategy, terms related to the central concepts of the investigation were selected, with emphasis on "Maturity Model" and "Industry 4.0" In the consulted databases, the descriptors were combined using Boolean operators (AND, OR) and proximity operators appropriate to each platform.

In Scopus, the PRE/1 operator was applied to locate occurrences in which the terms "matur*" and "model*" appear one word apart, while in Web of Science, the NEAR/1 operator was used to identify records with the same terms in close proximity, regardless of order. Additionally, terms related to I4.0 were included in the title fields, such as: ("i4.0" "fourth AND industr* AND revolut*" AND "industr* 4.0"), with the aim of broadening the search scope and capturing different forms of reference to the topic, as detailed in Figure 1.

The next stage involved extracting the articles from the databases and importing them into Rayyan (Ouzzani et al., 2016), a free web application used for reference management and to support study inclusion and exclusion decisions. Searches conducted in the Scopus and Web of Science databases identified 288 records. After removing 88 duplicate records, 200 studies remained for screening. During screening, 98 conference papers were excluded according to the predefined criteria, leaving 102 studies for eligibility assessment. The selection process followed three sequential filters. In the first filter, titles and abstracts were reviewed, excluding 39 studies that did not present maturity models, focused only on assessment processes, lacked case studies, relied exclusively on interviews, or corresponded to bibliometric, exploratory, or review studies. Thus, 63 studies remained. In the second filter, introductions and conclusions were analyzed to verify whether the studies developed and applied maturity models in real contexts rather than only discussing or reusing existing models. At the end of this stage, 34 studies met the selection criteria. Finally, full-text analyses were conducted, resulting in a final sample of 16 articles included in the qualitative synthesis. These studies provided the most relevant evidence regarding geographic contexts, analyzed dimensions, sectoral applications, and practical implementation of Industry 4.0 maturity models. The extracted data were systematically analyzed to address the research question and identify the main benefits, limitations, opportunities, and threats associated with applying Industry 4.0 maturity models. Based on these findings, a SWOT matrix was developed to consolidate the most recurrent internal and external factors affecting model implementation.

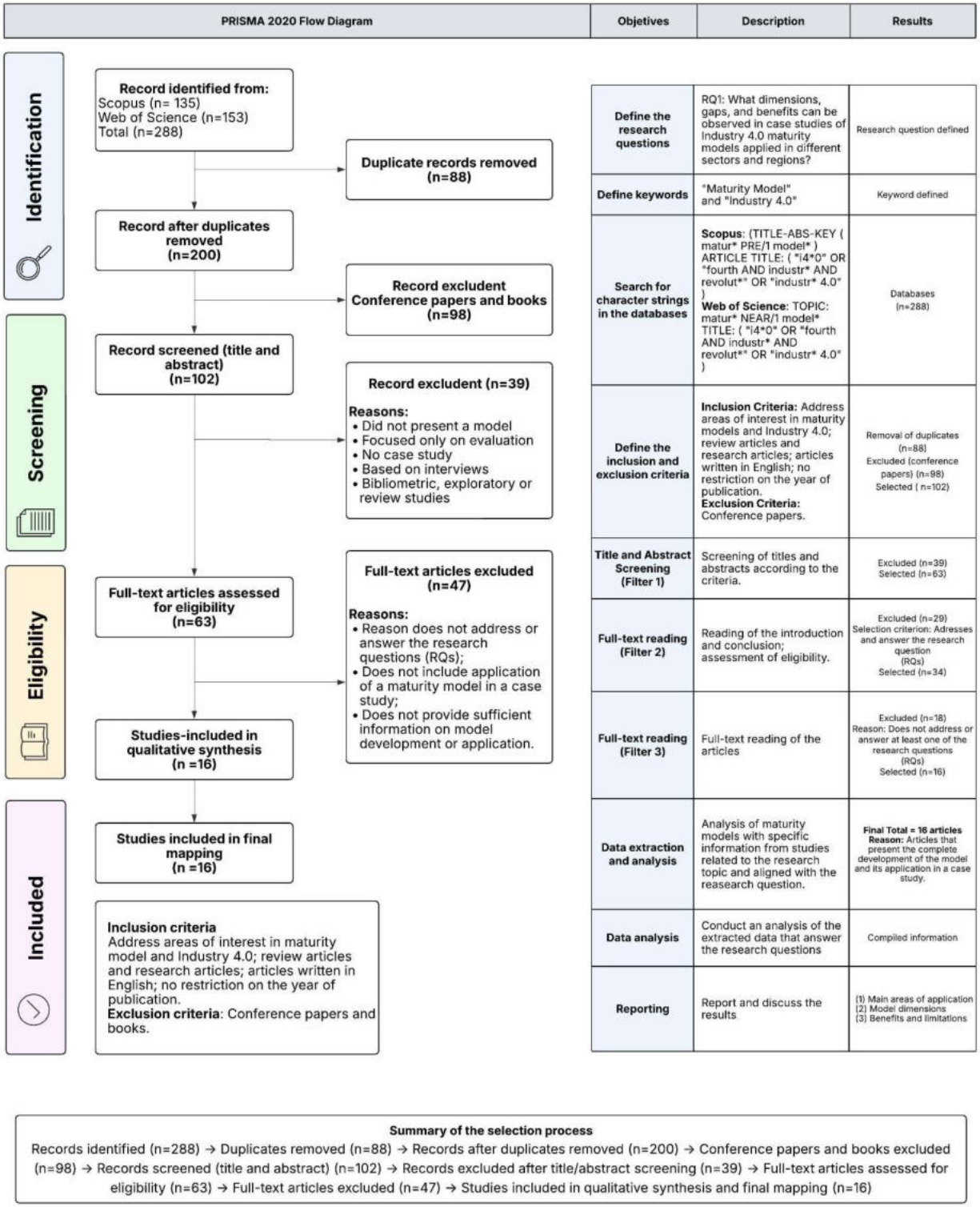


Fig. 1. Literature search and filtering process. Adapted from the PRISMA protocol (Page et al., 2021)

IV. Results and Discussion

This section presents the main results of the study, integrating the analysis of I4.0 maturity models with a discussion of their applications, limitations, and theoretical contributions. The findings are based on 16 selected articles that present dimensions, maturity levels, limitations, and benefits identified, collected, supporting critical analyses between the models. There is a strong representation of the automotive sector (25% of the articles), one of the sectors most impacted by I4.0. This is mainly due to the adoption of collaborative robots, automated control systems, and digital technologies,

enabling greater production flexibility, reduced logistics costs, and improved quality control (Rüßmann et al., 2015). These results suggest that I4.0 maturity models are more readily applicable to traditional industrial sectors, such as automotive, because of their higher levels of automation and digitalization. Consequently, they may serve as references for adaptation in less explored sectors (Galizia et al., 2023). Another 25% of the studies address diverse models applied in sectors such as plastics, metallurgy (Chonsawat and Sopadang, 2021), aerospace, textile, heavy industries (Sajjad et al., 2024), and services, commerce, and construction (Castelo Branco et al., 2022). These variations show that no single model fits all sectoral needs, reinforcing the need for continuous adaptation to different industrial contexts. Geographic patterns also influence the development and application of I4.0 maturity models. Asia leads with 43.75% of publications, reflecting strong industrial policies and advanced technological infrastructure, consistent with Tripathi and Gupta's (2021) identification of East Asia as a global reference. South America accounts for 25% of the studies, with Brazil particularly well represented, supported by national initiatives such as the More Production Plan (CNI, 2024). Collaborative studies involving Asia and Europe contribute 12.5% of the publications, while Europe alone also represents 12.5%. The Asia-Pacific region accounts for 6.25%, indicating smaller yet context-specific contributions. Overall, these results emphasize that organizational and regional contexts shape methodological choices and influence which dimensions receive greater emphasis, highlighting the importance of flexible and context-sensitive models. In addition to regional differences, maturity model analysis depends on the dimensions used for assessment. These dimensions help identify strengths, weaknesses, and improvement priorities (Głuszek, 2018). Across the 16 studies, 99 dimensions were identified, with 42 appearing only once and 57 recurring in multiple models. This dispersion reflects diverse theoretical approaches and limited standardization, while recurring dimensions indicate emerging areas of convergence. Figure 2 presents a Pareto chart illustrating the frequency and relative importance of these dimensions.

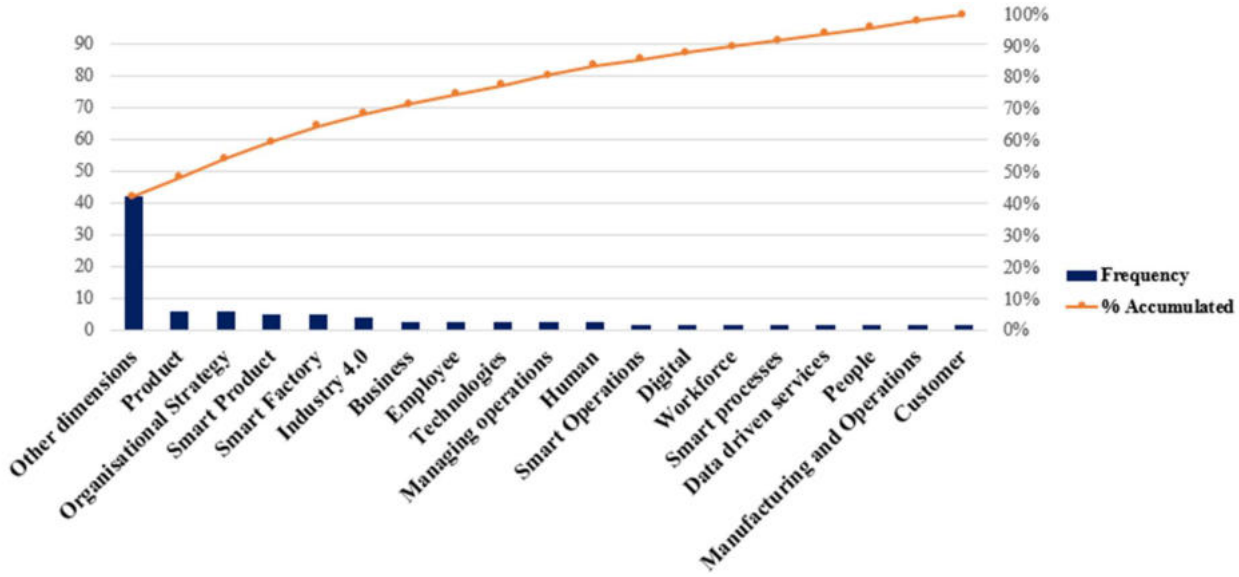


Fig. 2. Pareto Analysis of the Main Dimensions.

The application of Pareto analysis to the dimensions of the analyzed maturity models shows that a few categories concentrate most of the attention in the I4.0 literature. The dimensions "Product," "Organisational Strategy," "Smart Product," "Smart Factory," "Industry 4.0", "Business," "Employee," "Technologies," and "Managing Operations" together account for over 80% of the recorded occurrences. In contrast, the strong presence of the "Other dimensions" category indicates diverse approaches and a lack of consensus among models, highlighting limitations in standardization and the need for adaptation to organizational specificities. The Pareto analysis also shows that models are tailored to sector, size, and business context, reinforcing their dynamic and adaptable nature, while revealing difficulties in comparison and the importance of selecting dimensions aligned with each organization's strategy. Table 1 presents the most recurrent maturity dimensions, their definitions, and the respective authors identified in the analyzed sample. This structure facilitates comparison across approaches and highlights how each dimension relates to I4.0 maturity in different organizational contexts.

TABLE I. MAIN DIMENSIONS AND DEFINITIONS OF MATURITY MODELS

Dimensions	*Concepts*	*Authors*
Smart products	These are sensors and intelligent devices equipped with connectivity, interconnected in networks that enable continuous communication and real-time data exchange.	Santos and Martinho (2020), Sajjad et al. (2024)
Smart Factory	It enables real-time communication among machines, products, people, and infrastructure, optimizing operations through mobile devices and flexible layouts.	Alcácer et al., (2022), Vasconcellos et al. (2021)
Organisational Strategy	It refers to the organizational guidelines that support maturity and technological transformation, highlighting management's strategic role.	Chonsawat and Sopadang (2021), Wagire et al. (2021)
Product	It involves customization and the integration of digital capabilities to enhance customer value and operational efficiency.	Ávila-Bohórquez and Gil-Herrera (2022), Ramanathan and Samaranayake (2022), Hajoary, Amrita and Garza-Reyes (2024)
Industry 4.0	It represents an integrated set of technologies, strategies, and digital competencies that drive business innovation and competitiveness.	Dal Forno et al. (2023), Sajjad et al. (2024), Wagire et al. (2021)
Business	It involves business models, marketing, digital practices, and cloud computing to strengthen the organizational and technological foundation for digital transformation.	Chonsawat and Sopadang (2021), Faisal et al. (2024)
Employee	It includes training and development to adapt workers to the technological environment, reinforcing the role of people in consolidating organizational maturity.	Alcácer et al. (2022), Faisal et al. (2024)
Technologies	It covers the adoption and integration of technologies and information systems to support production operations.	Chonsawat and Sopadang (2021), Dal Forno et al. (2023), Wankhede and Vinodh (2023)

Smart Products, Product, and Technologies are the dimensions most directly linked to technological integration and the digitalization of production processes. They reflect the adoption of smart devices, interconnected systems, and digital solutions, enabling companies to assess automation, integration, and strategic alignment levels. These dimensions indicate not only technological adoption but also progress toward more efficient, agile, and data-driven operations aligned with I4.0 principles. The I4.0 and Smart Factory dimensions share similar concepts and play a structuring role by integrating advanced technologies, intelligent systems, and real-time connectivity. Together, they highlight companies' potential to develop autonomous, adaptable, and data-driven production environments, fostering gains in efficiency, innovation, and competitiveness.

Building on this foundation, the dimensions Organisational Strategy, Business, Employee, and Managing Operations extend the maturity perspective beyond operations. They are essential to organizational development, reflecting the ability to align strategies, business models, and practices with I4.0 requirements, while emphasizing the role of people. These elements support accurate diagnostics and the planning of actions aimed at continuous improvement, competitiveness, and efficiency. To analyze these dimensions in an integrated way, evaluative tools are essential. According to Benzaghta et al. (2021), SWOT analysis (Strengths, Weaknesses, Opportunities, and Threats) is a simple and widely used tool, particularly effective in the evaluation stage. It supports the identification of strengths, weaknesses, opportunities, and threats related to model implementation. Adapting the SWOT matrix to I4.0 maturity models enables a clearer distinction between internal (strengths and weaknesses) and external factors (opportunities and threats), improving analytical consistency (Table 2).

TABLE II. TERMS AND MEANINGS OF THE SWOT MATRIX IN THE CONTEXT OF THIS STUDY

Terms	*Concepts*
Strength	Benefits of the models for companies.
Weakness	Limitations and/or gaps that hinder the application of the models.
Opportunity	Chances to expand usage and generate new research.
Threat	Risks that compromise the adoption of the models.

Figure 3 presents the most frequently cited benefits in the literature, classified as strengths in the SWOT analysis, and indicates their relative frequency.

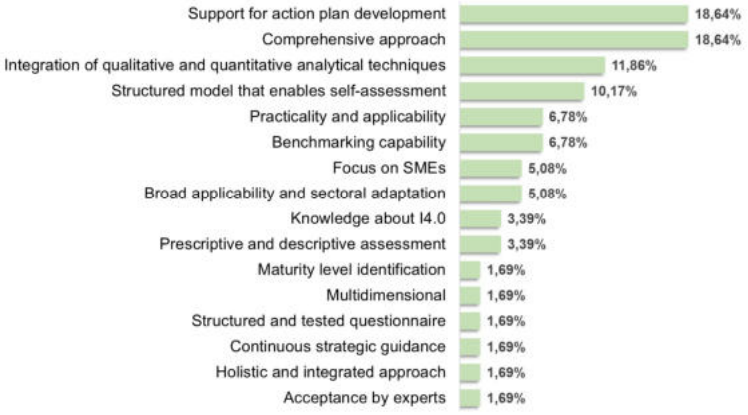


Fig. 3. Main benefits of Industry 4.0 maturity models.

As illustrated in Figure 3, the most frequently mentioned benefits are support for developing action plans and a comprehensive approach. Other relevant aspects include the integration of analytical techniques, self-assessment, and applicability in different contexts, especially in SMEs. These findings reinforce Schumacher et al. (2016) and Mittal et al. (2018), who argue that I4.0 maturity models should encompass organizational, strategic, and human dimensions rather than focusing only on technology. The predominance of benefits related to action planning and comprehensive assessment suggests that organizations increasingly seek integrated managerial tools to support digital transformation. Furthermore, the relevance of self-assessment mechanisms confirms the importance of organizational autonomy, particularly for SMEs with limited resources and technical support. The results also indicate that the most effective models integrate strategy, processes, culture, people, and technology, aligning with the systemic perspective of recent I4.0 literature. Figure 4 presents the main limitations of I4.0 maturity models, especially restricted scope, low generalizability, dependence on specialists, and the need for context-specific adjustments. These limitations corroborate previous studies criticizing the contextual dependence and operational complexity of maturity assessment frameworks, particularly in SMEs. Moreover, the absence of universally applicable frameworks highlights an important gap in the literature, suggesting that future models should prioritize flexibility, scalability, and ease of application.

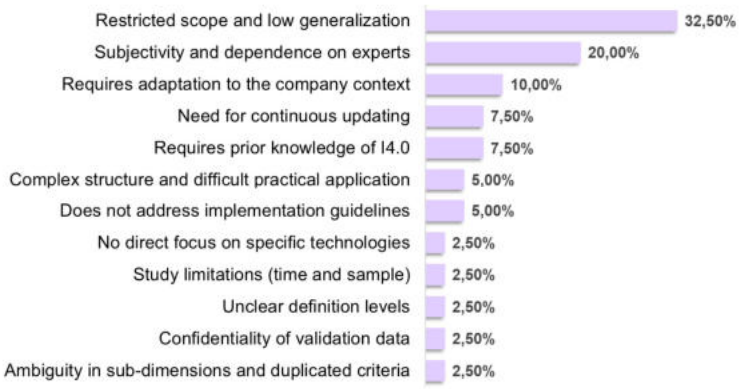


Fig. 4. Main limitations of Industry 4.0 maturity models.

Despite their advantages, these models present limitations that affect their applicability. Restricted scope and low generalizability are among the most common challenges, as many models are developed for specific contexts and require adaptation for broader use. This finding suggests that I4.0 maturity assessment frameworks still lack sufficient flexibility for heterogeneous organizational realities. Accuracy is also affected by reliance on subjective data from self-assessments and interviews, which introduces bias and reduces reliability. This result is consistent with previous studies criticizing the contextual dependence and operational complexity of maturity models, particularly in SMEs. Another challenge is the need to tailor models to each company's context, considering differences in digitalization, infrastructure, and internal integration. Additionally, their high structural complexity, characterized by multiple variables and unclear criteria, makes them difficult for non-specialists to apply. Figure 5 highlights opportunities related to adaptability, benchmarking, continuous improvement, learning support, and roadmap development. This contributes to the literature by reinforcing the strategic potential of adaptable maturity models for technological advancement

and organizational learning. Moreover, these opportunities indicate that simplified and scalable frameworks may enhance practical implementation while supporting more effective digital transformation initiatives globally.

Fig. 5. Main opportunities of Industry 4.0 maturity models.

Maturity models offer strategic opportunities that enhance their relevance. Scalable platforms can be adapted across sectors, company sizes, and geographic contexts, making them suitable for diverse industrial processes. This finding suggests that flexibility is a central requirement for supporting digital transformation in heterogeneous organizational realities. In addition, these models enable benchmarking, allowing companies to compare performance, identify best practices, and develop tailored roadmaps for technological advancement. This result is consistent with previous I4.0 studies emphasizing the strategic role of maturity assessments in guiding investments and continuous improvement initiatives. Figure 6 presents the main threats to the adoption of maturity models, especially limited adoption, technological obsolescence, and inconsistencies in expert assessments. Other relevant barriers include insufficient knowledge of I4.0 concepts and resistance to organizational change. This contributes to the literature by demonstrating that simplified, adaptable, and scalable frameworks may increase reliability, facilitate implementation, and strengthen the practical contribution of maturity models to digital transformation processes. Furthermore, these opportunities indicate that organizations require accessible tools capable of supporting long-term strategic decision-making effectively.

Fig. 6. Main threats of Industry 4.0 maturity models.

The widespread adoption of maturity models faces barriers related to market diversity and organizational differences. Variations in digitalization levels, infrastructure, and integration between sectors hinder the application of standardized solutions. These results suggest that I4.0 maturity models still require greater flexibility to support heterogeneous industrial realities effectively. The lack of coordination across organizational processes and value chains also reduces implementation reliability. This result is consistent with previous studies emphasizing that technological transformation depends on organizational integration and leadership engagement. In addition, rapid technological evolution contributes to model obsolescence, requiring constant revisions to maintain relevance. Cultural resistance, limited knowledge of I4.0 concepts, and subjectivity in expert assessments further compromise diagnostic accuracy and decision-making processes. This contributes to the literature by demonstrating that the effectiveness of maturity models depends not only on technological dimensions, but also on continuous updating, technical training, and organizational readiness. Figure 7 synthesizes the most recurrent SWOT factors affecting model effectiveness across different industrial sectors and regional business environments worldwide.

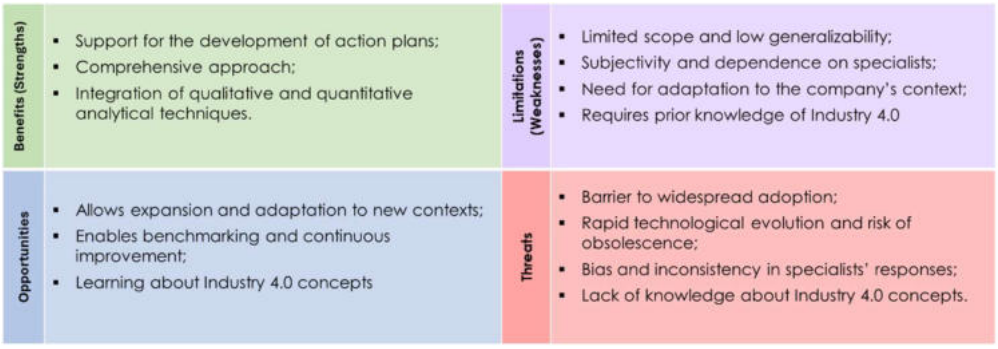


Fig. 7. Synthesis of the SWOT analysis applied to the Industry 4.0 maturity model.

The effectiveness of I4.0 maturity models depends on integrating technical, human, and organizational dimensions while establishing measurable indicators that support consistent assessments across sectors. Their main benefits include structured guidance for action planning, comprehensive evaluations combining qualitative and quantitative techniques, and the capacity to address multiple aspects of digital transformation. These results suggest that adaptable and multidimensional frameworks are essential for supporting organizational evolution toward I4.0. This result is consistent with previous studies emphasizing the strategic role of maturity models in benchmarking, continuous learning, and investment planning. Despite these advantages, limitations such as restricted scope, low generalizability, subjective assessments, and dependence on specialists reduce applicability and require contextual adaptation. External threats, including limited organizational understanding, resistance to adoption, and rapid technological evolution, further compromise implementation effectiveness. This contributes to the literature by demonstrating that continuous updating, technical training, and stronger organizational alignment are fundamental for maintaining reliable and effective maturity assessments across different industrial contexts. These measures may also improve long-term digital transformation performance worldwide

V. Conclusion

Industry 4.0 (I4.0) has driven profound transformations in production processes and organizational management, requiring greater flexibility, technological integration, and innovation capacity. In this context, maturity models emerge as strategic tools to support the assessment and development of organizations toward industrial digitalization. A systematic literature review was conducted to analyze these models across different contexts, focusing on their applications, benefits, and limitations. The analysis, based on an adaptation of the SWOT matrix, shows that although the number of models is increasing, theoretical and methodological gaps still limit their applicability across diverse environments. The findings indicate that effective models must integrate technological, human, and organizational dimensions. They should also include clearly defined maturity levels, objective criteria, and measurable indicators to ensure consistent diagnoses. Flexibility and adaptability are essential, enabling models to keep pace with rapid industrial changes and to be applied across different sectors and company sizes.

From a strategic perspective, these models should support decision-making, investment prioritization, and the development of organizational capabilities. However, key challenges remain, including data reliability, lack of standardization in evaluation criteria, and limited empirical validation of theoretical models. This study contributes by systematizing existing approaches and identifying research gaps in I4.0 maturity models. Nevertheless, some limitations must be acknowledged. The restricted number of analyzed publications may not capture the full diversity of models, and the focus on academic literature may exclude practical case studies. In addition, the predominance of models developed in specific sectoral and geographic contexts limits the generalization of the findings. Future research should expand the analysis to include multiple industries, countries, and comparative approaches. It is also recommended to explore readiness models and conduct empirical validations to assess their real impact on I4.0 implementation. Such efforts can strengthen the link between theory and practice, generate applicable knowledge, and support the continuous improvement of maturity models. Overall, maturity models are promising tools for guiding organizations in their transition to I4.0, provided they are continuously refined and validated. Future advances in this field are expected to contribute not only to theoretical development but also to practical solutions that enhance innovation, competitiveness, and sustainability in digital transformation.

References

Abdullah, A., Saraswat, S. and Talib, F. (2025), "A maturity model for assessing industry 4.0 implementation using data envelopment analysis and best and worst method approaches", International Journal of Productivity and Performance Management, Vol. 74 No. 5, pp. 1533-1563. https://doi.org/10.1108/IJPPM-12-2023-0668

Alcácer, V., Rodrigues, J., Carvalho, H. and Cruz-Machado, V. (2022), "Industry 4.0 maturity follow-up inside an internal value chain: a case study", The International Journal of Advanced Manufacturing Technology, Vol. 119, pp. 5035–5046, doi: 10.1007/s00170-021-08476-3.

Alsheyadi, A.K. (2024), "A literature review on industry 4.0 implementation and performance: a contingency perspective", The TQM Journal, Vol. 36 No. 7, pp. 2048-2070. https://doi.org/10.1108/TQM-03-2023-0073

Asadollahi-Yazdi, E., Couzon, P., Nguyen, N.Q., Ouazene, Y. and Yalaoui, F. (2020), "Industry 4.0: revolution or evolution?", American Journal of Operations Research, Vol. 10 No. 6, pp. 241-268. https://doi.org/10.4236/ajor.2020.106014

Ávila-Bohórquez, J.H. and Gil-Herrera, R.J. (2022), "Proposal and validation of an Industry 4.0 maturity model for SMEs", Journal of Industrial Engineering and Management, Vol. 15 No. 3, pp. 433–454, doi: 10.3926/jiem.3673.

Balkan, D. and Akyuz, G.A. (2023), "Labour productivity analysis of manufacturing sector in Turkey against EU", Journal of Business Economics and Management, Vol. 24 No. 2, pp. 245–273, doi: 10.3846/jbem.2023.19059.

Bayrak, İ.T. and Cebi, F. (2024), "Procedure model for Industry 4.0 realization for operations improvement of manufacturing organizations", *IEEE Transactions on Engineering Management*, Vol. 71, pp. 7901–7912, doi: 10.1109/TEM.2023.3292337.

Beisekenov, I., Suleiman, Z., Tokbergenova, A., Shaikholla, S., Dikhanbayeva, D., El-Thalji, I., Emiris, D., and Turkyilmaz, A. (2022), "Maturity assessment of industry 4.0 implementation in Kazakhstani and Norwegian oil and gas contexts", Journal of Industrial Integration and Management, Vol. 7 No. 4, pp. 455-477. https://doi.org/10.1142/S2424862222500191

Benzaghta, M.A., Elwalda, A., Mousa, M.M., Erkan, I. and Rahman, M. (2021), "SWOT analysis applications: an integrative literature review", Journal of Global Business Insights, Vol. 6 No. 1, pp. 55-73. https://doi.org/10.5038/2640-6489.6.1.1148

Bititci, U.S., Garengo, P., Ateş, A. and Nudurupati, S.S. (2015), "Value of maturity models in performance measurement", International Journal of Production Research, Vol. 53 No. 10, pp. 3062-3085. https://doi.org/10.1080/00207543.2014.970709

Calabrese, A., Dora, M., Levialdi Ghiron, N. and Tiburzi, L. (2022), "Industry's 4.0 transformation process: how to start, where to aim, what to be aware of", Production Planning & Control, Vol. 33 No. 5, pp. 492-512. https://doi.org/10.1080/09537287.2020.1830315

Castelo-Branco, I., Oliveira, T., Simões-Coelho, P., Portugal, J. and Filipe, I. (2022), "Measuring the fourth industrial revolution through the industry 4.0 lens: the relevance of resources, capabilities and the value chain", Computers in Industry, Vol. 138, Article 103633. https://www.sciencedirect.com/science/article/pii/S0166361522000343?via%3Dihub

Chonsawat, N. and Sopadang, A. (2021), "Smart SMEs 4.0 maturity model to evaluate the readiness of SMEs implementing industry 4.0", Chiang Mai University Journal of Natural Sciences, Vol. 20 No. 2, e2021027. https://doi.org/10.12982/CMUJNS.2021.027

Colli, M., Berger, U., Bockholt, M., Madsen, O., Møller, C. and Wæhrens, B.V. (2019), "A maturity assessment approach for conceiving context-specific roadmaps in the industry 4.0 era", Annual Reviews in Control, Vol. 48, pp. 165-177. https://doi.org/10.1016/j.arcontrol.2019.06.001

Confederação Nacional da Indústria (2024), *Financing for New Industry Brazil (NIB): available financing lines*, CNI, ISBN 978-85-7957-271-5, available at: https://www.portaldaindustria.com.br/publicacoes/2024/11/nucleo-de-acesso-ao-credito-nac-2533/#bndes-credito-digital%20

Dal Forno, A.J., Bataglini, W.V., Steffens, F. and Souza, A.A.U. de (2023), "Maturity model tool to diagnose industry 4.0 in the clothing industry", Journal of Fashion Marketing and Management, Vol. 27 No. 2, pp. 201-219. https://doi.org/10.1108/JFMM-09-2021-0241

Denyer, D. and Tranfield, D. (2009), "Producing a systematic review", in Buchanan, D.A. and Bryman, A. (Eds.), The Sage handbook of organizational research methods, Sage Publications Ltd., pp. 671-689.

Di Bona, G., Cesarotti, V., Arcese, G. and Gallo, T. (2021), "Implementation of industry 4.0 technology: new opportunities and challenges for maintenance strategy", Procedia Computer Science, Vol. 180, pp. 424-429. https://doi.org/10.1016/j.procs.2021.01.258

Elibal, K. and Özceylan, E. (2024), "An industry 4.0 maturity model proposal based on total quality management principles: an application to an automotive parts manufacturer", IEEE Transactions on Engineering Management, Vol. 71, pp. 10815-10832. https://doi.org/10.1109/TEM.2024.3397555

Faisal, S.M.F., Banik, S.C. and Sen Gupta, P. (2024), "Development of a readiness model for industry 4.0 using analytical hierarchy process and fuzzy inference system: Bangladesh perspective", Heliyon, Vol. 10 No. 1, e23664. https://doi.org/10.1016/j.heliyon.2023.e23664

Forschungsbeirat Industrie 4.0 (2025), Industrie 4.0 – Forschungs- und Entwicklungsperspektiven, available at: https://www.acatech.de

Galizia, F.G., Bortolini, M. and Calabrese, F. (2023), "A cross-sectorial review of industrial best practices and case histories on industry 4.0 technologies", Systems Engineering, Vol. 26, pp. 908-924. https://doi.org/10.1002/sys.21697

Ghobakhloo, M. (2020), "Industry 4.0, digitization, and opportunities for sustainability", Journal of Cleaner Production, Vol. 252, Article 119869. https://doi.org/10.1016/j.jclepro.2019.119869

Głuszek, E. (2018), "Dimensions and stages of the CSR maturity", Research Papers of Wrocław University of Economics, No. 520, pp. 64–80, doi: 10.15611/pn.2018.520.06.

Gorkhali, A. (2022), "Industry 4.0 and enabling technologies integration: framework and challenges", Journal of Industrial Integration and Management. https://doi.org/10.1142/S2424862222500075

Hajoary, P.K., Amrita, M.A. and Garza-Reyes, J.A. (2024), "Industry 4.0 maturity assessment: a multi-dimensional indicator approach", International Journal of Productivity and Performance Management, Vol. 73 No. 4, pp. 981-1004. https://doi.org/10.1108/IJPPM-07-2022-0325

Häring, K., Pimentel, C. and Teixeira, L. (2023), "Industry 4.0 implementation in small- and medium-sized enterprises: recommendations extracted from a systematic literature review with a focus on maturity models", Logistics, Vol. 7 No. 4, p. 99. https://doi.org/10.3390/logistics7040099

Lichtblau, K.; Stich, V.; Bertenrath, R.; Blum, M.; Bleider, M.; Millack, A.; Schmitt, K.; Schmitz, E.; Schröter, M. Industrie 4.0 Readiness. Aachen; Cologne: IMPULS-Stiftung / VDMA, 2015. https://impuls-stiftung.de/wp-content/uploads/2022/05/Industrie-4.0-Readiness-english.pdf

Mittal, S., Khan, M.A., Romero, D. and Wuest, T. (2018), "A critical review of smart manufacturing and industry 4.0 maturity models: implications for small and medium-sized enterprises (SMEs)", Journal of Manufacturing Systems, Vol. 49, pp. 194-214. https://doi.org/10.1016/j.jmsy.2018.10.005

Mohanta, P. R., & Mahanty, B. (2024), "Assessing Industry 4.0 implementation maturity in manufacturing MSMEs – a graph theory and matrix-based approach", *The TQM Journal*. https://doi.org/10.1108/TQM-03-2024-0128

Moura, L.R. and Kohl, H. (2020), "Maturity assessment in Industry 4.0 – A comparative analysis of Brazilian and German companies", Emerging Science Journal, Vol. 4 No. 5, pp. 365-375. https://doi.org/10.28991/esj-2020-01237.

Ouzzani, M., Hammady, H., Fedorowicz, Z. and Elmagarmid, A. (2016), "Rayyan: a web and mobile app for systematic reviews", Systematic Reviews, Vol. 5, Article 210.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P. and Moher, D. (2021), "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews", *BMJ*, Vol. 372, n71, doi: 10.1136/bmj.n71.

Pech, M. and Vaněček, D. (2022), "Barriers of new technology introduction and disadvantages of Industry 4.0 for industrial enterprises", *Serbian Journal of Management*, Vol. 17 No. 1, pp. 197–206, doi: 10.5937/sjm17-30453.

Prim, M. F.; Gomes, J. d. O.; Kohl, H.; Orth, R.; Will, M.; Vargas, G. B. "Identifying the Dynamics of Intangible Resources for Industry 4.0 Adoption Process." IEEE Access, v. 10, p. 101029–101041, 2022. https://doi.org/10.1109/ACCESS.2022.3208250

Puchan, J., Zeifang, A. and Leu, J.D. (2018), "Industry 4.0 in practice – identification of industry 4.0 success patterns", in Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE, pp. 1091-1095. https://doi.org/10.1109/IEEM.2018.8607745.

Ramanathan, K. and Samaranayake, P. (2022), "Assessing Industry 4.0 readiness in manufacturing: a self-diagnostic framework and an illustrative case study", Journal of Manufacturing Technology Management, Vol. 33 No. 3, pp. 468-488. https://doi.org/10.1108/JMTM-09-2021-0339.

Ramos, L.F.P., Loures, E.F.R., Deschamps, F., Venâncio, A.L.A.C. and Leal, G.S.S. (2023), "Assessment framework for Proof of Concept (PoC) in Industry 4.0 – an interoperability approach", International Journal of Computer Integrated Manufacturing, Vol. 36 No. 10, pp. 1482-1506. https://doi.org/10.1080/0951192X.2023.2189310.

Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P. and Harnisch, M. (2015), *Industry 4.0: the future of productivity and growth in manufacturing industries*, Boston Consulting Group. https://web-assets.bcg.com/img-src/Industry_40_Future_of_Productivity_April_2015_tcm9-61694.pdf

Sadeghi-Niaraki, A. (2020), "Industry 4.0 development multi-criteria assessment: an integrated fuzzy DEMATEL, ANP and VIKOR methodology", IEEE Access, Vol. 8, pp. 23689-23704. https://doi.org/10.1109/ACCESS.2020.2965979.

Sajjad, A., Ahmad, W., Hussain, S., Chuddher, B.A., Sajid, M., Jahanjaib, M., Ali, M.K. and Jawad, M. (2024), "Assessment by lean modified manufacturing maturity model for Industry 4.0: a case study of Pakistan's manufacturing sector", IEEE Transactions on Engineering Management, Vol. 71, pp. 6420-6434. https://doi.org/10.1109/TEM.2023.3259005.

Santos, R. C. and Martinho, J. L. (2020), "An Industry 4.0 maturity model proposal", Journal of Manufacturing Technology Management, Vol. 31 No. 5, pp. 1023-1043. https://doi.org/10.1108/JMTM-09-2018-0284.

Schumacher, A., Erol, S. and Sihn, W. (2016), "A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises", Procedia CIRP, Vol. 52, pp. 161-166. https://doi.org/10.1016/j.procir.2016.07.040.

Shabur, M. A. (2024). "A comprehensive review on the impact of Industry 4.0 on the development of a sustainable environment". Discover Sustainability, 5, 97. Springer. https://doi.org/10.1007/s43621-024-00290-7.

Silva, V.L., Kovaleski, J.L., Pagani, R.N. and Gomes, M.A.S. (2024), "Technology transfer model oriented to industry 4.0: proposal and application in the food industry", Brazilian Journal of Operations and Production Management, Vol. 21 No. 2, e20241777. https://doi.org/10.14488/BJOPM.1777.2024.

Singh, V.K., Singh, P., Karmakar, M., Leta, J. and Mayr, P. (2021), "The journal coverage of Web of Science, Scopus and Dimensions: a comparative analysis", Scientometrics, Vol. 126, pp. 5113–5142. https://doi.org/10.1007/s11192-021-03948-5.

Tripathi, S. and Gupta, M. (2021), "A holistic model for Global Industry 4.0 readiness assessment", Benchmarking: An International Journal, Vol. 28 No. 10, pp. 3006-3039. https://doi.org/10.1108/BIJ-07-2020-0354.

Ünlü, H., Demirörs, O. and Garousi, V. (2023), "Readiness and maturity models for Industry 4.0: A systematic literature review", Journal of Software: Evolution and Process, e2641. https://doi.org/10.1002/smr.2641.

Vaidya, S., Ambad, P. and Bhosle, S. (2018), "Industry 4.0 – A Glimpse", Procedia Manufacturing, Vol. 20, pp. 233-238. https://doi.org/10.1016/j.promfg.2018.02.034.

Vasconcellos, L. H. R., Gobo Junior, P. and Rodrigues, F. (2021), "An Industry 4.0 Maturity Model Applied to the Automotive Supply Chain", Revista Gestão & Tecnologia, v. 21, n. 4, p. 230–258. https://doi.org/10.20397/2177-6652/2021.v21i4.2189.

Wagire, A.A., Joshi, R., Rathore, A.P.S. and Jain, R. (2021), "Development of maturity model for assessing the implementation of Industry 4.0: learning from theory and practice", Production Planning & Control, Vol. 32 No. 8, pp. 603–622, doi: 10.1080/09537287.2020.1744763.

Wankhede, V.A. and Vinodh, S. (2023), "Benchmarking Industry 4.0 readiness evaluation using fuzzy approaches", Benchmarking: An International Journal, Vol. 30 No. 1, pp. 281-306. https://doi.org/10.1108/BIJ-08-2021-0505

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Integrating Cost Management into Maturity Models for Public Healthcare Systems: A Framework of Dimensions and Competencies

Tiago Rocha [1], Naja Santana [2] and Larissa Araújo [3]

[1] Federal University of São Carlos – UFSCar (Brazil); tiagosoares@estudante.ufscar.br
[2] Federal University of São Carlos – UFSCar (Brazil); naja@ufscar.br
[3] Federal University of São Carlos – UFSCar (Brazil); laaraujo@ufscar.br

***Abstract*— This study proposes a conceptual framework that integrates cost management into maturity models for public healthcare systems. Although maturity models are widely used in the healthcare sector to support process improvement and organizational development, they rarely incorporate cost management as a central dimension. Based on the technical literature on the addressed topics and on seminal studies that develop maturity models, this study identifies key dimensions and competencies that contribute to increased transparency, control, and decision-making capacity related to costs. The proposed framework addresses a relevant gap by integrating the cost management perspective into maturity assessments. The findings contribute to the advancement of maturity models and offer practical implications for healthcare managers operating in resource-constrained environments.**

Keywords – Cost Management; Maturity Models; Public Healthcare; Operations Management

I. Introduction

Considering the Brazilian policy context, in which resources allocated to the healthcare sector are becoming increasingly scarce, the commitment of leaders - whether in public, private, or philanthropic institutions, across different levels of operation - to organizational management models, cost management, and their impacts on the feasibility of different work processes is crucial for the delivery of safe, efficient, effective, and economically sustainable healthcare services (Brancalion and Lima, 2022). Currently, there is a shortage of financial resources in the hospital environment. The pursuit of management excellence is becoming increasingly prominent and growing (Ceretta et al., 2023).

In this context, modern healthcare management and practice rely heavily on data and scientific evidence to ensure that hospital management is both scientific and refined (Jiang et al., 2022). As an alternative, Gomes et al. (2019) define the mapping and costing (budgeting) of internal investment projects as fundamental to service success and quality, since multiple steps must be undertaken before regular and centralized cost information in healthcare services can be obtained.

Emerging in recent decades, New Public Management (NPM) represents a paradigm shift in public sector management. NPM advocates, among other practices, the adoption of managerial techniques from the private sector into the public sector, aiming to enhance efficiency and productivity, and, in some cases, to reduce the direct involvement of the state in activities previously under its responsibility. Supported by an optimistic discourse, governments across different regions of the world have sought to implement NPM principles in order to improve public sector performance (Blonski et al., 2017).

One of the alternatives for innovating organizational management is the use of maturity models which, according to Shaygan and Daim (2019), are organizational management tools that have been developed and applied for decades as a means for organizations to respond to the constant pressure to achieve and sustain competitive advantage through simultaneous innovation, quality improvement, and cost reduction. Managers in the healthcare sector have not been an exception in reaping the benefits and addressing the weaknesses of strategies through the systematic quality improvement provided by maturity models. However, the existing studies identified for this sector still lack a specific focus on cost management or any other approach that explicitly addresses the effective and efficient management of financial resources.

This study aims to address this gap by proposing a conceptual framework that integrates cost management into maturity models for public healthcare systems. Specifically, it identifies key dimensions and competencies that support the development of cost-related capabilities. The study contributes both theoretically and practically by enhancing the applicability of maturity models in resource-constrained environments.

II. Literature Review

A. Cost Management in Public Healthcare Systems

The evolution of management accounting has enabled a shift from a traditional role focused on cost determination and financial control to a more sophisticated role centered on value creation for organizations. New techniques have been developed to support modern technologies and emerging management processes, aiming to achieve competitive advantage in response to the challenges of global competition (Abdel-Kader and Luther, 2008).

According to Rehman et al. (2021), cost management plays a crucial role in both strategic and operational decision-making by providing insights into the profitability of products, services, and projects. Accurate cost information enables improved resource allocation, pricing strategies, and performance evaluation. Furthermore, understanding cost drivers supports the identification of opportunities for process improvement and cost reduction. In the healthcare context, for instance, cost management informs decisions related to resource allocation, treatment protocols, and investments in new technologies (Cortesi et al., 2015). Cardoso (2020) further argues that, in addition to being challenged by the uncertainty associated with healthcare demand and service outcomes, healthcare managers must also operate under constrained budgets and limited resources, while managing a large number of variables (e.g., scheduling, staffing, bed capacity, among others). As a result, healthcare managers are continuously evaluating the effectiveness of existing systems and exploring opportunities for improvement.

Recent research in healthcare cost management highlights several relevant themes. These include the pursuit of more integrated and patient-centered care models aimed at reducing hospitalizations and improving health outcomes; the assessment of the impact of health policies on costs and quality of care, with studies evaluating the effectiveness of different interventions and the need for policy adjustments to optimize outcomes; the growing interest in digital health technologies and value-based payment models; and the need to develop quality and efficiency indicators to monitor intervention outcomes and support cost management strategies (Wilson et al., 2020). In addition, the management of chronic diseases—representing a significant share of healthcare expenditures—has emerged as a key area of focus (Watson and Wilkinson, 2021).

Regarding cost management in public systems, it is strongly influenced by the principles of New Public Management. Ditterich et al. (2012) emphasize that the pursuit of greater efficiency and cost reduction is one of its central objectives. Akbar et al. (2022) observe that governments have adopted enterprise resource planning (ERP) systems and business process management (BPM) systems as mechanisms to reduce costs and enhance productivity, in alignment with New Public Management principles. In this context, maturity models represent a viable approach for supporting such initiatives. According to Cleven et al. (2014), maturity models assist in assessing process management maturity across dimensions such as organizational culture, strategy, structures, practices, and information technology. Across multiple sectors—including healthcare management—these models support the identification of improvement opportunities in organizational processes.

One of the tools to be introduced in this process of seeking greater operational efficiency is the use of maturity models. Their application in healthcare management, according to Gomes and Romão (2018), aims to optimize processes, improve service quality, and reduce costs.

B. Maturity Models

To explain the purpose of maturity models, Röglinger et al. (2012) state that a maturity model is a framework that provides organizations with a systematic approach to measure and assess their current level of capability and performance in each domain. Furthermore, maturity models are used to support the development of strategies aimed at progressing toward specific objectives.

Regarding the development of maturity models, Santos et al. (2020) emphasize that their core lies in well-defined dimensions and competencies. Dimensions represent the main areas of focus, while competencies describe the specific capabilities required within each dimension to achieve a certain maturity level. The selection of dimensions and competencies is highly dependent on the specific context in which the model is applied.

Maturity is commonly assessed across distinct stages and multiple dimensions. Most maturity models adopt a five-level structure, following the original concept proposed by Crosby (1984) and later popularized by the Capability Maturity Model for Software (CMM-SW) (Paulk et al., 1993). These approaches are based on the premise that improved processes lead to better outcomes; in other words, they assume that increasing maturity in project management results in more consistent performance and higher project success rates. Maturity models typically comprise a sequence of levels that define a logical progression from an initial state to a mature stage.

A major challenge associated with these models is that, for certain domains in which organizations operate, ready-made maturity models are not available. As a result, it becomes necessary to conduct research and adapt maturity models from other domains, considering the most relevant variables to ensure that the results accurately reflect the organizational reality (Cleven et al., 2014). Additionally, Gomes and Romão (2018) argue that, although research efforts have increasingly focused on developing more rigorous approaches for maturity model design, there has been a general lack of attention to the effective implementation of these models in complex organizational environments, such as hospitals.

III. METHODOLOGY

This research adopts a qualitative and exploratory approach based on a structured literature review. The objective was to identify how cost management has been addressed in the context of maturity models and to derive relevant dimensions and competencies.

The research process involved three main stages: (1) identification of relevant literature, (2) analysis and synthesis of key concepts, and (3) development of a conceptual framework. The analysis focused on identifying recurring themes and gaps related to cost management within maturity models.

Based on this synthesis, a set of dimensions and competencies was developed. These elements form the basis of the proposed framework, which aims to integrate cost management into maturity assessments in public healthcare systems.

Among the main structural references, the study by De Bruin et al. (2005) stands out, as it presented consolidated methodological approaches for the development of Maturity Models and proposed a development framework applicable across multiple domains. The concepts and definitions explored, synthesized in the Literature Review of this study, were considered in the construction of the competencies and dimensions to be presented subsequently.

IV. Proposed Dimensions and Competences

The proposed framework is structured around four key dimensions, each associated with specific competencies that support cost management capabilities.

A. Cost Planning and Strategy

The Cost Planning and Strategy dimension addresses the existence of formal strategies for cost management, the alignment of these strategies with organizational objectives, and the use of planning tools to forecast and control costs. The definition of precise goals and indicators to assess performance is also encompassed within this dimension (Abdel-Kader and Luther, 2008). For this dimension, it is important to consider the following competencies:

- Define cost reduction objectives aligned with the organizational strategy.
- Develop detailed budgeting plans and control mechanisms.
- Use break-even analysis tools and other cost forecasting techniques.

B. Cost Management Process

The Cost Management Processes dimension focuses on the formalization, standardization, and control of processes related to the identification, measurement, analysis, and interpretation of costs. It includes the adoption of costing methodologies such as full costing and Activity-Based Costing (ABC) to improve accuracy in cost allocation (Rehman et al., 2021). For this dimension, the following competencies are highlighted:

- Formalize the managerial function and designate personnel responsible for cost management.
- Implement accurate and comprehensive cost data collection and recording systems.
- Apply costing methodologies appropriate to the complexity of services (e.g., ABC).
- Establish processes for cost control across different areas of the organization.

C. Technology and Information Systems for Cost Management

The Technology and Information Systems for Cost Management dimension assesses the use of information systems and technologies to support the collection, analysis, and dissemination of cost information. This involves the ability of systems to integrate data, generate reports, and provide insights for decision-making. The overall digital maturity of the organization is also relevant (Santos et al., 2020). The competencies highlighted for this dimension are as follows:

- Implement and integrate information systems to track and analyze costs.
- Use data analytics and reporting tools to generate cost-related insights.
- Ensure the quality and integrity of cost data.

D. Culture and Training in Cost Management

The Culture and Training in Cost Management dimension examines the awareness and engagement of healthcare and administrative professionals in cost management, as well as the existence of training programs and incentives aimed at promoting efficiency and resource optimization. The focus on effectiveness and efficacy in decision-making, along with the pursuit of a continuous improvement culture that can lead to cost optimization, is essential. Establishing an organizational culture that values cost management and investing in employee training programs are fundamental to ensuring the success of cost management initiatives (Gökalp and Martinez, 2022).

The study by Vahedi et al. (2024) demonstrates that competent leadership has a positive and significant correlation with organizational culture, particularly in hospital settings, highlighting that the synergy between strong leadership and a robust organizational culture enhances the quality of healthcare services and improves operational outcomes, including financial and cost management. Leadership fosters a collaborative work environment and empowers professionals, facilitating the adoption of best practices and appropriate methodologies for hospital cost management. Regarding the competencies to be considered in this dimension, the following are presented:

- Promote, through organizational leaders and responsible departments, awareness of the importance of cost management among employees, particularly in operations and cost management areas.
- Provide, through leaders responsible for cost management, training and capacity building in cost management tools and techniques.
- Encourage the identification of opportunities for efficiency improvement and waste reduction by leaders and all individuals involved in the processes.

E. Performance Analysis and Cost-Based Decision-Making

Finally, the Performance Analysis and Cost-Based Decision-Making dimension assesses the organization's ability to analyse cost data, compare performance against targets and benchmarks, and use this information to support more informed strategic and operational decisions. This includes the use of quality and efficiency indicators (Sepasgozar et al., 2022). For this dimension, the following competencies are considered:

- Define and monitor cost performance indicators (Key Process Indicators – KPIs).
- Conduct variance analyses and identify the causes of deviations from the budget.
- Use cost information to assess the feasibility of new services and investments.

V. Development of Maturity Levels

Some studies, such as De Bruin et al. (2005), provide guidelines for the development of maturity models, whether domain-specific or more general in nature. However, in general, it is observed that there are no specific definitions or consensus, and the authors consistently emphasize the importance of drawing on multiple sources of evidence to define the steps and guidelines to be followed.

The levels presented in Table I aim to assess the previously defined dimensions and were established based on the work of De Bruin, Freezeet al. (2005), which proposes an applicable development framework consistent with those found in other studies, such as Gomes and Romão (2018) and Cleven et al. (2014).

TABLE I. Maturity Model Levels for Cost Management in Healthcare Systems

	Development of Levels
Level 0	Basic process and documentation required for necessary analyses are not non-existent.
Level 1: Initial	Processes or actions are reactive and unstructured; Lack of formal processes; Documentation is basic and frequently inaccurate.
Level 2: Developing	Basic processes are established but remain inconsistent; Some specific practices exist but are applied irregularly; Documentation is generated regularly, but analysis is limited.
Level 3: Defined	Processes are well defined and clearly understood by stakeholders; Practices are reviewed and periodically adjusted; Documentation is regularly verified to identify areas for improvement.
Level 4: Managed	Processes are actively monitored and managed; Well-established indicators are in place for control and evaluation; Integration between decision-making processes and strategic planning is evident.
Level 5: Optimized	Processes are continuously improved based on data and advanced analytics; Indicators are systematically used for control, evaluation, and efficiency; Advanced tools and analytical techniques are employed for optimization.

It is observed that the levels begin with the absence of processes, even basic ones, that could support the development of each identified dimension. As these processes evolve, the assigned level increases, with Level 5 (Optimized) representing the stage in which processes and data are used in the most efficient manner, being continuously improved and supported using advanced analytical techniques.

VI. Development of Cost Management Dimensions Across Maturity Levels

To achieve the objective of this study—namely, to develop a conceptual framework that integrates cost management into maturity models for public healthcare systems—reference studies were identified to support the definition of the conceptual foundations presented and the development of dimensions across maturity levels. One such study, Williams et al. (2019), presents the development of dimensions across maturity levels and serves as a reference for constructing the proposed model. The authors develop an action map demonstrating how the main involved areas evolve through their activities and objectives, providing a basis for defining maturity levels and classifying dimensions within these levels.

For the development of competencies within each dimension across maturity levels, the structure initially proposed by Paulk et al. (1993), and widely adopted in models across different fields, was used as a reference. The structure is also grounded in Campos et al. (2024), who based their work on Gearin et al. (2013), which in turn is founded on the QI Roadmap (Roadmap to a Culture of Quality Improvement), a framework endorsed by the National Association of County and City Health Officials (NACCHO). This roadmap is recognized as an easily applicable model that can be disseminated across hospitals and associated with a structured approach for developing quality improvement strategies. The structure is also suitable for incorporating cost management concepts, thereby forming the foundation of the model proposed in this study.

This development constitutes the main contribution of this study, as presented below according to each identified dimension, in Tables II, III, IV, V and VI.

TABLE II. Development of the Cost Planning and Strategy Dimension Across Maturity Levels

Cost Planning and Strategy	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
Define cost reduction objectives aligned with the organizational strategy.	Cost reduction objectives are not defined.	Cost reduction objectives exist but are not formalized.	Cost reduction objectives are formalized, but there is no evidence that they are aligned with the organizational strategy.	Cost reduction objectives are formalized, understood, and aligned with the organizational strategy.	Cost reduction objectives are formalized and actively managed to ensure continuous alignment with the organizational strategy.	Cost reduction objectives are continuously reviewed based on advanced analyses to support the organizational strategy.
Develop detailed budgeting plans and control mechanisms.	No budgeting plans exist.	Budgeting plans are not detailed and do not include control mechanisms.	Budgeting plans have the necessary level of detail but do not include control mechanisms.	Budgeting plans are detailed, and control mechanisms are established.	Detailed budgeting plans are monitored, controlled, and actively managed.	Detailed budgeting plans are controlled, continuously managed, and reviewed based on advanced analyses and decision-support tools.
Use break-even analysis tools and other forecasting techniques.	No cost forecasting techniques are utilized.	Cost forecasting is discussed, but no tools are utilized	Limited use of forecasting techniques, which are generally not formalized or fully understood.	Break-even analysis tools and other forecasting techniques are used for cost management.	Break-even analysis tools and other forecasting techniques are formalized, well understood, and supported by clearly defined analysis procedures.	Break-even analysis tools and other forecasting techniques are continuously managed and refined to enhance performance.

TABLE III. DEVELOPMENT OF THE COST MANAGEMENT PROCESSES DIMENSION ACROSS MATURITY LEVELS

Cost Management Process	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
Formalize the managerial function and designate personnel responsible for cost management.	No costing methodologies are applied.	There are individuals handling cost data operationally, but without generating outcomes for management.	There are individuals responsible for cost management within other management areas of the organization.	There is a dedicated cost management unit with personnel assigned to cost data collection and recording.	There is a cost management unit with dedicated personnel, formalized processes, and cost data collection and reporting results communicated to top management.	There is a cost management unit whose personnel actively participate in the organization's strategic decision-making, using generated cost information and analyses.
Implement accurate and comprehensive cost data collection and recording systems.	There are no processes for cost control in any area of the organization.	The process of collecting and recording cost data is not formalized.	The system for cost data collection and recording is formalized, but with limited scope and detail.	The cost data collection and recording system has a scope and level of detail consistent with organizational processes.	The cost data collection and recording system supports organizational processes and is actively managed.	The cost data collection and recording system is actively managed, employing advanced techniques and continuous improvement practices.
Apply costing methodologies appropriate to the complexity of services (e.g., Activity-Based Costing – ABC).	There is no cost management function or personnel assigned to this role.	The application of costing methodologies is informal and unstructured.	The application of costing methodologies is structured; however, it does not fully address the complexity of services.	The application of costing methodologies is appropriate to the complexity of services.	The application of costing methodologies is aligned with the complexity of services, which are monitored using well-established indicators.	The application of costing methodologies fully addresses the complexity of services, ensuring control through well-established indicators supported by advanced tools.
Establish processes for cost control across different areas of the organization.	There is no formalized system for cost data collection and recording.	Cost control processes are defined, but not structured across all areas of the organization.	Cost control processes are structured at a basic level across different areas of the organization.	Cost control processes are structured and understood across different areas of the organization, but are not actively monitored.	Cost control processes across organizational areas are actively monitored and integrated into the decision-making process.	Cost control processes across different areas of the organization are continuously improved and make use of advanced techniques.

TABLE IV. DEVELOPMENT OF THETECHNOLOGY AND INFORMATION SYSTEMS FOR COST MANAGEMENT DIMENSION ACROSS MATURITY LEVELS

Technology and Information Systems for Cost Management	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
Implement and integrate information systems to track and analyze costs.	There are no information systems to track and analyze costs.	Information systems are basic, inaccurate, and used infrequently.	Information systems are used regularly; however, they generate basic and imprecise information.	Information systems are understood by all stakeholders involved in the processes, ensuring integration.	Information systems are actively managed, integrating processes, including alignment with strategic planning.	Information systems ensure integration and are continuously updated and improved in their functionalities.

Use data analytics and reporting tools to enhance cost-related insights.	There are no data analytics or reporting tools providing cost-related information.	Data analytics and cost reporting tools are rarely used and applied informally.	Data analytics and cost reporting tools are basic and imprecise.	Data analytics and cost reporting tools are defined and understood by stakeholders, providing useful information.	Data analytics and cost reporting tools are actively managed and provide insights, resulting in well-established information.	Data analytics and cost reporting tools are continuously updated and improved, resulting in precise information.
Ensure the quality and integrity of cost data.	There is no assurance of the quality and integrity of cost data.	The assurance of cost data quality and integrity is not formalized.	Cost data quality and integrity are ensured through basic and inconsistent practices.	The quality and integrity of cost data are ensured through practices that are understood and executed by those involved.	The assurance of cost data quality and integrity is well established, formalized, and actively managed by a designated responsible party.	The quality and integrity of cost data are ensured through consolidated practices that support continuous improvement.

TABLE V. DEVELOPMENT OF THE CULTURE AND TRAINING IN COST MANAGEMEN DIMENSION ACROSS MATURITY LEVELS

Culture and Training in Cost Management	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
Promote awareness, through organizational leaders and responsible departments, of the importance of cost management among personnel involved in operations.	No promotion of awareness regarding the importance of cost management.	Awareness of the importance of cost management is observed empirically.	Awareness is promoted in alignment with organizational management.	The importance of cost management is communicated to operational staff in a structured manner.	Operational personnel are familiar with cost management processes and related established indicators.	Operational personnel adopt processes and practices that contribute to improving cost management indicators and proactively propose advancements.
Provide, through leaders responsible for cost management, training and capacity building in cost management tools and techniques for personnel involved in operations and cost management.	No training or capacity building in cost management tools and techniques is provided.	Training is offered for specific demands, in a non-systematic and non-goal-oriented manner.	Training is provided for broader demands, aligned with management objectives.	There is a regular provision of training in cost management tools and techniques, supporting the design of processes and activities.	Organizational management ensures training in cost management for operations and cost management areas, supporting continuous improvement initiatives.	Training ensures that top management decisions are supported by high-level analyses developed by personnel capable of driving innovation in operations and cost management processes.
Encourage the identification of opportunities for efficiency improvement and waste reduction by leaders and all individuals involved in processes.	No encouragement for identifying opportunities for efficiency improvement and waste reduction.	Employees are encouraged to identify opportunities for efficiency improvement and waste reduction.	Encouragement is fostered through training programs.	There is a continuous process for identifying opportunities, with defined indicators.	The use of advanced continuous improvement tools promotes efficiency gains and waste reduction, supported by well-established and actively monitored indicators.	Identified opportunities are translated into action plans, and their impacts are measured and may be incentivized proportionally, contributing to the organization's strategic objectives.

TABLE VI. DEVELOPMENT OF THE PERFORMANCE ANALYSIS AND COST-BASED DECISION-MAKING DIMENSION ACROSS MATURITY LEVELS

Performance Analysis and Cost-Based Decision-Making	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
Define and Monitor Cost Performance Indicators (KPIs)	No cost performance indicators (KPIs) are defined.	Cost performance indicators (KPIs) are defined without alignment with the management process.	Cost KPIs are simplistically formulated and used only within the respective functional department.	Cost indicators are established in coordination with organizational management and are formally monitored.	Cost indicators are formally monitored and continuously improved based on operational enhancements.	Cost performance indicators (KPIs) are more sophisticated and integrated with other managerial indicators, serving as key support for decision-making.
Perform Variance Analysis and Identify the Causes of Budget Deviations	No variance analysis is conducted to identify the causes of budget deviations.	Budget analysis is performed only during closing and planning periods by responsible personnel, preventing timely identification of deviations.	Periodic monitoring of budget deviations is conducted by the responsible department, which performs reactive variance analysis.	Budget deviations are promptly identified by the responsible department, enabling technical variance analysis.	Budget execution is monitored according to management-defined intervals, ensuring that any variance is identified and promptly analyzed, including the involvement of top management, thereby improving understanding of deviation causes.	Budget execution is monitored in real time by management, enabling early detection of variance trends and the implementation of action plans to enhance system efficiency and mitigate known

						causes of deviations.
Use Cost Information to Evaluate the Feasibility of New Services and Investments	Cost information is not used to evaluate the feasibility of new services and investments.	Feasibility assessments rely on generalized cost estimates and imprecise data.	Feasibility analysis is based on cost data formally established by the cost management function.	Forecasting techniques, supported by well-managed cost information repositories, are used to assess the feasibility of new services and investments.	Cost information, supported by well-established performance indicators, is used for strategic decision-making and feasibility analysis of new services and investments.	Cost information is combined with advanced foresight techniques to anticipate future scenarios and trends in new services and investments.

VII. Discussion

Based on the previous sections, the evolution of the dimensions across maturity levels can be established. It is important to emphasize that, after classifying the system under analysis at a given current level, the target level of evolution (i.e., the level to be achieved) should be defined according to the organization's capabilities and needs.

Although Level 5 is generally regarded as the most advanced stage, representing the highest level of process efficiency, this does not necessarily imply that organizations should aim to achieve it. For instance, if the resources required to reach Level 5 are not expected to yield proportional benefits, the organization may instead define Level 3 or Level 4 as its target. Similarly, in cases of low organizational complexity, the adoption of advanced techniques or tools for cost measurement and analysis may not be justified, making a maturity level below Level 5 more appropriate.

Additionally, a gradual evolution strategy may be adopted, planned in accordance with the current resource availability of each organization. The maturity evolution process itself may require progressive development of resources of different natures (e.g., capital, human resources, equipment, and information systems), such that a previously defined future state becomes the current state, enabling further advancement to be planned.

It is also important to highlight that, to effectively designate the framework developed in this section as a "Maturity Model," a validation process is required. This process will rely on structured assessment instruments, potentially derived from the references adopted as the foundation of this study.

VIII. Conclusion

Considering the complexity of the healthcare sector and the need for its management to be both scientific and refined, and considering maturity models as an effective tool for this purpose, it was identified that existing studies do not specifically address cost and financial resource management. In this context, the present study aimed to propose a conceptual framework that integrates cost management into maturity models for public healthcare systems.

The literature review conducted in this study provided a synthesis of the works used as a foundation for the definitions and concepts that supported the identification of dimensions and maturity levels. Emphasis is given to studies focused on the development of maturity models, which served as key references guiding the approach adopted by the authors.

In integrating Cost Management into the concept of Maturity Models for Public Healthcare Systems, the competencies defined within each dimension were structured across maturity levels, indicating how they can be classified and how they may evolve throughout these levels. This development represents the main contribution of the study, aiming to improve processes related to Cost Management in healthcare systems. Therefore, the proposed objective is considered to have been achieved.

This research expands the scope of Cost Management studies by emphasizing its interface with Operations Management, incorporating maturity model developments as a relevant managerial tool. Thus, the study contributes to both Cost Management and Operations Management fields, reinforcing the importance of this interdisciplinary perspective.

Through the framework proposed in this paper, it is expected to contribute to the development of public policies in healthcare system management, aiming to increase productivity while ensuring the required level of efficiency.

As highlighted in the references adopted in this study, the validation process is an essential step for consolidating the proposed framework as a Maturity Model for Cost Management in Public Healthcare Systems, and it is therefore identified as the main direction for future research. Once validated and consolidated, the model may also be adapted for application in private sector organizations. In this regard, New Public Management principles provide relevant guidelines transferable from the private to the public sector, such as management professionalization, workforce training, user satisfaction, and organizational learning—representing one avenue for further investigation.

Another possible extension, to be explored after the consolidation of the model, is its application in other sectors, given the importance of cost management across different types of organizations, whether public or private.

Regarding the limitations of this study, the lack of prior research specifically addressing cost management within maturity models is highlighted. While this gap provided an opportunity for the development of this research, it also limited the availability

of references aligned with the specificity of the proposed model. Future studies, with sufficient time for validation and application in the target organizations, may further consolidate the model, thereby enhancing its contribution to the literature.

References

Abdel-Kader, M. and Luther, R. (2008), "The impact of firm characteristics on management accounting practices: A UK-based empirical analysis", The British Accounting Review, Vol. 40 No. 1, pp. 2-27.

Akbar, M. et al. (2022), "Towards roadmap to implement blockchain in healthcare systems based on a maturity model", Journal of Software: Evolution and Process, Vol. 34 No. 12, pp. 1-15.

Blonski, F. et al. (2017), "O Controle Gerencial na Perspectiva do New Public Management: O Caso da Adoção do Balanced Scorecard na Receita Federal do Brasil", Administração Pública e Gestão Social, Vol. 9 No.1, pp. 15–30

Brancalion, F. and Lima, A. C. (2022), "Process-based Management aimed at improving health care and financial results", Revista da Escola de Enfermagem da USP, Vol. 56 No. 20210333.

Campos, S, Gomes, E. and Grande, M. (2024), "Uso de modelo de maturidade para melhoria da qualidade de hospitais de média complexidade", Gestão em Organizações de Saúde, Vol. 57 No. 2, pp. 1-7.

Cardoso, W. (2020), "O mapeamento do fluxo de valor como ferramenta do Lean Healthcare para enxergar desperdícios e pontos de melhoria: o caso do pronto atendimento de um hospital universitário", Revista de Gestão em Sistemas de Saúde, Vol. 9 No. 2, pp. 360-380.

Ceretta , J., Seibert , R. and Callegaro, A. R. (2023), "Gestão hospitalar: a auditoria operacional como ferramenta estratégica para o controle de desperdícios", Revista De Gestão E Secretariado, Vol.14 No. 3, pp. 2663–2675.

Cleven, A., Winter, R., Wortmann, F. and Mettler, T. (2014), "Process management in hospitals: an empirically grounded maturity model", Business & Information Systems Engineering, Vol. 7 No. 2, pp. 191-216.

Cortesi, P. et al. (2015), "Management of treatment-naïve chronic hepatitis C genotype 1 patients: a cost-effectiveness analysis of treatment options", Journal of Viral Hepatitis, Vol. 22 No. 2, pp. 175-183.

Crosby, P. (1984), Qualidade é investimento: a arte de garantir a qualidade, José Olympio, Rio De janeiro, RJ, Brazil.

De Bruin, T., Freeze, R., Kulkarni, U. and Rosemann, M. (2005), "Understanding the Main Phases of Developing a Maturity Assessment Model", in 16th Australasian Conference on Information Systems 2005, proceedings of the 16th Australasian Conference on Information Systems, Sydney, 2005, Association for Information Systems.

Ditterich, R., Moyses, S.and Moyses, S. (2012), "O uso de contratos de gestão e incentivos profissionais no setor público de saúde", Caderno de Saúde Pública, Vol. 28 No.4, pp. 615-627.

Gearin, Miner et al.(2023), "Monitoring QI Maturity of Public Health Organizations and

Systems in Minnesota: Promising Early Findings and Suggested Next Steps", Frontiers in Public Health Services& Systems Research, Vol. 2 No. 3.

Gomes, C., De Liberal, M. and Rached, C. (2019), "Cost management in a multi- professional small-scale clinic of popular health services", International Journal of Entrepreneurship, Vol. 23 No. 2, pp. 1-13.

Gomes, J. and Romão, M. (2018), "Information System Maturity Models in Healthcare", Journal of Medical Systems, Vol. 42 No. 12, p. 235.

Gökalp, E. and Martinez, V. (2022), "Digital transformation maturity assessment: development of the digital transformation capability maturity model", International Journal of Production Research, Vol. 60, No. 20.

Jiang, Q. et al. (2022), "Research on Time-Driven Activity-

Based Management System of Public Hospitals", Frontiers in Public Health, Vol. 9 No. 27.

Paulk, M. et al. (1993), "Capability maturity model, version 1.1", IEE Software, Vol. 10 No. 4, pp. 18-27.

Rehman, S. et al. (2021), "MCS package and entrepreneurial competency influence on business performance: the moderating role of business strategy", European Journal of Management and Business Economics, Vol. 32 No. 1, pp. 1-23

Röglinger, M., Pöppelbuß, J. and Becker, J. (2012), "Maturity models in business process management", Business Process Management Journal, Vol. 18 No. 2, pp. 328-346.

Santos, D. et al. "Proposal for a Maturity Model in Sustainability in the Supply Chain", Sustainability, Vol. 12 No. 22, pp. 1-37

Sepasgozar, S. et al. (2022), "BIM and Digital Tools for State-of-the-Art Construction Cost Management", Buildings, Vol. 12 No. 4, pp. 1-28.

Shaygan, A. and Daim, T. (2019), "Technology Management Maturity Assessment Model: An Exploratory Multi-Criteria Approach for Healthcare Organizations", in Portland International Conference on Management of Engineering and Technology (PICMET) 2019, proceedings of the Portland International Conference on Management of Engineering and Technology (PICMET), Institute of Eletrical and Eletronics Engineers, Portland, USA, 2019.

Vahedi, H., Shivyari, S., Soltanmohamadi, S. and Babakhanian, F. (2024), "The Role of Leadership Competence and Organizational Culture in Enhancing Healthcare Service Quality", Journal of Archives in Military Medicine, Vol. 12, No. 3.

Watson, A. and Wilkinson, T. (2022), "Digital healthcare in COPD management: a narrative review on the advantages, pitfalls, and need for further research", Therapeutic Advances in Respiratory Disease, Vol. 16 pp. 1-23.

Williams, P. et al. (2019), "Improving Digital Hospital Transformation: Development of an Outcomes-Based Infrastructure Maturity Assessment Framework", Journal of Medical Internet Research, Vol. 7, No. 1.

Wilson, M. et al. (2020), "The impacts of accountable care organizations on patient experience, health outcomes and costs: a rapid review", Journal of Health Services Research & Policy, Vol. 25 No. 2, pp. 130-138.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Where to work outside the home: mapping coworking spaces and evaluating their quality in the context of remote work

Lorena BASTOS [1], Marlene AMORIM [1] , Mário RODRIGUES [2] and Cláudia SILVA [3]

[1] GOVCOPP & DEGEIT– Governance Competitiveness and Public Policies & Department of Economics, Management, Industrial Engineering and Tourism, University of Aveiro; lorenabbastos@ua.pt; mamorim@ua.pt

[2] IEETA & ESTGA - Institute of Electronics and Informatics Engineering of Aveiro & Águeda School of Technology and Management, University of Aveiro; mjfr@ua.pt

[3] GOVCOPP& ESTGA - – Governance Competitiveness and Public Policies, Águeda School of Technology and Management, University of Aveiro; claudia.margarida@ua.pt

***Abstract* — The rise of remote work has driven the expansion of coworking spaces as a hybrid alternative between the traditional office and the home office. The study analyzes the quality dimensions associated with coworking spaces in the context of remote workers, operationalizing criteria such as technological infrastructure, comfort, ergonomics, conditions for concentration, community, security, accessibility, and cost-effectiveness, based on the dimensions perceived by users, namely: connectivity, physical comfort, noise levels, availability of facilities, cleanliness, accessibility, and support services. This study aims to develop a quality assessment model for coworking spaces for remote work, using Lisbon and Porto as empirical study sites to operationalize quality dimensions, analyzing their characteristics, types, and geographical distribution, as well as the perceived quality in the context of remote work. As a result, the article presents (i) an updated overview of the coworking ecosystem in the cities of Porto and Lisbon, (ii) an analysis of the dimensions of coworking space quality in the context of remote work, and (iii) a quality assessment model applicable to decision-making by remote professionals. The study discusses the role of coworking in reshaping contemporary work practices, contributing to a better understanding of urban and labor dynamics in Portugal.**

Keywords – Coworking, Remote work, Dimensions of quality, Quality of services.

I. INTRODUCTION

Over the past decade, we have witnessed the emergence, expansion, and global spread of coworking spaces—shared workspaces where knowledge workers, working independently, come together to produce and exchange knowledge, while maintaining an individual approach within a collective context. This phenomenon accompanies broader transformations in the organization and distribution of work (Spinuzzi et al., 2019).
There has been a continuous growth in the number of independent professionals, freelancers, and remote workers; at the same time, the supply of shared and collaborative work environments, known as coworking spaces, has expanded rapidly (Clifton et al., 2022).

Coworking will be analyzed as an emerging urban social practice, associated with new ways of organizing work and fostering cooperation among freelancers and independent professionals. As flexible environments with relatively affordable costs and a strong community focus, these spaces foster encounters, interactions, and enriching exchanges among different communities, work cultures, and forms of knowledge. In this sense, they can be understood as a new form of urban social infrastructure, capable of fostering connections and collaborations among diverse people, ideas, and contexts (St George, 2015).

It has been found that freelancers tend to thrive in coworking spaces, as they find greater meaning in their work, enjoy more autonomy, and experience a sense of belonging to a community. On the other hand, they analyze the potential of these spaces for large organizations, concluding that they can stimulate the generation of new ideas, contribute to reducing facility costs, and increase employee satisfaction (Howell, 2022).

In this regard, Lisbon and Porto have been establishing themselves as key hubs in the digital and innovation ecosystem, in line with broader trends observed in European cities. The literature highlights that the urban concentration of talent, collaboration networks, and digital infrastructure plays a central role in promoting innovation (OECD, 2023). At the same time, institutional initiatives at the European level have bolstered the development of startup ecosystems and digital hubs (European Commission, 2020). In this context, cities with strong connectivity and international appeal tend to establish themselves as centers of innovation and technological entrepreneurship (Startup Genome, 2023).

However, this study expands our current understanding of coworking spaces in the context of remote work, helping to identify gaps in the literature, particularly regarding the spatial dimension of coworking spaces and the user experience in a remote work setting. By focusing on the cities of Lisbon and Porto, the study offers a contextualized and comparative perspective, thereby making a significant contribution to the understanding of contemporary work dynamics. Thus, the study seeks to answer the following research questions: (i) Which quality dimensions are most relevant for coworking spaces geared towards remote work? (ii) How can these dimensions be

operationalized in an evaluation framework? (iii) To what extent do the coworking spaces identified in Lisbon and Porto show different quality configurations?

In summary, this article seeks to contribute to a more holistic and in-depth analysis of how coworking spaces influence the quality of remote work—specifically in terms of productivity, well-being, and professional networking—thereby highlighting their emerging role in the new geographies of work. To this end, the research conducted a systematic literature review on coworking spaces in the cities of Porto and Lisbon in the context of remote work, with the aim of mapping coworking spaces in these cities and analyzing the dimensions of perceived quality in the context of remote work. Accordingly, the article is structured as follows: (1) Introduction; (2) Conceptual background of the study; (3) Data and Methods; (4) Discussion and Results; (5) Conclusion; and (6) References.

II. CONCEPTUAL BACKGROUND OF THE STUDY

A. Coworking: definitions, typologies, and value proposition

Coworking refers to a shared work arrangement in which individuals from different professional backgrounds—often independent professionals, freelancers, entrepreneurs, or remote workers—use the same workspace without necessarily working for the same organization or project. Beyond sharing physical infrastructure and resources, coworkers typically pay a membership fee to access an environment that promotes flexibility, interaction, networking, and a sense of community and social belonging (Garrett, L. E., 2017; Spinuzzi, Clay, C, 2012).

Following the 2007–2008 financial crisis and the ensuing global recession, there was a rapid expansion of collaborative and shared workspaces, known as coworking, in various cities around the world. The term coworking describes the practice of working collectively in flexible environments, where workstations can be rented on a daily, weekly, or monthly basis (St George, 2015).

The first coworking spaces, in their contemporary form, emerged between 2005 and 2006, although open and collaborative work environments with similar characteristics had existed since the early 1990s, albeit without that specific designation. Initially, these spaces were set up as shared offices intended for professionals who did not need a fixed workplace, such as freelancers and remote workers, who sought to integrate into collaborative environments (Howell, 2022; Orel & Bennis, 2021; Spinuzzi et al., 2019).

The number of coworking spaces worldwide has grown significantly, rising from around 600 locations in 2010 to approximately 15,500 in 2017 (Spinuzzi et al., 2019). The coworking phenomenon experienced rapid growth, reaching approximately 2,498 spaces by mid-2013, after posting an annual growth rate of nearly 100% between 2007 and 2012 (Bruno Moriset, 2013).

The concept of coworking encompasses a diverse audience, ranging from independent professionals to remote workers integrated into corporate structures. Individuals share collaborative environments based on the principle of "working alone, but together." In this study, the term "coworkers" is adopted as a unifying category for all users of these spaces, without ignoring the heterogeneity of their motivations and profiles (Clifton et al., 2022; Spinuzzi, 2019).

The coworking business model is largely based on a rental brokerage model. Companies in this sector enter into long-term contracts with property owners, reconfigure the spaces by incorporating common areas, social zones, and community-oriented services, and subsequently make these spaces available to users under more flexible terms, such as monthly contracts, typically at higher prices. The main users include entrepreneurs, freelancers, remote workers, and other independent professionals who, in many cases, lack the resources to maintain their own office. Beyond providing physical infrastructure, coworking fosters a social and collaborative environment, bringing together professionals who, although working on different projects, share the same space and benefit from interactions and synergies. In this way, coworking represents an innovation in both organizational structure and business model, offering unique solutions that emerge from the concentration of professionals in a single environment (Howell, 2022).

Coworking spaces differ according to user profiles, forms of interaction, available infrastructure, and community orientation. Factors such as collaboration, networking, technological resources, workspace design, flexibility, location, and shared values shape different coworking models and influence whether spaces prioritize collaboration, entrepreneurial support, creativity, or innovation (Bouncken, 2018).

According to Orel & Bennis (2021), a taxonomy of contemporary coworking spaces was proposed, classifying them based on essential conditions and distinguishing the original model—here referred to as coworking spaces geared toward individual purposes—from three alternative typologies that reflect different dimensions of this phenomenon, namely: spaces geared toward creation, group work, and startups.

Coworking spaces focused on individuals align with the model's original concept, catering to individual users rather than teams. In these settings, geographical and professional autonomy play a central role, and it is also common to foster interactions and a sense of community among members. In turn, spaces geared toward creativity share similarities with the former in terms of valuing the community and the individual, but differ in the type of activities carried out, which are generally more specialized and often associated with creative practices or hobbies, with professional autonomy not being the predominant element. Group-oriented spaces, on the other hand, tend to serve organized teams, including corporate structures, prioritizing separate areas or offices over shared open spaces. Nevertheless, they maintain the principle of fostering connections between individuals and organizations, offering physical and social conditions that encourage such interactions. Finally, coworking spaces geared toward startups have the primary purpose of supporting the development and success of new businesses. In these environments, networking and collaboration among organizations are essential, particularly for pooling resources and expertise. Access to these spaces is often temporary and based on selection criteria, although other models may incorporate startup support components without time limits or formal evaluation processes (Orel & Bennis, 2021).

B. Quality dimensions in coworking spaces and remote work

A relevant perspective for understanding quality in coworking environments derives from studies on Quality of Working Life (QWL) and workplace wellbeing. Ciccarelli (2023), in a systematic review of coworking literature, argues that coworking spaces can function as quasi-organizations by reproducing elements traditionally associated with employers, including physical infrastructure, social support, work routines, learning opportunities, and community integration. The author highlights that coworking environments affect workers' wellbeing through material, technological, and social conditions that shape the quality of work experiences, particularly for freelancers and remote workers. Dimensions such as ergonomic conditions, work-life balance, access to infrastructure, social interaction, opportunities for collaboration, flexibility, and perceived support emerge as recurring determinants of workplace quality.

From a workplace assessment perspective, Riratanaphong (2021) demonstrates that user satisfaction in coworking spaces is influenced by a combination of physical work environment characteristics and support services. Based on empirical evidence from coworking spaces in Bangkok, the study identifies relevant quality-related factors including indoor environmental quality (lighting, acoustics, temperature, and air quality), workspace arrangement, furniture, IT infrastructure, supporting facilities, and operational services such as security and cleaning. The findings reinforce the importance of both functional and experiential dimensions in shaping coworking quality and user satisfaction.

III. METHODS

This study employs a qualitative, empirical, and exploratory-descriptive methodology with the aim of developing a quality assessment model for coworking spaces for remote work, using Lisbon and Porto as empirical study sites to operationalize quality dimensions. The choice of a qualitative methodology is justified by the need to conduct an in-depth analysis of the spatial, functional, and social dynamics associated with these work environments. The study seeks to answer the following research questions: (i) Which quality dimensions are most relevant for coworking spaces geared towards remote work? (ii) How can these dimensions be operationalized in an evaluation framework? (iii) To what extent do the coworking spaces identified in Lisbon and Porto show different quality configurations?

This study focuses on the development of a quality assessment model for coworking spaces for remote work. The methodology was structured in three main stages: The research combines a literature-based conceptual phase with an empirical comparative analysis of coworking spaces located in Lisbon and Porto.

The methodological process was organized into three sequential phases: (1) conceptual identification of quality dimensions; (2) empirical mapping and characterization of coworking spaces: evidence from Lisbon and Porto; and (3) framework development and comparative analysis.

.

A. Data collection

Data collection focuses on mapping coworking spaces in the cities of Lisbon and Porto, considering the municipal boundaries of both cities as a territorial unit. This delimitation allows for a comparable and consistent analysis, focused on central urban areas where the concentration of coworking spaces is most significant. Data collection was based on multiple secondary sources, with the aim of ensuring the comprehensiveness and reliability of the data. The main sources used include:

- Platforms and directories specializing in coworking: Workin.space (https://workin.space/); Regus (https://www.regus.com/pt), Coworker (https://www.coworker.com/), Nomad List (https://nomads.com/), Coworkbooking (https://www.coworkbooking.com/), Google maps Coworking Lisbon (https://maps.app.goo.gl/bYqtLE1EtUpXPLXMA) and Google maps Coworking Porto (https://maps.app.goo.gl/tvG1Cwpv3PkACdL27)
- Google Maps, for geographic identification and location validation
- Official websites of coworking spaces
- Social media (e.g., Instagram, LinkedIn), to verify recent activity
- Review platforms (e.g., user reviews), to gather qualitative insights

The use of multiple sources allowed for data triangulation and greater robustness in the construction of the database. For each identified space, the following variables were recorded: Name of the space; Address and urban area; Type of space (open space, private office, hybrid); Hours of operation; Reference prices (when available); Infrastructure and amenities (internet, meeting rooms, common areas, etc.); Links (website, social media); User reviews and ratings. This set of variables allowed for a detailed and comparable characterization of the analyzed spaces. The list of coworking spaces in the cities of Lisbon and Porto is available in Appendices 1 and 2, used for empirical screening and validation of analytical dimensions.

B. Quality dimensions of coworking spaces in remote work: evidence from Lisbon and Porto

1) Mapping Coworking Spaces in Lisbon: evidence for quality dimensions

The city of Lisbon represents a particularly relevant empirical setting for examining the quality dimensions of coworking spaces in the context of remote work due to its consolidated ecosystem, high concentration of international professionals, and strong integration with digital entrepreneurship and innovation networks. The Lisbon case was used as empirical evidence to identify how quality-related attributes are manifested in different coworking

configurations. Through an exploratory screening based on multiple secondary sources, including specialized coworking platforms, official websites, Google Maps, social media, and user review platforms, coworking spaces were identified and analyzed according to predefined quality criteria. A total of 122 coworking spaces were identified within Lisbon's municipal boundaries, enabling a heterogeneous empirical base for analytical comparison.

The analysis focused on observable indicators associated with the proposed quality dimensions of coworking spaces for remote work. First, digital infrastructure was assessed through the identification of connectivity-related features, including Wi-Fi availability, dedicated internet services, phone booths, meeting rooms for virtual communication, and technological support resources. Second, comfort and ergonomics were examined based on workspace organization, lighting conditions, ventilation, furniture quality, acoustic comfort, and cleanliness, as described in publicly available information and user reviews.

Third, focus conditions were evaluated through indicators associated with quiet environments, availability of private or semi-private work areas, occupancy organization, and the existence of spaces designed to reduce interruptions during focused work activities. Fourth, collaboration and community dimensions were analyzed through the presence of networking initiatives, events, collaborative activities, and references to community engagement promoted by coworking operators.

Additional dimensions were operationalized through documentary evidence. Services and operations included reception services, technical support, kitchens, cafés, lockers, meeting facilities, and common areas designed to support daily work routines. Security and privacy were assessed through references to controlled access, private offices, secure internet access, and designated areas for confidential meetings or calls. Flexibility was examined through the diversity of plans available, such as daily, monthly, or 24/7 access options, reflecting the adaptive needs of remote professionals. Accessibility and mobility were considered through spatial location, proximity to transport systems, bike parking, and accessibility features. Finally, value and transparency were examined based on publicly available information concerning pricing structures, included services, and perceived cost-benefit relationships.

2) Mapping Coworking Spaces in Porto: evidence for quality dimensions

The city of Porto provides a complementary empirical setting for examining the quality dimensions of coworking spaces in the context of remote work, due to its emerging ecosystem, creative orientation, and strong local identity. Rather than merely identifying coworking spaces, Porto was analyzed to observe how quality dimensions are configured in a smaller-scale yet diversified coworking ecosystem. Following the same documentary procedures applied in Lisbon—including specialized coworking platforms, official websites, Google Maps, social media, and user reviews—47 coworking spaces were identified within Porto's municipal boundaries, providing a heterogeneous empirical basis for analysis.

The analysis focused on observable indicators associated with quality dimensions relevant to remote work. Digital infrastructure was assessed through internet reliability, Wi-Fi availability, meeting rooms, videoconferencing facilities, and technological resources supporting professional activities. Comfort and ergonomics were examined through workspace organization, furniture, lighting, thermal comfort, cleanliness, and acoustic conditions, while focus conditions considered quiet environments, private offices, and layouts designed to support concentration and reduce interruptions.

Collaboration and community emerged as particularly relevant dimensions in Porto, where coworking spaces frequently emphasized local identity, networking opportunities, creative activities, and community-oriented initiatives. Additional dimensions were operationalized through documentary analysis, including services and operations (meeting facilities, kitchens, cafés, lockers, and support amenities), security and privacy (access control and confidential workspaces), flexibility (daily, monthly, and 24/7 access models), accessibility and mobility (proximity to transport and central areas), and value and transparency, particularly reflected in affordable pricing structures and flexible entry options for remote workers.

C. Comparative framework

The comparative table was developed to examine how the quality dimensions of coworking spaces are configured in Lisbon and Porto, in the context of remote work, as shown in Table 1. Instead of focusing on the number of spaces or their geographical concentration, the analysis compares how coworking environments respond to the needs of remote professionals by operationalizing nine quality dimensions: (1) digital infrastructure, (2) comfort and ergonomics, (3) concentration conditions, (4) collaboration and community, (5) services and operations, (6) security and privacy, (7) flexibility, (8) accessibility and mobility, and (9) value and transparency.

Documentary evidence suggests that coworking spaces in both cities generally provide adequate technological infrastructure, operational support, and flexible access models. However, relevant differences emerge in the configuration of quality dimensions. Lisbon coworking spaces tend to emphasize structured work environments, stronger technological infrastructure, professional layouts, and diversified service offerings, reflecting the demands of a more internationalized and mature ecosystem. Porto coworking spaces, in contrast, more frequently combine functional infrastructure with informal, creative, and community-oriented environments, emphasizing collaboration, accessibility, and affordability.

Differences are also visible in focus conditions and community dynamics. Lisbon coworking spaces often prioritize acoustic comfort, private areas, and professional networking opportunities, whereas Porto spaces tend to promote

informal interaction, local identity, and collaborative environments. Regarding value and transparency, Lisbon coworking spaces generally offer broader service portfolios at higher price levels, while Porto spaces present lower entry barriers and more accessible pricing structures.

Overall, the comparative framework highlights that coworking quality is shaped by distinct urban and organizational configurations. While Lisbon demonstrates a more structured and service-intensive model, Porto reflects a more flexible and community-oriented approach. These differences reinforce the importance of a multidimensional framework for evaluating coworking quality in the context of remote work.

TABLE I. COMPARATIVE QUALITY DIMENSIONS OF COWORKING SPACES IN LISBON AND PORTO

Quality Dimension	Main theoretical foundation	Lisbon	Porto
Digital infrastructure	Riratanaphong (2021); Ciccarelli (2023)	Strong emphasis on structured technological resources, videoconferencing rooms, phone booths, and professional connectivity	Adequate technological infrastructure, with flexible and multifunctional solutions for independent workers and small teams
Comfort and ergonomics	Riratanaphong (2021); Ciccarelli (2023)	Professional layouts, ergonomic furniture, acoustic insulation, and highly structured work environments	Comfortable and functional environments, often combining professional infrastructure with creative and informal settings
Focus conditions	Riratanaphong (2021); Ciccarelli (2023)	Greater emphasis on quiet zones, private offices, and environments designed for concentration	More hybrid environments balancing focused work with informal interaction
Collaboration and community	Ciccarelli (2023); Spinuzzi et al. (2019); Howell (2022)	Networking oriented toward startups, international professionals, and business ecosystems	Stronger emphasis on local identity, informal interaction, creativity, and community engagement
Services and operation	Riratanaphong (2021); Howell (2022)	Diversified and formalized service offerings, including meeting rooms, reception, support services, and premium amenities	Functional operational support with emphasis on practical services and collaborative environments
Security and privacy	Riratanaphong (2021)	Greater presence of controlled access, private offices, and confidential workspaces	Secure environments with privacy mechanisms, although generally less corporate-oriented
Flexibility	Ciccarelli (2023); Orel & Bennis (2021)	Flexible access models combined with diversified workspace configurations and premium services	Flexible plans, lower entry barriers, and adaptable access models for freelancers and small teams
Accessibility and mobility	Riratanaphong (2021); St George (2015)	Strong integration with business districts, transport hubs, and innovation ecosystems	Integration with central neighborhoods, walkable areas, and creative districts
Value and transparency	Howell (2022); Ciccarelli (2023)	Broader service portfolio associated with higher price levels	More affordable pricing structures with satisfactory infrastructure and operational support

IV. DISCUSSION AND RESULTS

A. Dimensions of quality in coworking for remote work

Building upon these contributions, this study conceptualizes coworking quality as a multidimensional construct particularly relevant in remote work settings, where coworking spaces substitute or complement traditional offices and home environments. Accordingly, nine analytical dimensions were operationalized: (1) digital infrastructure, reflecting technological readiness and connectivity conditions; (2) comfort and ergonomics, associated with workspace functionality and environmental comfort; (3) focus conditions, related to concentration and interruption management; (4) collaboration and community, concerning social interaction and networking opportunities; (5) services and operations, referring to operational support and workplace amenities; (6) security and privacy, involving access control and confidential work conditions; (7) flexibility, associated with adaptable access and workspace configurations; (8) accessibility and mobility, referring to spatial integration and transport access; and (9) value and transparency, reflecting the perceived relationship between costs, services, and workplace benefits. Rather than emerging independently from the empirical cases, these dimensions were derived from previous literature on quality of working life, workplace wellbeing, coworking satisfaction, and physical work environments, and subsequently adapted to the specific requirements of remote work in coworking settings. This theoretical grounding supports the operationalization of quality dimensions employed in the comparative analysis of Lisbon and Porto. The proposed dimensions can be assessed using a 1–5 Likert scale for each quality dimension, allowing for a comparative analysis across different spaces, as shown in Table 2.

TABLE II. PROPOSED QUALITY MODEL (EVALUATION GRID)

Quality Dimension	Description
Digital infrastructure	This dimension refers to the availability and reliability of technological resources essential for remote work. It includes the quality and redundancy of the Wi-Fi network, access to Ethernet cable, connection stability, suitable rooms for video calls, and the availability of soundproof booths. These elements are fundamental to ensuring operational continuity and avoiding interruptions in remote work.
Comfort and ergonomics	This dimension assesses the physical conditions of the workspace, including the quality of chairs and desks, natural and artificial lighting, temperature control, ventilation, and noise levels. Spaces with good ergonomic conditions tend to promote greater well-being and reduce physical fatigue during long periods of work.

Quality Dimension	Description
Focus conditions	This dimension analyzes the existence of environments conducive to concentration, such as quiet zones, individual booths, rules for phone calls, and occupancy density. These factors are particularly relevant for remote workers who require extended periods of concentration.
Collaboration and community	This dimension considers the social role of coworking, including the organization of events, networking opportunities, collaborative culture, and onboarding processes for new members. These elements contribute to reducing the isolation associated with remote work and promoting social capital.
Services and operation	This dimension includes services that support daily operations, such as reception, cleaning, technical support, lockers, kitchen, café, and common areas. The availability of these services improves the user experience and reduces operational friction.
Security and privacy	This dimension assesses access control mechanisms, physical security, data protection, and the existence of private spaces for confidential calls or meetings. These aspects are particularly relevant for professionals who handle sensitive information.
Flexibility	This dimension refers to the diversity of plans available, including daily passes, part-time, 24/7 access, and ease of upgrading or downgrading between different options. Flexibility is a key factor for remote workers with varying needs.
Accessibility and mobility	This dimension considers the location of spaces, proximity to public transport, bicycle parking, universal accessibility, and ease of movement. These factors directly influence the attractiveness of spaces.
Value and transparency	This dimension assesses the clarity of the pricing structure, additional costs, included services, and perceived value for money. Price transparency is especially important for freelancers and independent professionals.

The proposed model has implications for various stakeholders. For users, it serves as a decision-support framework, allowing them to compare coworking spaces based on their specific needs. For operators, the model identifies priorities for improvement and opportunities for competitive differentiation. For cities, the development of high-quality coworking spaces contributes to attracting talent, promoting remote work, and strengthening urban innovation ecosystems. Price transparency and plan diversification are also key factors in attracting different user profiles.

V. Conclusion

This study contributes to the understanding of coworking spaces in the context of remote work by proposing and operationalizing a multidimensional quality assessment framework grounded in the literature on quality of working life, workplace wellbeing, and coworking user satisfaction. Rather than focusing solely on the identification of coworking spaces, the study examined how quality-related attributes are configured and manifested in different coworking environments, using Lisbon and Porto as empirical settings for comparative analysis. The findings demonstrate that coworking quality should be understood as a multidimensional construct encompassing digital infrastructure, comfort and ergonomics, focus conditions, collaboration and community, services and operations, security and privacy, flexibility, accessibility and mobility, and value and transparency. Together, these dimensions provide a structured basis for evaluating coworking spaces as environments capable of supporting productivity, wellbeing, concentration, and social interaction in remote work settings. The comparative analysis suggests that Lisbon and Porto present distinct configurations of coworking quality. Lisbon is characterized by more structured and service-intensive environments, stronger technological infrastructure, and a more internationalized ecosystem, whereas Porto demonstrates more community-oriented, accessible, and flexible coworking configurations, frequently associated with creative and collaborative environments. These findings reinforce the importance of contextual and multidimensional approaches when evaluating coworking quality in different urban settings.. From a practical perspective, the proposed framework may support remote professionals in selecting coworking spaces according to their specific work needs, while also providing coworking operators with guidance for service improvement and differentiation. The study presents some limitations, particularly the reliance on documentary and publicly available data, the dynamic nature of coworking ecosystems, and the absence of direct user-based empirical validation. Future research may expand the proposed framework through surveys, interviews, and longitudinal studies examining the relationship between coworking quality dimensions and indicators such as user satisfaction, wellbeing, productivity, retention, and professional collaboration. Comparative studies involving other Portuguese and European cities may also contribute to refining and validating the proposed model.

Acknowledgment

This work was financially supported by the research unit on Governance, Competitiveness and Public Policy (UIDB/04058/2020)+(UIDP/04058/2020), funded by national funds through FCT - Fundação para a Ciência e a Tecnologia.

References

Bouncken, Ricarda B., R. B., & Reuschl, A. J. (2018). Coworking-spaces: How a phenomenon of the sharing economy builds a novel trend for the workplace and for entrepreneurship. Review of Managerial Science, 12(1), 317–334. https://doi.org/10.1007/s11846-016-0215-y

Bruno Moriset (2013). Building new places of the creative economy. The rise of coworking spaces. 2013. ⟨halshs-00914075⟩

Ciccarelli, F. C. (2023). Exploring the potential of coworking spaces for quality of working life and wellbeing: A systematic review of academic literature. *Cidades, Comunidades e Territórios, 46*, 1–18. https://journals.openedition.org/cidades/7155

Clifton, N., Füzi, A., & Loudon, G. (2022). Coworking in the digital economy: Context, motivations, and outcomes. Futures, 135. https://doi.org/10.1016/j.futures.2019.102439

Deskmag. (2019). "2019 COWORKING FORECAST FINAL RESULTS", available at: https://www.dropbox.com/s/jjor71mecwqbxdy/2019%20Complete%20Coworking%20Forecast.pdf?dl=0. (accessed 02 April 2026).

European Commission. (2020). Startup Europe: Connecting European startup ecosystems. Available in: https://digital-strategy.ec.europa.eu/en/policies/startup-europe

Garrett, L. E., Spreitzer, G. M., & Bacevice, P. A. (2017). Co-constructing a sense of community at work: The emergence of community in coworking spaces. Organization Studies, 38(6), 821–842. https://doi.org/10.1177/0170840616685354

Howell, T. (2022). Coworking spaces: An overview and research agenda. Research Policy, 51(2). https://doi.org/10.1016/j.respol.2021.104447

OECD. (2023). OECD SME and Entrepreneurship Outlook 2023. Paris: OECD Publishing. Available in: https://www.oecd.org/en/publications/oecd-sme-and-entrepreneurship-outlook-2023_342b8564-en.html

Orel, M., & Bennis, W. M. (2021). Classifying changes. A taxonomy of contemporary coworking spaces. In Journal of Corporate Real Estate (Vol. 23, Number 4, pp. 278–296). Emerald Group Holdings Ltd. https://doi.org/10.1108/JCRE-12-2020-0061

Riratanaphong, C. (2021). Factors influencing user satisfaction: Case studies of ten coworking spaces. In Proceedings of International Structural Engineering and Construction, 8(1), FAM-04-1–FAM-04-6. ISEC Press. https://doi.org/10.14455/ISEC.2021.8(1).FAM-04

Spinuzzi, Clay, C. (2012). Working alone together: Coworking as emergent collaborative activity. Journal of Business and Technical Communication, 26(4), 399–441. https://doi.org/10.1177/1050651912444070

Spinuzzi, C., Bodrožić, Z., Scaratti, G., & Ivaldi, S. (2019). "Coworking Is About Community": But What Is "Community" in Coworking? Journal of Business and Technical Communication, 33(2), 112–140. https://doi.org/10.1177/1050651918816357

Startup Genome. (2023). Global Startup Ecosystem Report 2023. Available in: https://startupgenome.com/reports/gser2023

St George, C. (2015). Coworking in the city. In Ephemera (Vol. 15, Number 2). http://openaccess.city.ac.uk/publications@citystgeorges.ac.ukPermanentrepositorylink:https://openaccess.city.ac.uk/id/eprint/14478/

APPENDIX I. LIST OF COWORKING SPACES IN THE CITY OF LISBON

Name of the coworking space	Address	Type (open space, private office, hybrid)	Opening hours	Reference prices	Infrastructure and amenities (Products and Services)	Links (website, social networks)	User reviews and ratings	Google Rating (1 to 5)
Impact Hub Lisbon - Baixa	Tv. das Pedras Negras 1 1º Esquerdo, 1100-404 Lisboa	Private office	09:00–18:00	167€ / month	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Heating, Kitchen, Pet-friendly, Personal Lockers, Catering Kitchen, Free Coffee, Free Drinking Water, Free Tea, Free Wine.	https://lisbon.impacthub.net/	Pleasant atmosphere, well located (city center), modern equipment, friendly staff, clean and organized environment.	5 stars
Venture X Portugal	Av. da Liberdade 224 2nd Floor, 1250-147 Lisboa	Private office	09:00–18:00	150€ / month	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Air Conditioning, Heating, Kitchen, Outdoor Terrace, Wheelchair Accessibility, Personal Lockers, Catering Kitchen, Event Space for Rent, Free Coffee, Free Drinking Water, Free Snacks, Free Tea, Parking/Storage Parking, Bike Parking, Plug in for Electric Cars.	https://venturex.pt/	Strategically located, this space offers a professional and modern environment, with private offices and fully equipped meeting rooms. The common areas are very comfortable, with soundproof booths.	5 stars
Coworking Business Networking Lounge	R. Mário Gomes Páscoa 8C, 1600-824 Lisboa	Private office	08:00–20:00	150€ / month	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Air Conditioning, Heating, Kitchen, Outdoor Terrace, Wheelchair Accessibility, Personal Lockers, Catering Kitchen, Event Space for Rent, Free Coffee, Free Drinking Water, Free Snacks, Free Tea, Parking/Storage Parking, Bike Parking, Plug In for Electric Cars.	http://bn-lounge.com/	The internet is super-fast, the kitchen is equipped, the location is good, and the environment is very quiet. It's excellent for meetings with business partners or clients.	5 stars
Impact Hub Lisbon - Penha	R. Neves Ferreira 13, 1170-273 Lisboa	Private office	09:00–18:00	225€ / month	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Air Conditioning, Heating, Kitchen, Outdoor Terrace, Wheelchair Accessibility, Personal Lockers, Catering Kitchen, Event Space for Rent, Free Coffee, Free Drinking Water, Free Snacks, Free Tea, Parking/Storage Parking, Bike Parking, Plug in for Electric Cars.	https://lisbon.impacthub.net/	Super welcoming space, very friendly people, well-lit and with great facilities. Very bright space, tranquil atmosphere, ergonomic chairs.	5 stars
Quarto Andar - Coworking	R. Padre Luís Aparício 9 4º Direito, 1150-248 Lisboa	Private office	09:30–18:00	125€ / month	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Air Conditioning, Heating, Kitchen, Outdoor Terrace, Wheelchair Accessibility, Personal Lockers, Catering Kitchen, Event Space for Rent, Free Coffee, Free Drinking Water, Free Snacks, Free Tea, Parking/Storage Parking, Bike Parking, Plug In for Electric Cars.	https://quartoandar.space/	The space is very clean, organized, and well-equipped. The workstations are spacious and have excellent quality chairs. A welcoming environment. Comfortable chairs. The office space has everything you need.	5 stars
Eureka Coworking - Lisboa	Palácio Rosa, R. Seara Nova 5, 1250-002 Lisboa	Private office	09:00–18:00	140€ / month	Wheelchair accessible entrance. LGBTQ+ friendly. Safe space for transgender people. On-site parking.	http://eurekacoworking.com/	Very attentive team and great facilities. Excellent quality standard. Very quiet and clean workspace. Great coworking space! Very cozy.	5 stars
Avila Spaces - Parque das Nações	Av. Dom João II 52 Piso 0, 1990-096 Lisboa	Private office	08:00–19:00	220€ / month	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Air Conditioning, Heating, Kitchen, Outdoor Terrace, Wheelchair Accessibility, Personal Lockers, Catering Kitchen, Event Space for Rent, Free Coffee, Free Drinking Water, Free Snacks, Free Tea, Parking/Storage Parking, Bike Parking	http://www.avilaspaces.com/	Very welcoming space, excellent workspace, very well located in Parque das Nações. The facilities are modern, well-maintained and very functional, with very comfortable phone booths. Beautiful decor.	5 stars
Office Evolution	Edifício, R. Soeiro Pereira Gomes 8 andar 11, 1600-024 Lisboa	Private office	08:30–17:30	240€ / month	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Air Conditioning, Heating, Kitchen, Outdoor Terrace, Wheelchair Accessibility, Personal Lockers, Catering Kitchen, Event Space for Rent, Free Coffee, Free Drinking Water, Free Snacks, Free Tea, Parking/Storage Parking, Bike Parking, Gym, Massages, Yoga	https://officeevolution.pt/locations/portugal/lisbon/entrecampos/?utm_source=google&utm_medium=organic&utm_campaign=gbp_listing	The design is bright and well-planned, with plenty of natural light, striking a good balance between professionalism and informality. The location is super convenient, well-organized, and easy to use.	5 stars

Name of the coworking space	Address	Type (open space, private office, hybrid)	Opening hours	Reference prices	Infrastructure and amenities (Products and Services)	Links (website, social networks)	User reviews and ratings	Google Rating (1 to 5)
					Studio			
Lisbon-Cowork	Rua da Bela Vista à Graça 27A, 1170-058 Lisboa	Private office	Open 24 hours	130€ / month	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Air Conditioning, Heating, Kitchen, Outdoor Terrace, Wheelchair Accessibility, Personal Lockers, Catering Kitchen, Event Space for Rent, Free Coffee, Free Drinking Water, Free Snacks, Free Tea, Parking/Storage Parking, Bike Parking	https://lisbon-cowork.com/	The place is amazing, it has a very well-organized structure, open spaces with cozy and relaxed decor, and a deck with a breathtaking view of the Tagus River. The chairs and tables are very comfortable.	4,9 stars
Avila Spaces	Av. da República 6 1º esq, 1050-191 Lisboa	Private office	09:00–19:00	150€ / month	24 Access, Parking, Printing, Kitchen.	http://www.avilaspaces.com/	A very pleasant, calm environment, ideal for working. A combination of modernity, elegance, convenience, and dynamism. The outdoor patio is perfect for meals and recharging.	4,9 stars
Kube Coworking | Lisboa	R. do Centro Cultural 27A, 1700-106 Lisboa	Private office	08:30–19:30	225€ / month	Outdoor spaces, Cultural Club: Our coworking café, Community events, Pet-friendly	https://kubecoworking.pt/	A welcoming, professional environment, ideal for working with focus and motivation, calm and inspiring, which facilitates concentration and enhances productivity. The space is spacious, well-lit and very well organized.	4,9 stars
Heden Rossio	Largo do Duque de Cadaval 17 Fracção I, 1200-160 Lisboa	Private office	09:00–18:00	225€ / month	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Heating, Kitchen, Pet-friendly, Personal Lockers, Catering Kitchen, Free Coffee, Free Drinking Water, Free Tea, Free Wine.	https://heden.co/coworking-rossio/	Bright, comfortable space with excellent working conditions. Pleasant and clean environment. It has a fantastic interior design and is in a prime location, well connected by public transport.	4,6 stars
IDEA Spaces - Saldanha	Av. Defensores de Chaves 4, 1000-117 Lisboa	Private office	09:00–18:00	42€ / month (Virtual offices)	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Air Conditioning, Heating, Kitchen, Outdoor Terrace, Wheelchair Accessibility, Personal Lockers, Catering Kitchen, Event Space for Rent, Free Coffee, Free Drinking Water, Free Snacks, Free Tea, Parking/Storage, Parking, Bike Parking, Swimming Pool	https://ideaspaces.pt/	Excellent service, super cool work environment, with rooms for all profiles, professionals and different types of business. Beautiful, clean, super functional place, good rooms and a small cinema room that is amazing.	4,5 stars
IDEA Spaces - Parque das Nações	Av. Dom João II 13, 1990-077 Lisboa	Private office	09:00–18:00	140€ / month	Wi-Fi, Standing Desks, Printing, 24 Access, Phone Booth, Lounge Area, Air Conditioning, Heating, Kitchen, Outdoor Terrace, Wheelchair Accessibility, Personal Lockers, Catering Kitchen, Event Space for Rent, Free Coffee, Free Drinking Water, Free Snacks, Free Tea, Parking/Storage Parking, Bike Parking	https://www.ideaspaces.pt/	Very well-equipped space: lounge, phone booths, meeting rooms, secure Wi-Fi. Excellent location and working conditions. Space with good infrastructure that provides a range of essential services.	4,7 stars

APPENDIX II. LIST OF COWORKING SPACES IN THE CITY OF PORTO

Name of the coworking space	Address	Type (open space, private office, hybrid)	Opening hours	Reference prices	Infrastructure and amenities (Products and Services)	Links (website, social networks)	User reviews and ratings	Google Rating (1 to 5)
Bõn Coworking	R. Joaquim António de Aguiar 150, 4300-465 Porto	Private office	08:00–18:00	69€ / month	24/7 access. Dogs allowed. Public transport nearby. Wheelchair accessible. Shared accommodation on site. English-speaking staff. Copying machine. Ergonomic chairs, Printer, Scanner, Quiet area, Kitchen, Free coffee/tea, Refrigerator, Microwave.	https://www.boncoworking.com/	Amazing, cozy, and well-equipped space. The space is very well maintained, clean, and has a great energy for working. The internet is fast, and the environment is quiet.	5 stars
Synergy Hub - Coworking, Eventos Corporativos, Estúdio de Vídeo	Av. de Fernão de Magalhães 613, 4350-154 Porto	Hybrid	09:00–20:00	15€ / day	Coffee and water included. A4 and A3 color Wi-Fi printer available.	http://hub.synergy-porto.com/	Welcoming and relaxed atmosphere, homemade meals, strategic location in Porto. Excellent facilities, easy access, easy parking, in an area with access to restaurants, hotels and everything else.	4,9 stars
Typographia Cowork	Campo dos Mártires da Pátria 144 A, 4050-368 Porto	Private office	09:00–13:00 14:00–18:00	65€ / month	Wi-Fi, Heating, Lounge Area, Outdoor Terrace, Kitchen, Personal Lockers, Event Space For Rent, Printing, Bike Parking	http://www.typographia-cowork.com/	A pleasant, elegant, and well-equipped place. A welcoming and enjoyable shared office. The location is great.	4,7 stars
CRU Creative Hub	R. do Rosário 211, 4050-524 Porto	Hybrid	10:00–19:30	75€ / month	Internet Wi-Fi, Flexible or dedicated desks, Meeting rooms (limited or unlimited depending on the plan), Printer and mail reception, Access to common areas and creative community	https://www.crucreativehub.com/	Wonderful coffee, unbeatable friendliness, shop with a great selection of Portuguese products. Very welcoming and geared towards the creative industries! At the entrance there is a small gallery/shop selling local products, many of them developed by artists and designers from the coworking space itself.	4,7 stars
Work Wise Co-Working & Offices	R. de Sá da Bandeira 706 6D, 4000-432 Porto	Shared open space	08:00–20:00	90€ / month	Private offices, Open space, Meeting rooms, Dedicated internet, Reception, Printing, Common areas.	https://workwise.today/	Freedom, tranquility, and security. The space is pleasant. The quality of the facilities and the professionalism of the management team are excellent. A modern and welcoming environment.	4,9 stars
Vertical Coworking Bolhão	R. Guedes de Azevedo 131, 4000-272 Porto, Portugal	Hybrid	Open 24 hours	100€ / month	24/7 Access, Wi-Fi, Flexible Desks, Meeting Rooms Common Areas, Freelancer Community	https://verticalcoworking.com/	Good working conditions, leisure opportunities, and a great location. The environment is modern and well-decorated with varied workspaces that meet all needs, from private offices to collaborative areas.	4,7 stars
PortoSoul Cowork	R. de Mouzinho da Silveira 226, 4050-417 Porto	Hybrid	Open 24 hours	15€ / day	24-hour access, Wi-Fi, Shared tables, Lounge areas, Networking, Community atmosphere	https://www.nortesoul.com/coworking-spaces/	Clean space. Good Wi-Fi. Great location. Good value for money. Very cozy and pleasant.	5 stars
WOW - Coworking Porto	Estr. da Circunvalação 9543, 4250-148 Porto	open space	Open 24 hours	120€ / month	Free parking, fiber internet, daily cleaning, auditoriums, meeting rooms, lounge, equipped kitchen, restaurant, and green areas. There is also a package reception area, video surveillance, and individual lockers, creating a	https://www.wowbyfinsa.com/cowork/	Fantastic location, lots of natural light, parking, comfortable chairs. The atmosphere is very friendly. Dynamic and creative space.	4,5 stars

Name of the coworking space	Address	Type (open space, private office, hybrid)	Opening hours	Reference prices	Infrastructure and amenities (Products and Services)	Links (website, social networks)	User reviews and ratings	Google Rating (1 to 5)
					comfortable and safe environment for daily work.			
Facts Coworking	R. de 5 de Outubro 185, 4100-175 Porto	Hybrid	09:00–19:00	100€ / month	Users have access to high-speed internet, air conditioning, heating, printers, personal lockers, and a fully equipped kitchen with free coffee and tea. The space is pet-friendly, wheelchair accessible, and has gender-neutral restrooms. There is also paid parking nearby and space to store bicycles.	https://factscoworking.com/	Wonderful environment. Super comfortable and with everything you need. Clean place. High-speed wired (Ethernet) internet.	4,9 stars
Workin Porto	Via do Castelo do Queijo N.º 395 Lojas 22 e 23, 4100-429 Porto, Portugal	Hybrid	7:00 - 23:00	14€ / day 60€ / month	Space + Furniture, Meeting Rooms, Expenses Included, Cleaning, electricity and water, unlimited Wi-Fi included, Happy Hours and Community Benefits	https://www.workin.pro/	Excellent place. The location is ideal, the amenities are plentiful, and the spaces are ample. Incredible view!!!	4,9 stars
Re-work Porto Cowork	R. da Boavista 217, 4050-107 Porto	open space	8:00 - 19:30	90€ / month	Shared tables, Wi-Fi, networking	https://instagram.com/re_work_porto	A calm, comfortable place with friendly coworkers. Super comfortable chairs, cozy atmosphere, very fast Wi-Fi, air conditioning. There's also coffee available, a fully equipped kitchen, and an outdoor area. The location is great, just a few minutes from Trindade.	5 stars
DeHouse Downtown	R. de José Falcão 199, 4050-215 Porto	Hybrid	8:00 - 20:00	120€ / month	Meeting rooms, lounge, community	https://dehousecommunity.com/	Very nice coworking space. Amazing space.	4,9 stars
Fisga - Warehouse	Rua de Santos Pousada 826, 4000-485 Porto	open space	8:00 - 21:00	80€ / month	Creative studios, coworking spaces, events	http://fisga.space/	It's a coworking space with excellent facilities, kind and creative people, and a rich cultural program (art exhibitions, pop-up dinners, events, workshops, etc.).	4,9 stars
Solvian Coworking	R. Álvaro Gomes 106A, 4150-063 Porto	Hybrid	8:00 - 21:00	110€ / month	Meeting rooms, private offices	https://www.solvian.pt/	Incredible, modern, and super versatile space! The atmosphere conveys both focus and creativity, making it perfect for working quietly or exchanging ideas in a group. A very modern, welcoming, and well-organized space.	5 stars
shiftworkplace coworking	R. do Bom Sucesso 372 1º, 4150-148 Porto	Hybrid	Open 24 hours	150€ / month	24/7 access, meeting rooms	https://www.shiftworkplace.pt/	A super welcoming and practical space! The environment blends modernity and comfort, stimulating concentration and productivity. Superbly located. A well-equipped, pleasant, and tranquil space.	5 stars

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Digitalising Quality Audits to Enhance Data Management: Evidence from a Metallurgical Company

Mariana Pinto[1], Leonor Teixeira[1,2], Helena Alvelos[1,3]

[1] Department of Economics, Management, Industrial Engineering and Tourism (DEGEIT), University of Aveiro; marianafpinto@ua.pt
[2] Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Laboratório Associado de Sistemas Inteligentes (LASI), Department of Economics, Management, Industrial Engineering and Tourism (DEGEIT), University of Aveiro, Portugal, lteixeira@ua.pt
[3] Center for Research & Development in Mathematics and Applications (CIDMA), University of Aveiro, Department of Economics, Management, Industrial Engineering and Tourism (DEGEIT), University of Aveiro, Portugal, Portugal; helena.alvelos@ua.pt

Abstract- In manufacturing, quality audits are essential to ensure product conformity and compliance with strict standards. However, these processes are often supported by manual and fragmented data collection practices, which limit traceability, analysis, and effective decision-making. In this context, digitalisation plays a key role in improving information management and supporting decision-making. This study presents a case study conducted in the Quality Department of a multinational company that produces faucets, where BPMN was used to analyse the audit process and support its redesign. Based on this analysis, an application and a dashboard were developed, enabling structured audit data registration, centralised information management, and real-time analysis. The results show that the implemented solution improved data consistency and process transparency, while facilitating access to relevant information for decision-making. The analysis of audit data allowed the identification of trends and recurring non-conformities, enabling managers to act more effectively and in a more targeted way, contributing to continuous process improvement.

Keywords – BPMN; Quality Audits; Data Digitalization; Industry 4.0; Low Code

INTRODUCTION

The transition towards Industry 4.0 has profoundly reshaped how organisations collect, analyse, and use data to monitor and optimise their operations. In increasingly digitalised industrial environments, fast, accurate, and integrated data collection systems have become essential to strengthen process transparency, improve operational control, and enable earlier deviation detection. According to Yu et al. (2021), digitalisation through automated data flows, interconnected systems, and data-driven decision-making mechanisms is fundamental for improving organisational responsiveness and ensuring process stability and efficiency. At the same time, growing market demands require companies to ensure product conformity with greater precision, speed, and reliability, particularly in high-end manufacturing contexts, where quality is a core source of competitive advantage (*Rudolf et al., 2024)*. In this context, continuous data collection and analysis play a decisive role, as they support more proactive quality management, enable root-cause identification, and sustain continuous improvement initiatives (Bousdekis et al., 2023).

The present study addresses this need by analysing the current audit processes carried out in the Quality Department of Grohe, a company that belongs to the LIXIL group, a global leader in the sanitary industry. Through BPMN-based process mapping, the study identifies inefficiencies in the existing workflow and supports the design of a future-state process oriented towards digitalisation. Based on this redesigned process, an application was developed to centralise audit information in a single digital environment, thereby improving data accessibility, consistency, and traceability, while strengthening the organisation's ability to monitor quality in real time.

The next section presents the methodology adopted in this study. This is followed by the theoretical background, outlining the most relevant concepts related to quality control, digitalisation, and low-code technologies. The results section then presents the BPMN-based analysis of the AS-IS process, the design of the TO-BE model, and the digital solution developed to support process transformation. This is followed by a discussion of the main findings. Finally, the main conclusions and directions for future work are presented.

METHODOLOGY

This study adopts a qualitative case-study approach based on direct observation of the audit processes and iterative process modelling. The audits were observed in their real operational context by accompanying quality controllers during the execution of Standard Work, Laser, and Product audits. Based on these observations, AS-IS process models were developed and subsequently refined through repeated observation and validation, ensuring that they accurately represented the actual workflows. The validated models were then analysed to identify inefficiencies, particularly in data collection, registration, and information flow. Based on this analysis, improvement opportunities were defined and used to design the TO-BE process model, supported by digital solutions such as AppSheet to enable automated data collection, centralised information storage, and improved analysis capabilities.

Industry 4.0 and Quality Management

Digitalisation is a core pillar of the Industry 4.0 paradigm, enabling the integration of cyber-physical systems, industrial Internet of Things (IIoT), cloud computing and advanced analytics within manufacturing environments (Kagermann et al., 2013). Through the continuous acquisition and exchange of data, digitalised production systems support real-time monitoring, decentralised decision-making and increased operational transparency (Yu et al., 2021). In the context of quality management, this transformation has led to the emergence of Quality 4.0, which extends traditional quality assurance practices by embedding digital technologies into inspection, auditing and control activities (Bousdekis et al., 2023).

The most recent stage of quality management evolution is Quality 4.0, which emerges from the convergence between consolidated quality management practices and the technologies associated with Industry 4.0. According to Carvalho et al. (2021), Quality 4.0 results from the application of digital technologies, such as data analysis, automation, and real-time connectivity, to quality management processes, enhancing previous approaches without replacing them. Broday (2022) reinforces this perspective, highlighting that Quality 4.0 represents the digitalization of quality management, combining traditional control methods with the new capabilities of Industry 4.0, and that this evolution requires a cultural transformation in organizations, with an increasing focus on the use of data for decision-making.

Process digitalisation

Industry 4.0 shifts quality management from a reactive, inspection-oriented approach to a more proactive and predictive paradigm. Digital tools enable faster detection of deviations, improved traceability across process stages and stronger integration between operational data and managerial decision-making. However, the effectiveness of these approaches depends on the reliability and structure of the data collected. Many manufacturing organisations still rely on manual or fragmented data acquisition practices, which limit the potential benefits of digitalisation and prevent the full realisation of Industry 4.0 principles (Lourenço et al., 2022).

Process digitalisation should not be understood merely as the conversion of paper-based records into digital formats, but rather as a broader transformation in the way information is captured, structured, shared and used to support decision-making. In this sense, digitalisation plays an important role in increasing the speed and capacity of analysing information derived from production processes, while also improving data accessibility, consistency and transparency. Silva et al. (2025) argue that well-organised, relevant and user-friendly information adds value to organisations and supports managerial decision-making. In the addressed case study, in an automotive company, the authors show that paper-based defect monitoring generated excessive recording time, inconsistent data, errors and difficulties in analysing and following up corrective actions, mainly due to the lack of centralised and organised information. Therefore, the digitalisation of the process not only improved the recording and synthesis of information, but also strengthened problem-solving and supported continuous improvement (Silva et al., 2025).

Data Control in Manufacturing

Data control in manufacturing refers to the systematic collection, validation, storage, and use of operational data to support process stability, quality assurance, and continuous improvement. High-quality data is essential for compliance monitoring, deviation analysis, traceability, and audit effectiveness. When data is inaccurate, incomplete, or delayed, quality tools such as Six Sigma, PDCA, or statistical process control lose analytical power and may lead to delayed or incorrect decisions (Antony et al., 2021).

The literature consistently identifies manual and paper-based data collection as a major source of data quality issues, including transcription errors, missing values, selective reporting, and latency between observation and documentation (Lourenço et al., 2022; Yu et al., 2021). These limitations are particularly critical in audit processes, where data functions not only as evidence of conformity but also as a trigger for corrective actions and escalation mechanisms (Santos et al., 2022). Fragmented data sources further hinder consolidation, trend analysis, and performance comparison over time (Bousdekis et al., 2023).

Compared to manual or semi-automated systems, digitally integrated data collection improves data availability, validation, and traceability, forming the basis for more diagnostic, predictive, and prescriptive quality analytics (Qin et al., 2016; Bousdekis et al., 2023). Empirical studies also show that digitally enabled inspection and IIoT-based monitoring systems improve defect detection, increase process consistency, and support proactive quality management (Rožanec et al., 2022; Banerjee et al., 2021)

Business Process Modelling with BPMN

Business Process Model and Notation (BPMN) is a standardised graphical language used to represent business processes in a clear, structured and unambiguous manner (Object Management Group, 2014). BPMN provides a set of elements to model activities, events, decision points, organisational roles and information flows, enabling consistent interpretation of processes across different stakeholders and organisational contexts (Dumas et al., 2018).

Within the process management lifecycle, BPMN is widely applied to model both the current state (AS-IS) and the desired future state (TO-BE) of a process. AS-IS models describe how processes are actually executed, serving as a baseline for identifying inefficiencies, redundancies and sources of variability (Aguilar-Savén, 2004). In contrast, TO-BE models represent improved versions of processes, incorporating structural changes, task reallocation or automation to address the limitations identified in the current state (Dumas et al., 2018).

The distinction between AS-IS and TO-BE modelling is central to process improvement initiatives, as it enables a systematic comparison between existing practices and intended objectives. The literature emphasises that effective TO-BE process design should be grounded in a rigorous analysis of the AS-IS process, ensuring coherence and traceability between process versions (Reijers & Mansar, 2005; Mendling et al., 2010).

One of the main strengths of BPMN lies in its ability to represent, within a single model, workflow logic, organisational responsibilities and data-related aspects. Pools and lanes are used to clarify roles and responsibilities, while the representation of data objects makes explicit the points at which information is created or recorded throughout the process (Object Management Group, 2014). This clarity contributes to model quality and understandability and supports the analysis of traceability requirements (La Rosa et al., 2011).

The literature shows that well-structured BPMN models enhance communication among stakeholders, improve process understandability and support evidence-based redesign decisions (Reijers & Mendling, 2011; Bandara et al., 2015). For these reasons, AS-IS and TO-BE modelling using BPMN provides a solid methodological foundation for continuous improvement initiatives, as well as for the digitalisation and automation of business processes.

Additionally, the integration of Business Process Modelling (BPM) into Quality Management Systems and tools creates preconditions for the development of an effective and efficient organization. BPM and Quality Management (QM) should not be considered as separate management disciplines with different tools, as BPM can be used as a methodology that can be integrated into QM paradigms (Stravinskiene and Serafinas, 2020).

Low-Code Platforms and Process Digitalisation

Within the scope of this study, AppSheet functions as the digital backbone for transforming Grohe's quality audit processes. By automating data entry, timestamping, audit identification and reporting workflows, AppSheet directly addresses the main limitations of the current system, namely manual data collection, delayed corrective actions and lack of end-to-end traceability. This implementation enables a structured transition from the AS-IS manual audit system to a TO-BE digital audit management solution aligned with BPMN-designed workflows, supporting data-driven quality management and continuous improvement.

Low-code and no-code platforms have emerged as practical enablers of digital transformation, allowing organisations to develop customised applications rapidly and with limited programming expertise (Saha & Dey, 2022). Their flexibility is particularly valuable in manufacturing contexts, where operational and quality-related processes frequently evolve and often require tailored workflows. Gao & Fagerholm (2026) add that, although low-code does not profoundly modify the overall software development process, it allows for faster creation of solutions and more agile adjustments throughout the project.

Among these platforms, AppSheet has gained relevance due to its ability to automate data collection, trigger workflow notifications, generate reports, connect to cloud databases, provide audit trails, and enforce data validation rules. These features make it particularly suitable for supporting quality audit processes, where structured data capture, traceability, and timely reporting are critical. Although academic literature specifically focused on AppSheet remains limited, its adoption is aligned with broader research showing that low-code solutions improve operational agility and data accuracy in Industry 4.0 environments (Saha & Dey, 2022; Ludbrook et al., 2025). This relevance is reinforced by its adoption within the LIXIL Group, which has publicly highlighted the use of Google Cloud and AppSheet to democratise data access and support internal digital solutions developed by non-technical employees (Google Cloud, 2023).

CASE STUDY

Problem Caracterization and proposed solution

While Grohe is a global leader in the sanitary industry, the focus of this project lies on its production facility located in Albergaria-a-Velha, Portugal. This plant represents a strategic unit within the group and serves as the empirical setting for the present study. It is responsible for manufacturing a wide and diverse portfolio of products, including a significant share of essential faucets and sanitary components.

Notably, the facility produces thermostatic cartridges for the group's entire global manufacturing network and operates as a centre of excellence for Physical Vapor Deposition (PVD), a distinctive process that enables the production of high-end coloured faucets. In addition, the plant manufactures technologically advanced solutions and a complete range of bathroom products, including bidets, bathtubs, and showers. The diversity, complexity, and technological sophistication of these production activities require a high level of process stability and operational consistency.

To address these requirements, the Albergaria-a-Velha plant has been progressively implementing the Standard Work framework in the production line, to ensure consistent task execution, reduce process variability, and support quality assurance. Standard Work defines the sequence of operations, task allocation, and execution methods for each product and production line, thereby establishing a common and repeatable way of working. However, as this framework is still an ongoing implementation project at the site, it is currently applied only to selected products and production lines.

In this context, systematic verification of compliance with Standard Work has become essential to ensure its effective adoption and to prevent deviations during daily operations. This need has led to the establishment of the Standard Work Audit, which focuses specifically on assessing whether the defined task division and execution sequence are respected in practice. The audit evaluates compliance with the Standard Work Sheet, which is adaptable to different production scenarios depending on the number of operators assigned to the line. Monitoring compliance is particularly critical in such a complex manufacturing environment, as deviations from standardized practices can compromise process stability, hinder error traceability, and ultimately affect product quality.

To illustrate the process, the Standard Work audit will be used as a representative example, although the work developed also encompassed three additional audit processes with a similar structure and approach.

The Standard Work Audit process is currently performed at least four times per day within the Quality Department, reflecting its critical role in ensuring process stability and product conformity. Due to this high execution frequency, the audit process requires significant operational effort and generates a substantial volume of data daily. In its current configuration, the process relies heavily on manual data collection and documentation, which amplifies inefficiencies and increases the likelihood of delays and inconsistencies.

Given the repetitive nature of the audit and the growing demands for timely, reliable, and traceable quality information, the need to optimise and automate this process became evident. Manual execution not only consumes valuable time from quality controllers but also limits the organisation's ability to consolidate data, monitor trends, and respond proactively to deviations. As the volume of audits increases, these limitations become progressively more critical, reinforcing the necessity for a structured improvement approach.

As an initial step toward process improvement, this study focuses on identifying the stages at which inefficiencies occur. To achieve this, Business Process Model and Notation (BPMN) is used to model the current (AS-IS) audit process. BPMN provides a clear and standardised visual representation of the workflow, enabling a comprehensive understanding of task sequencing, decision points, data flows, and interactions between actors. By mapping the process using BPMN, it became possible to systematically analyse the workflow, identify redundant or non-value-added activities, and establish a solid foundation for subsequent automation and digitalisation initiatives.

The Standard Work Audit ensures that operations on the production line follow the defined methods and work instructions. Currently, this audit includes the following steps:

- Selection of the production line to be audited, based on the audit plan and the availability of Standard Work documentation.
- Preparation of the audit, including printing the physical audit sheet that will be used for on-site data collection.
- Movement to the production line and initiation of the audit.
- Direct observation of operators' activities during three consecutive production cycles, assessing compliance with the Standard Work Sheet, including task sequence, task allocation, and adherence to defined work instructions.
- Manual recording of audit results on the physical audit sheet, with each parameter classified according to predefined outcomes (e.g., OK or NOK).
- Evaluation of audit results, with immediate awareness and clarification provided to operators in cases where non-compliances are identified.
- Transcription of the collected data into a digital template, ensuring consistency between physical and digital records.
- Storage of each audit report in a shared Google Drive folder as a separate PDF file identified by its respective audit number.
- Entry of audit data into the general audit database, enabling historical storage and future analysis.
- Dissemination of the audit report, typically shared with relevant stakeholders via email.

The critical stages of the current (AS-IS) process model are highlighted in Fig. 1, particularly those associated with inefficiencies and data management limitations.

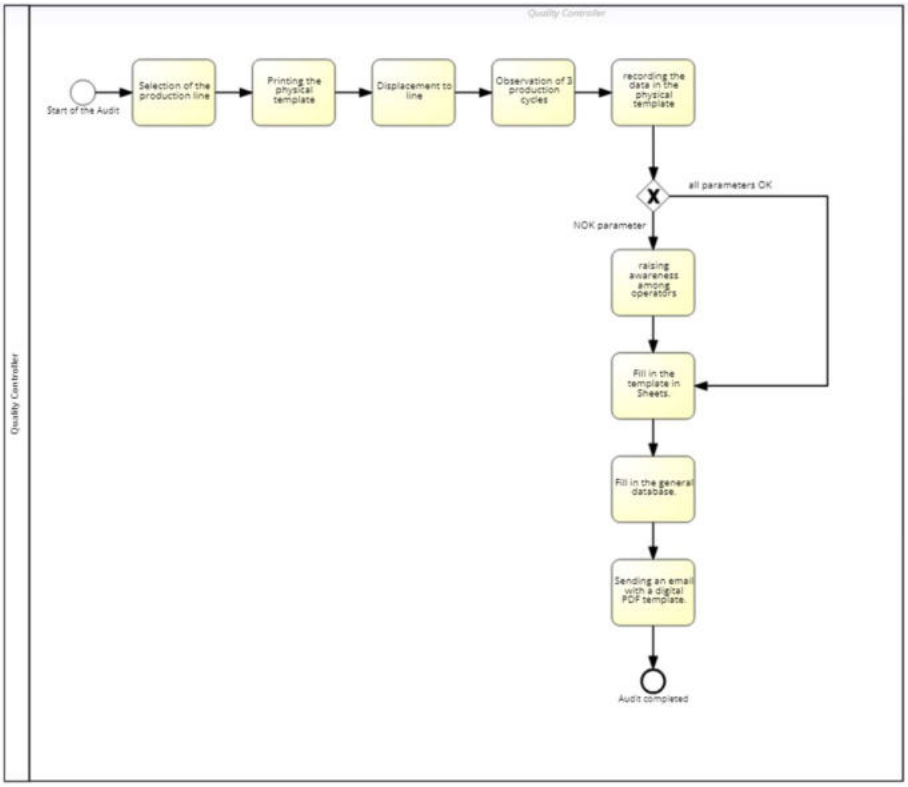


Fig. 1. AS-IS model.

The BPMN modelling of the Standard Work Audit process enabled a detailed visualisation of the existing workflow structure. The analysis of the AS-IS model revealed several operational characteristics and inefficiencies embedded within the current process. These findings relate primarily to (i) the distribution and repetition of activities throughout the Standard Work Audit workflow, (ii) the multi-step and manual nature of data registration, and (iii) the absence of integrated mechanisms for consolidating and analysing audit data.

To overcome these limitations, a TO-BE process model was designed to support a more integrated and efficient audit workflow as illustrated in Fig. 2 In this future-state process, audit records are entered directly into the AppSheet application and automatically stored in a central database, eliminating the need for manual transcription and improving data availability, traceability, and consistency. Additionally, once the audit record is submitted, the corresponding email notification is automatically sent, thereby streamlining communication with the relevant stakeholders and reducing delays in the reporting process.

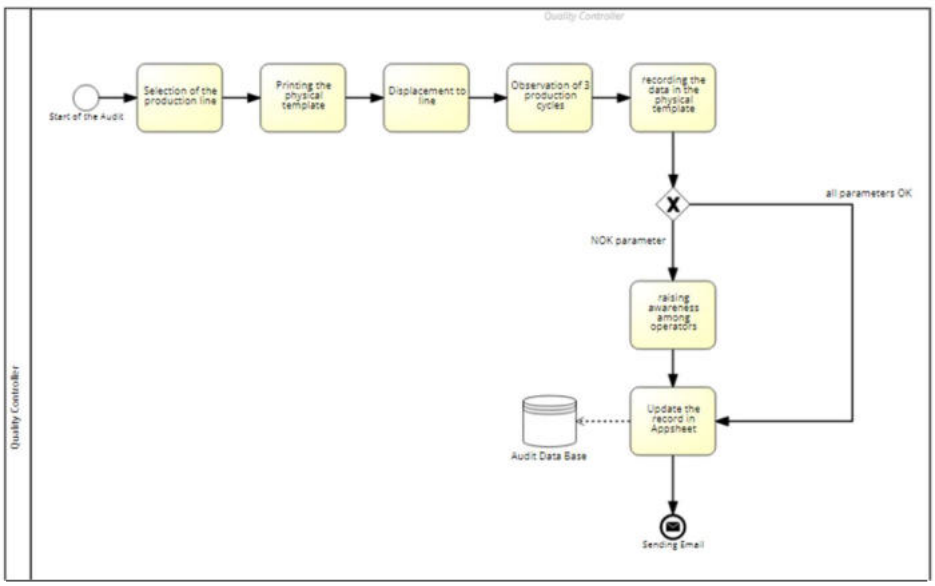


Fig. 2. TO-BE model.

In order to address the limitations identified in the current audit process and support its digitalisation, a solution was developed to improve data collection, storage, and analysis. As highlighted in the AS-IS analysis, the existing process relied heavily on manual data registration, fragmented information storage, and limited analytical capabilities, which constrained the organisation's ability to monitor performance and support decision-making.

A low-code application was then developed using AppSheet, aiming to centralise audit data within a single platform and streamline the overall audit workflow. The application was designed to integrate the different types of audits performed in the Quality Department, ensuring a structured and standardised approach to data registration.

The developed solution enables users to register audit data through a digital form, automatically store information in a central database, and support real-time analysis through an integrated dashboard. The main interface of the application, which provides access to the different audit types and functionalities, is illustrated in Fig. 3.

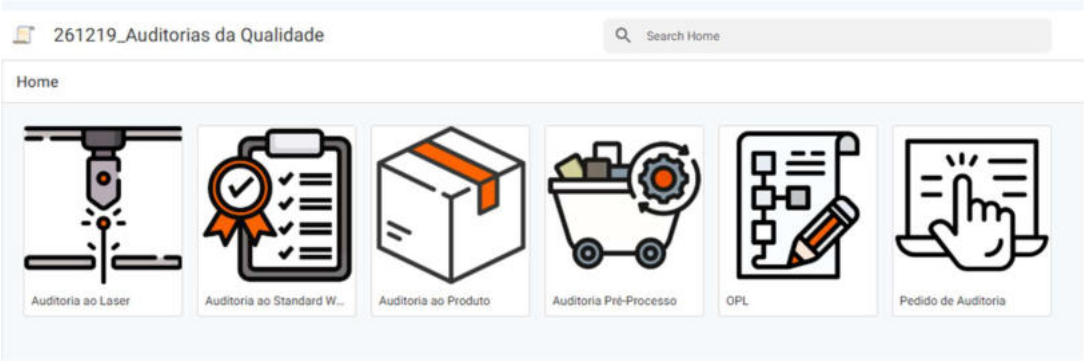


Fig. 3. Home page of the developed app.

Audit records are submitted through a form integrated into the application, which includes validation rules such as mandatory fields. As shown in Fig. 4, when a parameter is marked as NOK, the associated comment becomes mandatory, preventing incomplete data entry.

Fig. 4. Form of a new register.

In addition to the application, a data analysis dashboard was developed, as shown in Fig. 5. Connected to the same data source, it provides continuously updated records and enables a consolidated analysis of audit results. The dashboard allows users to monitor key indicators, apply filters by relevant parameters, and identify trends or recurring non-conformities, thereby improving process visibility and supporting data-driven decision-making.

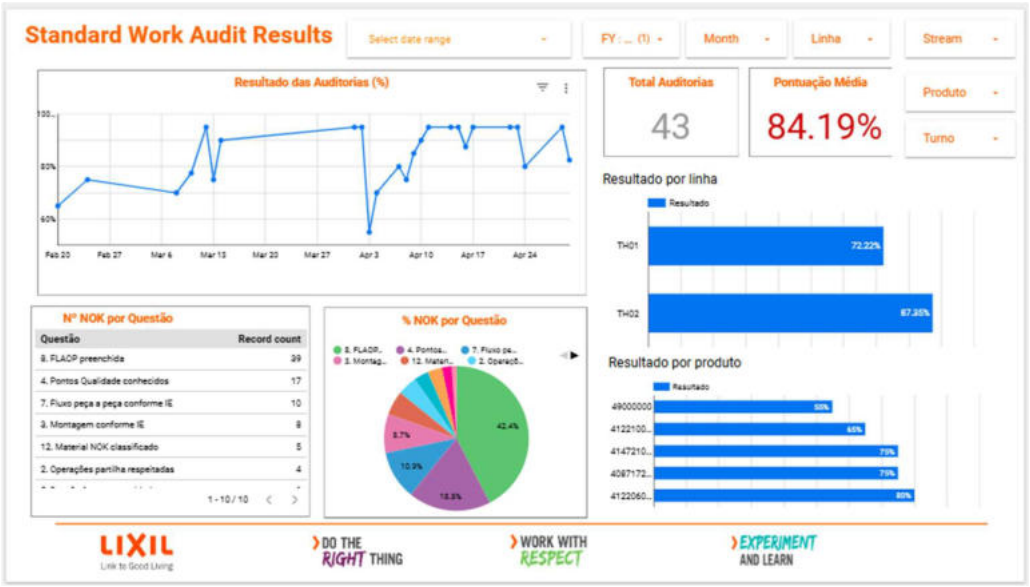


Fig. 5. Dashboard for data analysis

As illustrated in Figure 5, the dashboard enables the analysis of key performance indicators (KPIs), including the total number of registrations filtered by date, the average audit result, the most frequently failing parameters across the dataset, and the percentage of non-conformities (NOK) associated with each evaluated audit question. Furthermore, it supports the assessment of performance at both production line and product levels. These analytical capabilities provide robust and data-driven insights, supporting informed decision-making and facilitating the identification of priority areas for corrective action.

CONCLUSIONS

This study highlights the important role of BPMN-based process modelling in supporting the digitalisation of quality management processes. The AS-IS analysis enabled a systematic identification of several inefficiencies, namely information dispersion, lack of standardisation, and the absence of real-time monitoring mechanisms. By structuring and making process flows explicit, BPMN proved to be fundamental in guiding their redesign and promoting a more integrated and data-driven approach.

Building on this diagnostic phase, an AppSheet-based application was developed to operationalise the defined TO-BE model. This solution enabled the automation of data recording, standardised the way information is entered, and centralised all relevant data within a single system. As a result, there was a reduction in manual effort and the likelihood of human error, alongside a significant improvement in data consistency, traceability, and availability.

Furthermore, the integration of functionalities such as automated calculations, notifications, and near real-time dashboards contributed to increased process visibility and supported continuous monitoring of audit performance. This advancement translates into faster and more informed decision-making, in line with the principles of Industry 4.0 and Business Process Management. Rather than merely converting existing practices into digital formats, the developed solution promotes a structural transformation in the way information is handled and utilised within the organisation.

The results align with findings from several authors who emphasise that digitalisation of quality processes is a prerequisite for efficiency, control, and operational responsiveness. Yu et al. (2021) highlight that the absence of integrated data-collection systems prevents early deviation detection and impairs an organisation's ability to respond rapidly to failures. Similarly, Bousdekis et al. (2023) argue that manual data entry and the lack of analytical dashboards are typical limitations of traditional quality systems and constitute major barriers to the effective implementation of Industry 4.0 and Quality 4.0 practices.

Although the redesigned process substantially improves the structure, traceability, and consolidation of audit information, certain limitations remain. Despite the digitalisation of data processing and reporting, the initial data collection continues to rely on manual input, as quality controllers currently do not have access to mobile digital devices that would allow the direct recording of audit results at the point of inspection. As a result, some degree of manual transcription persists, which may still introduce delays or isolated inaccuracies. This limitation reflects a broader infrastructural constraint rather than a methodological one and should be addressed in future stages of the digital transformation.

Additionally, while this study focuses exclusively on the audit processes performed within the Assembly Department, similar audits are conducted in other areas of the organisation, such as Supplier Quality Management,

and Surface Treatment. These departments employ data-collection procedures that follow the same manual and fragmented logic identified in the mapped processes. Consequently, the benefits achieved through the development of the AppSheet solution are expected to be equally relevant in these contexts. Future improvement initiatives should therefore consider replicating or adapting the digital application to other audit processes, promoting standardisation across departments, and maximising the impact of digitalisation on the organisation's overall quality-management system.

Overall, the results demonstrate that the combination of process modelling and low-code digital solutions, such as AppSheet, represents an effective approach to bridging the gap between conceptual analysis and practical implementation. This integration is particularly relevant in industrial contexts, as it supports continuous improvement, strengthens quality management systems, and fosters a data-driven organisational culture.

ACKNOWLEDGMENTS

This work was supported by CIDMA (https://ror.org/05pm2mw36) under the Portuguese Foundation for Science and Technology (FCT, https://ror.org/00snfqn58), projects UID/04106/2025 (https://doi.org/10.54499/UID/04106/2025) and UID/PRR/04106/2025 (https://doi.org/10.54499/UID/PRR/04106/2025.

REFERENCES

Aguilar-Savén, R. S. (2004). Business process modelling: Review and framework. International Journal of Production Economics, 90(2), 129–149. https://doi.org/10.1016/j.ijpe.2003.12.002

Antony, J., Sony, M., Dempsey, M., Brennan, A., Farrington, T., & Cudney, E. A. (2021). A review of Lean Six Sigma for manufacturing quality. The TQM Journal, 33(2), 438–470. https://doi.org/10.1108/TQM-12-2019-0303

Bousdekis, A., Magoutas, B., Apostolou, D., & Mentzas, G. (2023). Quality 4.0: Data-driven quality management in Industry 4.0 environments. Computers & Industrial Engineering, 176, 109157. https://doi.org/10.1016/j.cie.2023.109157

Broday, E. E. (2022). The evolution of quality: from inspection to quality 4.0. International Journal of Quality and Service Sciences, 14(3), 368–382. https://doi.org/10.1108/IJQSS-09-2021-0121

Carvalho, A. M., Sampaio, P., Rebentisch, E., & Saraiva, P. (2021). 35 years of excellence, and perspectives ahead for excellence 4.0. Total Quality Management and Business Excellence, 32(11–12), 1215–1248. https://doi.org/10.1080/14783363.2019.1691915

Chiarini, A., & Kumar, M. (2022). What is Quality 4.0? An exploratory sequential mixed methods study of Italian manufacturing companies. International Journal of Production Research, 60(16), 4890–4910. https://doi.org/10.1080/00207543.2021.1942285

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). Fundamentals of business process management (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-56509-4

Gao, D., & Fagerholm, F. (2026). An Investigation of Low-Code Development Adoption in a Finnish IT Consulting Firm. Lecture Notes in Computer Science, 16361 LNCS, 169–185. https://doi.org/10.1007/978-3-032-12089-2_11

Google Cloud. (2023). LIXIL builds a data-driven culture with Google Cloud. https://cloud.google.com/customers/lixil

GROHE. (s.d.). Sobre a empresa. https://www.grohe.pt/pt_pt/sobre-a-grohe/sobre-a-empresa.html

Imai, M. (1986). Kaizen: The key to Japan's competitive success. McGraw-Hill.

Kagermann, H., Wahlster, W., & Helbig, J. (2013). Recommendations for implementing Industry 4.0. Acatech. https://en.acatech.de/publication/recommendations-for-implementing-industry-4-0/

Lourenço, L., Silva, A., & Figueiredo, R. (2022). Manufacturing data analytics for quality assurance. CEUR Workshop Proceedings, 3306, Paper 22. https://ceur-ws.org/Vol-3306/paper22.pdf

Ludbrook, F., et al. (2025). Business models for sustainable innovation in Industry 4.0: The role of digital platforms and low-code solutions. Journal of Manufacturing Technology Management. Advance online publication.

Rother, M. (2010). Toyota kata: Managing people for improvement, adaptiveness and superior results. McGraw-Hill.

Rožanec, A., et al. (2022). Automated visual inspection in manufacturing using artificial intelligence. Expert Systems with Applications, 198, 117080. https://doi.org/10.1016/j.eswa.2022.117080

Saha, A., & Dey, S. (2022). A systematic review of low-code and no-code development platforms. Journal of Systems and Software, 190, 111351. https://doi.org/10.1016/j.jss.2022.111351

Santos, G., et al. (2022). Internal quality audits in industrial environments: Practices and challenges. Journal of Manufacturing Technology Management. https://doi.org/10.1108/JMTM-09-2021-0287

Silva, M., Alvelos, H., & Xambre, A. R. (2025). Quality Problems Reporting: Digitalization of the processes in an Automotive Company. Procedia Computer Science, 253, 737–744. https://doi.org/10.1016/j.procs.2025.01.135

Stravinskiene, I., & Serafinas, D. (2020). The Link between Business Process Management and Quality Management. Journal of Risk and Financial Management 2020, Vol. 13, Page 225, 13(10), 225. https://doi.org/10.3390/JRFM13100225

Womack, J. P., & Jones, D. T. (1996). Lean thinking: Banish waste and create wealth in your corporation. Simon & Schuster.

Yu, W., Xu, X., & Liu, F. (2021). Digitalization in Industry 4.0 and real-time manufacturing control. Journal of Manufacturing Systems, 60, 266–280. https://doi.org/10.1016/j.jmsy.2021.06.010

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Aggregation Sequence in Multiple-Inspector Sampling Plans: Preliminary Outcomes and Practical Implications

Domenico A. Maisano[1,*], Lucrezia Ferrara[1] and Fiorenzo Franceschini[1]

[1] Department of Management and Production Engineering (DIGEP), Politecnico di Torino, Italy; [*] domenico.maisano@polito.it

Abstract: **In safety-critical or high-value sectors – such as aerospace, defence, nuclear production, safety devices and luxury manufacturing – incorrect conformity assessments can have unacceptable consequences. When 100% inspection is impractical and inspections are affected by non-negligible error, redundant inspection of a subset of sampled units by multiple independent inspectors becomes an attractive option. In multiple-inspector sampling plans for attributes, the inspectors' unit-level assessments can be arranged in an *n*·*m* matrix, where *n* is the number of sampled units and *m* is the number of inspectors. The sequence used to aggregate those assessments into a final *lot-disposition decision* (LDD) is crucial; two sequences are considered here. In the *rows-first* sequence, each inspector first applies a classical sampling-plan rule to the sampled units (i.e., acceptance if the observed composition of conforming and defective units satisfies the relevant reference thresholds, and rejection otherwise), generating a local LDD; the local LDDs are then aggregated by *majority*. In the *columns-first* sequence, the inspectors' assessments are first aggregated unit by unit, again by *majority* in the case studied, and the same sampling-plan rule is then applied to the resulting global assessments.**

This paper investigates the practical consequences of the aggregation sequence on the severity and discriminatory power of the resulting multiple-inspector sampling plan, as described by the relevant *operating-characteristic* (OC) curve. A broad set of operational configurations is examined by varying the sampling-plan parameters, the number of inspectors and their propensity to make inspection errors. Stability is assessed by comparing the multiple-inspector OC curves with the "ideal" OC curve of a corresponding sampling plan in the absence of inspection errors. The analysis shows that the two aggregation sequences can yield substantially different and sometimes counterintuitive outcomes, even when all input conditions are unchanged. The columns-first sequence is generally more stable and robust, and its performance typically improves as the number of inspectors increases. By contrast, the rows-first sequence often produces excessive severity, and – under some conditions – a larger group of inspectors can even worsen the deviation from the ideal benchmark. These findings can be interpreted through *Condorcet's Jury Theorem*, according to which majority voting amplifies the alternative that is individually more likely.

Keywords: Sampling plan; Multiple inspectors; Inspection error; Majority criterion

I. Introduction

In critical, high-value and highly customized contexts – such as aerospace, defence, nuclear production, automotive safety devices and luxury manufacturing – lot-acceptance decisions must be highly reliable. A wrong conformity assessment may have serious consequences, so organizations sometimes use more than one inspector on the same units. Although redundant inspection increases effort, it can improve decision reliability when individual assessments are not perfectly accurate (Chun, 2009). Moreover, the literature shows that conformity assessments are often subjective and prone to inspection error, whether inspections are manual or automated; differences in training, fatigue, procedures, instrumentation and familiarity with the product make it unrealistic to assume perfect assessment of every unit, especially for visual, aesthetic or otherwise ambiguous attributes (Cymek, 2024). Another practical issue is that 100% inspection is not always feasible; this happens, for example, when inspection is destructive, very time-consuming or prohibitively costly. Airbag systems are an example: a complete inspection may require dismantling and detailed examination of the device, which prevents reassembly and makes full inspection of all units impossible. Similar issues arise in many critical or expensive products; in such cases, statistical acceptance sampling is used: only a subset of sampled units is inspected and then a lot-level decision is derived.

In the classical attribute-sampling setting, a sampling plan is defined by several parameters: the number of sampled units, which may be fixed or may vary depending on how the inspection evolves, and a set of reference thresholds used to determine lot acceptance or rejection. The *n* sampled units may be drawn from a lot of *N* units either all at once or in successive stages, depending on the specific sampling-plan scheme. The final decision to accept or reject the lot depends on the comparison between the number of conforming and defective units detected within the sampled units and the relevant reference thresholds. This framework is well established in statistical quality control because it balances customer protection and inspection effort when a full inspection is either impossible or inefficient (Montgomery, 2019; Schilling and Neubauer, 2009; ISO/DIS 2859-1:2025, 2025). The problem becomes more complex when the same sampled units are inspected independently by *m* different inspectors. Their binary unit-level assessments can be organized into a conformity-assessment matrix of size *n*·*m*, where the columns correspond to sampled units (u_1, u_2, …, u_n) and the rows correspond to inspectors (i_1, i_2, …, i_m). The central question is how to aggregate those multiple assessments into a single global *lot-disposition decision* (LDD); the answer is not unique. Two aggregation sequences can be adopted. In the first one, called *rows-first*, each inspector is treated separately; every inspector applies the sampling-plan rule to the sampled units as they assessed them, thereby generating a local LDD. Those *m* local decisions are then aggregated by a column-wise rule, such as *majority* or *unanimity*, to obtain the final global LDD. In the second sequence, called *columns-first*, the order is reversed: the *m* inspectors' judgments on each sampled unit are first aggregated to produce one global assessment for that unit; after all sampled units have been assessed globally, the classical sampling-plan rule is applied once to determine the final LDD. Fig. 1 exemplifies the two sequences (Franceschini et

al., 2025; Deming, 2018); even when based on the same input data (i.e., the same assessment matrix), the two aggregation sequences do not necessarily yield the same final global LDD.

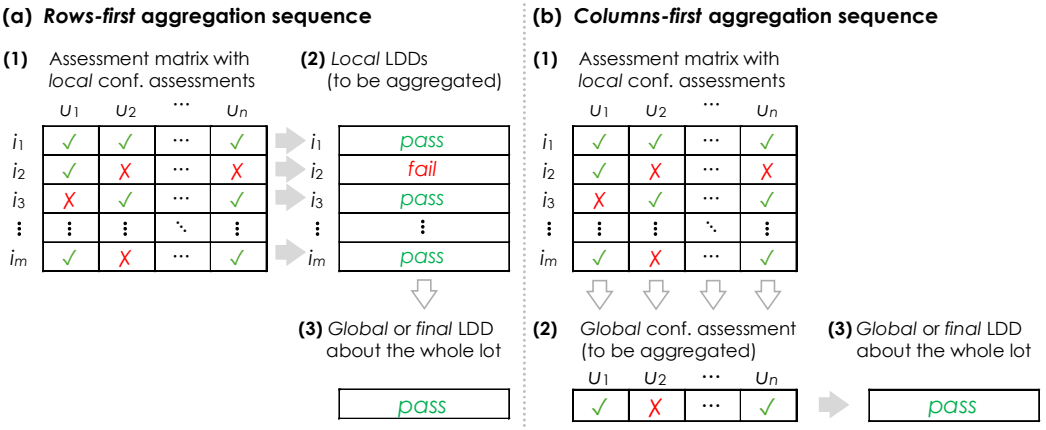


Fig. 1. Conceptual example of a conformity-assessment matrix with aggregation sequences (a) *rows-first* and (b) *columns-first*, used to obtain the final LDD for the lot. The symbols "✓" and "✗" denote units classified as conforming and defective, respectively. The column-wise aggregation step may rely on different criteria, such as *majority* or *unanimity* (Franceschini et al., 2025).

Previous work by Maisano et al. (2025) formalized multiple-inspector sampling plans statistically, by considering both aggregation sequences and different column-wise aggregation rules, notably *majority* and *unanimity*. That work showed that majority aggregation can produce *operating-characteristic* (OC) curves relatively close to those of an "ideal" sampling plan with a single error-free inspector, whereas unanimity is often too severe because even a small number of false-defective classifications makes unanimous agreement on conformity unlikely. What remains less clear, however, is the role of the aggregation sequence itself. If the same inspectors, the same sampled units, the same sampling-plan parameters and the same error propensities are kept fixed, can the final performance still change substantially, simply because the aggregation is carried out in a different sequence?

The present study addresses that question, with the aim of investigating whether the aggregation sequence affects the stability of multiple-inspector sampling plans and, if so, under what operating conditions the effect becomes most relevant. Particular attention is devoted to the resulting OC curves, as they summarize the severity and discriminatory power of the sampling plan. For the sake of simplicity, the analysis is restricted to *single-sampling plans* (SSPs). Moreover, to keep the investigation focused on practically meaningful designs, only *majority* aggregation is considered in the column-wise aggregation step; *unanimity* is excluded because prior evidence showed that, under non-negligible inspection error, it is justified only in extremely critical applications (Maisano et al., 2025; Duffuaa and El-Ga'aly, 2015). The work is guided by three research questions:

- *RQ1*: Which aggregation sequence (columns-first or rows-first) is generally more robust?
- *RQ2*: Under what operating conditions do the differences between the two sequences become most evident?
- *RQ3*: What theoretical lens best explains the observed differences?

To answer these questions, a broad set of SSP configurations is examined by varying the sampling-plan parameters and the inspectors' error rates. The results are then interpreted through *Condorcet's Jury Theorem* (CJT), a classical theorem in *social choice theory* according to which majority voting improves collective accuracy only when individual decisions are more likely to be correct than incorrect. As will be shown, that theorem offers a compact and convincing explanation for the empirical patterns that emerge (Condorcet, 1785; Grofman et al., 1983; List and Goodin, 2001; Franceschini et al., 2022).

The remainder of the paper is organized as follows. Section II recalls the theoretical model of attribute SSPs, first in the ideal single-inspector case and then in the multiple-inspector setting with inspection errors. Section III describes the methodological design of the comparison. Section IV presents the OC-curve results and interprets them through the CJT. Section V concludes by summarizing the practical implications, limitations and directions for future research.

II. Theoretical Background on SSPs

A. Ideal SSP

A classical SSP for attributes is defined by the sample size n and the acceptance number c. After a random sample of n units is drawn from a lot of size N, the number of defectives d observed in the sample is counted. If $d \leqslant c$, the lot is accepted; if $d > c$, it is rejected. In the standard case in which isolated lots with $n \ll N$ or trains of lots are considered, the probability of accepting the lot can be approximated by a binomial model. The corresponding Type-B OC curve is given by (Montgomery, 2019; Schilling and Neubauer, 2009):

$$P_a = \sum_{d=0}^{c} \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d}, \tag{1}$$

which represents the cumulative probability of observing at most c defectives in a sample of size n when the lot defectiveness is p. More generally, when the binomial approximation is not applicable, the hypergeometric model should be

used, yielding the Type-A OC curve. In this paper, however, only the binomial approximation is considered, as the aim is to isolate the effects of inspection error and aggregation sequence without overcomplicating the analytical models (Montgomery, 2019).

The OC curve is the fundamental tool for evaluating the discriminatory power of a sampling plan, that is, its ability to distinguish acceptable lots from unacceptable ones. Two reference points are particularly important. The first is the *Acceptable Quality Level* (*AQL*), a relatively small defectiveness that the supplier may produce while still expecting acceptance with relatively high probability; the complementary probability of rejection at the AQL is the "producer's risk", $\alpha = 1 - P_a(p = AQL)$. The second point is the *Lot Tolerance Percent Defective* (*LTPD*), a relatively high defectiveness at which the customer expects rejection with relatively high probability; the acceptance probability at that point is the "consumer's risk", $\beta = P_a(p = LTPD)$. In practice, the values of *n* and *c* are selected so that the resulting OC curve reflects an acceptable compromise between supplier and customer interests (Deming, 2018; Montgomery, 2019).

B. Modelling inspection errors

The classical SSP implicitly assumes a single inspector who makes no inspection error. Unfortunately, real inspections are rarely error-free, especially when they involve subjective judgments or subtle quality characteristics. Two error types are particularly relevant (Kenett et al., 2023; Montgomery, 2019):

- *False defectives*: conforming units incorrectly classified as defective; the corresponding probability is denoted by *a*.
- *False conformings*: defective units incorrectly classified as conforming; the corresponding probability is denoted by *b*.

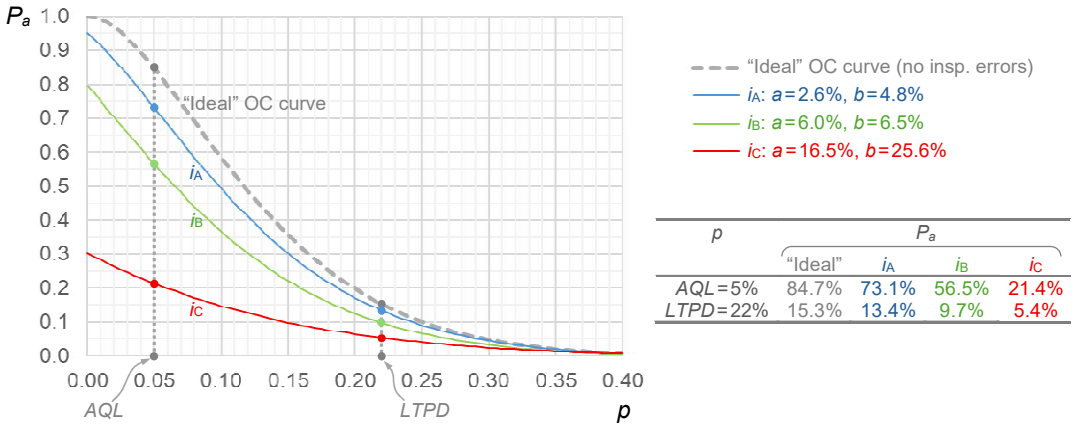
Empirical studies suggest that both probabilities are often only a few percentage points in industrial practice, with false-conforming rates sometimes slightly higher than false-defective rates. Even so, small inspection-error probabilities can materially distort the performance of a sampling plan because the final LDD is based on a small number of observations (Burke et al., 1995; Kenett et al., 2023). If the true defectiveness is *p*, an inspector with error rates *a* and *b* does not perceive *p* directly. Instead, the inspector perceives an apparent defectiveness *p'*, defined by:

$$p' = (1 - p)\cdot a + p\cdot(1 - b)\,. \tag{2}$$

Equation (2) states that a sampled unit is classified as defective either because it is truly defective and correctly identified as such, or because it is truly conforming but falsely classified as defective. Consequently, the lot acceptance probability in the presence of inspection error is obtained by replacing *p* with *p'* in the ideal SSP formula, as shown in (3) (Ferrell Jr and Chhoker, 2002; Jamkhaneh et al., 2011; Gadrich and Bashkansky, 2016):

$$P_a' = \sum_{d=0}^{c} \binom{n}{d} \cdot (p')^d \cdot (1 - p')^{n-d}\,. \tag{3}$$

The "distorted" OC curve implied by (3) generally lies below the ideal OC curve at low values of *p*, because false defectives make good lots appear worse than they are, and above the ideal OC curve at high values of *p*, because false conformings make bad lots appear better than they are. The net effect is a loss of discriminatory power. Fig. 2 illustrates these distortions for a SSP with *n* = 14 and *c* = 1. Even moderate error rates can reduce acceptance probabilities at low defectiveness by 10-20%; when error rates (*a* and *b*) approach 20%, the OC curve may flatten dramatically, yielding a sampling plan whose practical meaning is severely compromised (Burke et al., 1995; Ferrell Jr and Chhoker, 2002). A further difficulty is that estimating *a* and *b* in practice is not straightforward; reliable estimates require a "ground truth", such as re-inspection by a reference method or a superior inspector, and they may vary over time with fatigue, training, product type and operating conditions. Many organizations therefore ignore these parameters and design sampling plans as if inspections were error-free. This difficulty motivates the search for configurations that remain reasonably stable even when individual inspection errors cannot be estimated accurately (Burke et al., 1995; Jaraiedi and Iskander, 1989; Franceschini et al., 2025; Maisano et al., 2025).



p	P_a			
	"Ideal"	i_A	i_B	i_C
AQL = 5%	84.7%	73.1%	56.5%	21.4%
LTPD = 22%	15.3%	13.4%	9.7%	5.4%

Fig. 2. Comparison of Type-B OC curves for a SSP with a single inspector and parameters *n* = 14 and *c* = 1, in the absence of inspection errors (ideal OC curve) and in the presence of inspection errors. Three hypothetical inspectors (i_A, i_B and i_C) with different combinations of the *a* and *b* error rates are considered.

C. SSPs with multiple inspectors

Consider now the case in which the same sample units are inspected independently by multiple inspectors. Their assessments are stored in a conformity-assessment matrix like the one in Fig. 1. The rationale for using multiple inspectors is that independent judgments can cross-check one another and reduce the effect of individual misclassifications, albeit at a higher inspection cost; this trade-off is especially attractive in safety-critical or high-value applications. In what follows, inspectors are assumed to work independently, so that one inspector's assessment does not influence another's (Knight and Myers, 1993; Aurum et al., 2002; Chandra and Schall, 1998). The modelling problem is to derive the OC curve of the resulting multiple-inspector SSP while accounting for both the inspectors' error rates and the aggregation sequence used to transform local assessments into a global decision (Maisano et al., 2025).

D. Rows-first aggregation sequence

In the rows-first aggregation sequence, each *i*-th inspector evaluates the same *n* sample units and applies the classical SSP rule to the sample as perceived by that inspector; the result is a local LDD. The final global LDD is then obtained by aggregating the *m* local LDDs by majority (Franceschini et al., 2025; Maisano et al., 2025). To model this case, the apparent defectiveness perceived by each inspector must first be specified. For the *i*-th inspector, whose false-defective and false-conforming probabilities are a_i and b_i, the apparent defectiveness is given by:

$$p_i' = (1-p)\cdot a_i + p\cdot(1-b_i)\,. \tag{4}$$

Using that quantity in the binomial SSP model yields the probability that the *i*-th inspector finds at most *c* defectives in the sample, namely the local lot-acceptance probability shown in (5):

$$P_{a,\,i} = \sum_{d=0}^{c} \binom{n}{d} \cdot (p_i')^d \cdot (1-p_i')^{\,n-d}\,. \tag{5}$$

The global probability of accepting the lot under rows-first aggregation is then obtained by aggregating the *m* local LDDs through the majority criterion:

$$P_{a,\,\mathrm{agg}} = \sum_{d=0}^{n}\left\{ \sum_{s=\left\lfloor \frac{m}{2}\right\rfloor+1}^{m}\left[\sum_{A\subseteq\{1,\,\ldots,\,m\},\,|A|=s}\left\{\left[\prod_{i\in A}\left(\sum_{k=n-c}^{n}\left\{ \sum_{j=\max[0,\,k-(n-d)]}^{\min(k,\,d)}\left\{\left[\binom{d}{j}\cdot(1-b_i)^{j}\cdot {b_i}^{d-j}\right]\cdot\left[\binom{n-d}{k-j}\cdot {a_i}^{(k-j)}\cdot(1-a_i)^{(n-d)-(k-j)}\right]\right\}\right\}\right)\right]\right.\right.$$
$$\left.\left.\cdot\left[\prod_{i\notin A}\left(1-\left[\sum_{k=n-c}^{n}\left\{ \sum_{j=\max[0,\,k-(n-d)]}^{\min(k,\,d)}\left\{\left[\binom{d}{j}\cdot(1-b_i)^{j}\cdot {b_i}^{d-j}\right]\cdot\left[\binom{n-d}{k-j}\cdot {a_i}^{(k-j)}\cdot(1-a_i)^{(n-d)-(k-j)}\right]\right\}\right\}\right]\right)\right]\right\}\cdot\binom{n}{d}\cdot p^{d}\cdot(1-p)^{n-d}\right\}. \tag{6}$$

Equation (6) can be interpreted as the probability that the majority of inspectors, each subject to individual inspection error rates a_i and b_i, identifies no more than *c* defective units, conditional on the sample containing exactly *d* defectives and accounting for all possible values of *d*. Here, *A* denotes a possible subset of *s* inspectors (out of *m*) yielding a local acceptance LDD, while *k* and *j* are dummy summation variables. This aggregation scheme is mathematically complex because all inspectors evaluate the same sampled units, thereby introducing a common-sample constraint. Further details on the step-by-step derivation of (6) are provided in Maisano et al. (2025).

E. Columns-first aggregation sequence

In the columns-first sequence, aggregation takes place in two stages. First, the *m* assessments referring to the same sampled unit are aggregated into one global conformity assessment for that unit. Second, the classical SSP rule is applied to the resulting *n* global assessments. Under majority aggregation, a sampled unit is globally classified as conforming if at least *q* of the *m* inspectors judge it conforming. The threshold *q* is given by (Maisano et al., 2025):

$$q = \left\lfloor \frac{m}{2} \right\rfloor + 1\,. \tag{7}$$

For example, when *m* = 5, majority requires *q* = 3. A column such as [✓, ✓, ✓, ✗, ✗]^T is therefore globally classified as conforming. Once unit-level majority aggregation has been applied to all sampled units, the probability of lot acceptance is given by:

$$P_{a,\,\mathrm{agg}} = \sum_{d=0}^{c}\binom{n}{d}\cdot\left[p'_{\mathrm{agg}}\right]^{d}\cdot\left[1-p'_{\mathrm{agg}}\right]^{n-d}\,. \tag{8}$$

The term p'_{agg} in (8) is a multiple-inspector apparent defectiveness, which can be expressed as a function of two aggregated error rates, a_{agg} and b_{agg}, through (9):

$$p'_{\mathrm{agg}} = (1-p)\cdot a_{\mathrm{agg}} + p\cdot\left(1-b_{\mathrm{agg}}\right). \tag{9}$$

These error rates characterize a hypothetical "aggregate inspector" whose decisions result from majority aggregation of the individual inspectors' assessments. In other words, a_{agg} is the probability that a truly conforming unit is classified as defective by the majority, while b_{agg} is the probability that a truly defective unit is classified as conforming by the majority. Their exact expressions are given by (10) and (11):

$$1-a_{\mathrm{agg}} = \sum_{s=q}^{m}\left\{\sum_{\substack{A\subseteq\{1,\ldots,m\}\\ |A|=s}}\left[\prod_{i\in A}(1-a_i)\right]\cdot\left[\prod_{i\notin A}a_i\right]\right\} \quad\Rightarrow\quad a_{\mathrm{agg}} = 1-\sum_{s=q}^{m}\left\{\sum_{\substack{A\subseteq\{1,\ldots,m\}\\ |A|=s}}\left[\prod_{i\in A}(1-a_i)\right]\cdot\left[\prod_{i\notin A}a_i\right]\right\}, \tag{10}$$

$$b_{\text{agg}} = \sum_{s=q}^{m} \left\{ \sum_{\substack{A \subseteq \{1,\dots,m\} \\ |A|=s}} \left[\prod_{i \in A} b_i \right] \cdot \left[\prod_{i \notin A} (1 - b_i) \right] \right\}. \tag{11}$$

In (10), A denotes a possible subset of s inspectors (out of m) who classify a truly conforming unit as defective; similarly, in (11), A denotes a possible subset of s inspectors (out of m) who classify a truly defective unit as conforming. A step-by-step derivation of (10) and (11) is provided in Maisano et al. (2025). If all inspectors are error-free, then $a_{\text{agg}} = b_{\text{agg}} = 0$, so that $p'_{\text{agg}} = p$ and (8) collapses to the ideal SSP model in (1). This limiting case confirms that the multiple-inspector model is consistent with the classical model in the absence of inspection errors (Maisano et al., 2025).

III. Methodology

The empirical analysis consists of constructing and comparing a large set of OC curves for multiple-inspector SSPs affected by inspection errors. For each multiple-inspector configuration, the corresponding ideal Type-B OC curve with the same (n, c) parameters and no inspection error is used as a benchmark. The closer the multiple-inspector OC curve lies to that benchmark, the more stable and robust the design is considered to be.

Four factors are varied. First, the aggregation sequence is changed, comparing rows-first and columns-first. Second, the SSP parameters are varied by considering $n \in \{4, 9, 14\}$ and $c \in \{0, 1, 2\}$, which produces nine distinct (n, c) combinations ranging from relatively severe to relatively indulgent SSPs. Third, the number of inspectors is varied by considering groups of size $m = 3$, 5 and 7. Fourth, two error-rate scenarios are introduced: a low-error scenario *L*, meant to represent inspectors with strong familiarity with the product, and a high-error scenario *H*, intended as an extreme stress test.

A realistic set of seven hypothetical inspectors, denoted i_1 to i_7, is used throughout the analysis. In the scenario *L*, the values of a_i and b_i were generated from a normal distribution with mean 4% and standard deviation 2%, in line with empirical evidence that inspection errors in industrial settings are often only a few percent. In the scenario *H*, the values were generated from a normal distribution with mean 18% and standard deviation 4%. Those higher values are intentionally severe, because the objective is to test the model under highly unfavourable conditions. The three inspector subsets correspond to the first 3, the first 5, and all 7 inspectors listed in Table 1.

Table 1. Error rates (a_i and b_i) of seven hypothetical inspectors (i_1 to i_7) across two error-rate scenarios: low error rates (*L*), assuming that inspectors have strong familiarity with the inspected units, and high error rates (*H*), assuming only partial familiarity. Inspectors are grouped into three subsets of size m = 3, 5 and 7.

Inspector	*L* scenario		*H* scenario	
	a_i	b_i	a_i	b_i
i_1	5.3%	2.6%	17.4%	13.2%
i_2	4.3%	4.2%	13.4%	23.0%
i_3	6.5%	6.0%	14.1%	16.8%
i_4	5.5%	3.6%	13.6%	22.9%
i_5	2.9%	4.7%	13.2%	19.7%
i_6	4.1%	2.4%	15.5%	21.3%
i_7	3.1%	3.0%	25.5%	15.6%

group of m=3 inspectors

group of m=5 inspectors

group of m=7 inspectors

For each of the nine (n, c) configurations, thirteen OC curves are therefore constructed: one ideal OC curve and twelve multiple-inspector OC curves obtained from the combination of two aggregation sequences, three values of m and two error-rate scenarios. The total number of OC curves analysed is therefore given by:

$$(3 \cdot 3) \cdot [1 + 2 \cdot (2 \cdot 3)] = 9 \cdot 13 = 117. \tag{12}$$

The focus is on the overall shape of the OC curves, their deviations from the ideal benchmark at different levels of defectiveness, and the way those deviations change when the aggregation sequence, the number of inspectors or the error scenario change. This design was intended to address *RQ1* and *RQ2*, and to provide the empirical basis required for the CJT interpretation developed later.

IV. Results and Interpretation

A. *Presentation of the results*

For each of the nine (n, c) configurations, Fig. 3 compares the ideal Type-B OC curve with the twelve multiple-inspector OC curves obtained under the two aggregation sequences, the three values of m and the two error scenarios. The dashed line represents the ideal SSP. The red curves correspond to scenario *L* and the blue curves correspond to scenario *H*. For each (n, c), the upper graph refers to rows-first aggregation and the lower graph to columns-first aggregation. Some panels in Fig. 3 are shown on the full domain $p \in [0, 1]$, while others are restricted to $p \in [0, 0.4]$, because the largest deviations from the ideal benchmark are concentrated at relatively low defectiveness. This already points to the dominant role of false-defective classifications.

Let us consider first the rows-first sequence. In general, the deviations from the ideal OC curve are substantial, especially under scenario *H*, and they are almost always downward, meaning that the resulting plan is more severe than intended. Even under scenario *L*, noticeable deviations remain in several configurations. The discrepancy tends to be stronger for relatively

large n and relatively small c, that is, for sampling plans that are already intrinsically severe; the combination $n = 14$, $c = 0$ is one of the clearest examples. A first apparently counterintuitive result concerns the role of the number of inspectors m under rows-first aggregation. In scenario *L*, increasing m often makes the aggregated OC curve closer to the ideal benchmark. At first sight, this seems consistent with the intuitive idea that more inspectors should improve reliability. Yet, in scenario *H*, the pattern is not monotonic. For some severe plans, increasing m worsens the deviation from the ideal curve rather than reducing it. In other words, adding inspectors can amplify, rather than attenuate, the distortion. Overall, rows-first aggregation may yield multiple-inspector SSPs whose severity and discriminatory power differ markedly from those of the ideal reference SSP. In most cases the deviation is downward, meaning that good lots are rejected more frequently than expected. Rare exceptions appear for very indulgent plans, particularly when n is small and c is relatively large, as in the case $n = 4$, $c = 2$; there, some curves may lie above the ideal benchmark because false conformings partly offset the effect of false defectives and delay the decline of the acceptance probability as p increases.

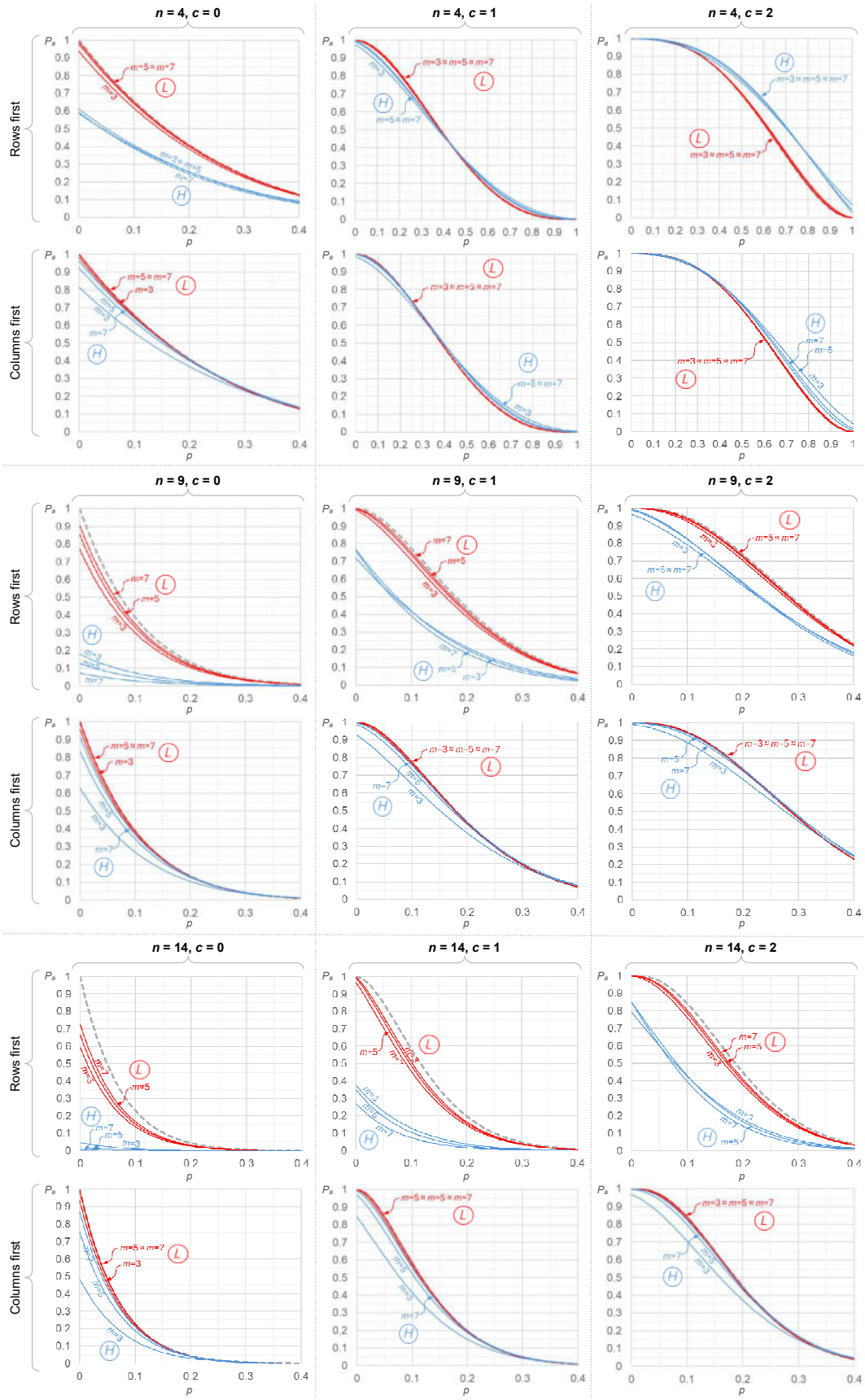


Fig. 3. Type-B OC curves for multiple-inspector SSPs involving m = 3, 5 and 7 inspectors, considering both aggregation sequences, rows-first and columns-first. Nine combinations of n and c are examined, and for each combination two error scenarios are analysed (L and H). The dashed curve represents the ideal SSP with no inspection errors; red curves refer to scenario L and blue curves to scenario H.

The picture changes sharply under columns-first aggregation. In scenario *L*, the corresponding multiple-inspector OC curves almost perfectly overlap the ideal benchmark for nearly all (*n*, *c*) combinations. Even in scenario *H*, the agreement remains good overall, especially when $m = 5$ and even more when $m = 7$. The only clearly visible discrepancies are typically associated with $m = 3$, and even then, they remain moderate compared with those found under rows-first aggregation. The practical implication is straightforward: of the two sequences, columns-first is clearly superior in terms of robustness and stability with respect to the ideal OC curve, especially when inspection errors are non-negligible. Moreover, in most configurations the increase of *m* has a stabilizing effect under columns-first aggregation, whereas under rows-first aggregation the same increase can be either beneficial or detrimental depending on the operating conditions. At this point, however, three issues still require explanation. First, why can a mere change in aggregation order produce such different outcomes when all input data are identical? Second, why does increasing *m* usually help under columns-first aggregation but sometimes hurt under rows-first aggregation? Third, why are deviations from the ideal OC curve mostly downward, with only rare upward exceptions? A purely descriptive reading of Fig. 3 does not provide a satisfactory answer. The next subsections address that gap by using CJT as an interpretive framework.

B. Concise summary of CJT

The majority criterion is classically justified in *social choice theory* through the CJT. This theorem is a standard formal argument behind the idea that, under appropriate conditions, groups can identify the truth more reliably than individuals. In the present context, it provides a useful bridge between the mathematics of multiple-inspector SSPs and the behaviour observed in the OC curves (Condorcet, 1785; Grofman et al., 1983; Dietrich and Spiekermann, 2019).

Consider a binary state $\theta \in \{0, 1\}$, for example the true status of a sampled unit as conforming or defective. Each inspector $i = 1, \ldots, m$ expresses a binary assessment $J_i \in \{0, 1\}$. In the classical CJT setting, three assumptions are made: (i) one alternative is objectively correct and the other is objectively wrong; (ii) every inspector has the same probability *w* of giving the correct answer; and (iii) inspectors' judgments are mutually independent. Under those assumptions, if *s* denotes the total number of correct judgments among the *m* inspectors, then *s* follows a binomial distribution and the probability that the majority is correct is given by (Boland, 1989; Ladha, 1992; Ladha, 1995):

$$W = P(s \geq q) = \sum_{s=q}^{m} \binom{m}{s} \cdot w^s \cdot (1-w)^{m-s}, \tag{13}$$

where the majority threshold *q* is analogous to that defined in (7). Fig. 4 shows the family of curves defined by (13); the horizontal axis reports the probability *w* that a single inspector is correct, while the vertical axis reports the probability *W* that the majority of *m* inspectors is correct.

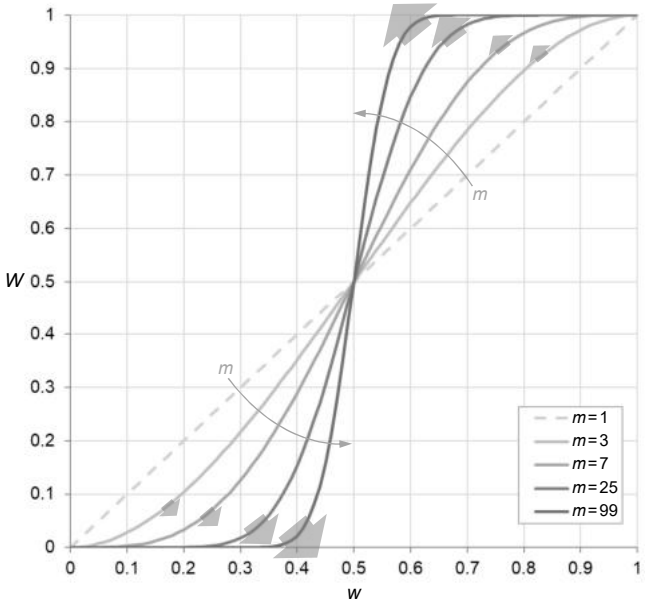


Fig. 4. Graphical representation of the probability *W* as a function of *w* for different values of *m*, according to the model in (13). On the horizontal axis, *w* denotes the probability that a single inspector gives the correct assessment in a binary decision problem; on the vertical axis, *W* denotes the probability that the majority of *m* inspectors is collectively correct. Under independence and equal competence, increasing *m* makes *W* increasingly similar to a step function centered at $w = 0.5$.

The theorem implies that *W* is an increasing function of *w*, but the crucial feature is the threshold at $w = 0.5$. If $w > 0.5$, increasing the group size *m* increases *W*, and *W* tends to 1 as *m* grows. If $w < 0.5$, the opposite happens: increasing *m* drives *W* toward 0. If $w = 0.5$, majority voting is no better than a random guess (Boland, 1989; Fey, 2003). The operational meaning is simple but powerful: majority voting amplifies the individually more likely alternative; if each individual is, on average, more likely to be correct than wrong, then majority improves collective performance. If each individual is, on average, more likely to be wrong than correct, majority makes the collective decision even worse. Moreover, for larger *m*, the curve around $w = 0.5$ becomes steeper, so small changes around the threshold can produce abrupt changes in group performance. This

property is especially important for interpreting the multiple-inspector SSP results, because the probability being aggregated is not the same in the two aggregation sequences (List and Goodin, 2001; Dietrich and Spiekermann, 2019).

C. Interpretation of the results in light of the CJT

To apply the CJT to the present problem, it is useful to distinguish three binary decision problems. In the rows-first sequence, majority is applied to local LDDs. In the columns-first sequence, majority is applied to unit-level assessments, and two distinct cases must be considered: a truly conforming unit and a truly defective unit. Table 2 summarizes this mapping (Condorcet, 1785; Boland, 1989).

Table 2. Interpretation of the two aggregation sequences in light of the CJT.

Aggregation sequence	**Binary decision problem and definition of θ**	**Probability (w) that the individual inspector assigns $\theta = 1$**	**Probability (W) that the majority of inspectors assign $\theta = 1$**
Columns-first	**Problem 1:** Assessment of a sampled unit that is truly conforming **Definition:** $\theta = 0$: unit classified as defective; $\theta = 1$: unit classified as conforming. **Note:** In this problem, the objectively correct alternative is $\theta = 1$.	**Probability:** $1 - a_i$ **Note:** The parameter a_i may vary across inspectors and across product types, and is generally difficult to estimate. Even under relatively unfavourable conditions, the probability $1 - a_i$ is unlikely to fall below 0.5.	**Probability:** $1 - a_{agg}$ **Note:** Since $1 - a_i$ is expected to be greater than 0.5 in most practical situations, the CJT yields a beneficial aggregation effect: as m increases, $1 - a_{agg}$ tends to increase.
Columns-first	**Problem 2:** Assessment of a sampled unit that is truly defective **Definition:** $\theta = 0$: unit classified as conforming; $\theta = 1$: unit classified as defective. **Note:** The objectively correct alternative is again $\theta = 1$, since the alternatives are the exact opposite of those in Problem 1.	**Probability:** $1 - b_i$ **Note:** The parameter b_i may vary across inspectors and across product types, and is generally difficult to estimate. Even under relatively unfavourable conditions, the probability $1 - b_i$ is unlikely to fall below 0.5.	**Probability:** $1 - b_{agg}$ **Note:** Since $1 - b_i$ is expected to be greater than 0.5 in most practical situations, the CJT yields a beneficial aggregation effect: as m increases, $1 - b_{agg}$ tends to increase.
Rows-first	**Problem 3:** Lot acceptance/rejection **Definition:** $\theta = 0$: lot rejection; $\theta = 1$: lot acceptance. **Note:** Unlike the standard CJT setting, this problem does not involve an objectively correct versus incorrect alternative. Only some reference (complementary) probabilities can be defined as a function of p: $1 - P_a$ for state $\theta = 0$ and P_a for state $\theta = 1$, with respect to the ideal OC curve agreed upon by customer and supplier (cf. Sect. II.A).	**Probability:** $P_{a,\,i}$ **Note:** $P_{a,\,i}$ is affected by inspection errors a_i and b_i, which may vary across inspectors. For relatively low values of p, the effect of a_i tends to dominate, leading to a decrease in $P_{a,\,i}$ relative to the ideal OC curve. For relatively high values of p, the effect of b_i tends to dominate, leading to an increase in $P_{a,\,i}$ relative to the ideal OC curve.	**Probability:** $P_{a,\,agg}$ **Note:** $P_{a,\,agg}$ is computed according to (6).

To simplify the description, first consider the columns-first sequence. In Problem 1 of Table 2, the inspected unit is truly conforming; the objectively correct alternative is therefore to classify the unit as conforming. The probability that the i-th inspector makes the correct assessment is $w = 1 - a_i$, while the probability that the majority makes the correct assessment is $W = 1 - a_{agg}$. In the special case in which all inspectors have the same false-defective probability $a_i = a$, (10) collapses to the standard binomial form of the CJT. More generally, even when inspectors are heterogeneous, the qualitative implication remains the same: majority amplifies the individually more likely alternative (Boland, 1989; Grofman et al., 1983). In realistic inspection settings, $1 - a_i$ is typically greater than 0.5 for all inspectors. Therefore, majority aggregation acts in the favourable regime of the CJT, and the aggregated false-defective probability decreases as the number of inspectors increases. This limiting behaviour is expressed by (14) (Condorcet, 1785; Boland, 1989).

$$\lim_{m\to\infty} = 1 - a_{\mathrm{agg}} = 1 \text{ or } \lim_{m\to\infty} = a_{\mathrm{agg}} = 0, \tag{14}$$

The same logic applies to Problem 2, in which the inspected unit is truly defective. Here the correct alternative is to classify the unit as defective, so the individual probability of correct assessment becomes $w = 1 - b_i$, while the probability that the majority is correct is $W = 1 - b_{agg}$. Under homogeneous error rates, (11) again collapses to the classical CJT expression. Because $1 - b_i$ is also normally greater than 0.5 in plausible industrial settings, the majority aggregation reduces b_{agg} as m increases, as expressed in (15) (Condorcet, 1785; Boland, 1989).

$$\lim_{m\to\infty} = 1 - b_{\mathrm{agg}} = 1 \text{ or } \lim_{m\to\infty} = b_{\mathrm{agg}} = 0, \tag{15}$$

Taken together, these two effects explain why columns-first aggregation is so stable. When both aggregated error rates a_{agg} and b_{agg} decrease, the apparent defectiveness p'_{agg} in (9) approaches the true defectiveness p. Consequently, the OC curve in (8) approaches the ideal OC curve in (1). This is exactly what Fig. 3 shows: near-perfect adherence under scenario *L* and good robustness even under the deliberately severe scenario *H*, especially when m is larger. In other words, when majority voting is applied at the unit level, the multiple-inspector system behaves like a progressively better ("aggregate") inspector as the number of individual inspectors increases. The only situation in which the CJT would work against this mechanism would be one in which inspectors were, on average, worse than a random classifier, that is, if a_i or b_i exceeded 0.5. Such a condition is unrealistic in the contexts for which redundant industrial inspection is contemplated (Boland, 1989; Dietrich and Spiekermann, 2019).

The rows-first sequence is structurally different; majority is not applied to unit-level assessments but to LDDs. The relevant quantity for the i-th inspector is therefore not $1 - a_i$ or $1 - b_i$, but the local acceptance probability $P_{a,\,i}$ given by (5). This probability depends simultaneously on the true defectiveness p, the inspection-error rates (a_i, b_i) and the intrinsic severity of the sampling plan (n, c). As a result, the probability being aggregated can move above or below the critical CJT threshold of 0.5 depending on the operating region. For relatively low values of p, false-defective classifications tend to dominate; even when the lot is good, some sampled units can be perceived as defective, which reduces $P_{a,\,i}$ compared with the ideal OC curve. In severe plans, or under high-error conditions, the reduction can be strong enough to push $P_{a,\,i}$ below 0.5 over a meaningful portion of the low-defectiveness region. Once this happens, majority aggregation enters the unfavourable regime of the CJT: the individually more likely alternative becomes rejection, so majority amplifies rejection. As the number of inspectors increases, the global acceptance probability $P_{a,\,agg}$ can collapse rapidly toward zero. This explains why rows-first OC curves are often excessively severe and why, under scenario *H*, increasing m may worsen the deviation. For higher values of p, the ideal acceptance probability is already small. In that region, many inspectors naturally have $P_{a,\,i} < 0.5$, so majority again pushes the aggregated acceptance probability downward. This is why the deviations in rows-first aggregation are so often negative over wide portions of the domain. The result is a plan that rejects lots far more aggressively than the benchmark agreed upon by customer and supplier (cf. Sect. II.A).

The rare upward deviations observed for highly indulgent plans are also understandable in this framework. When n is small and c is relatively large, false-conforming errors b_i can partly neutralize the effect of false defectives and delay the decline of $P_{a,\,i}$ as p increases. If $P_{a,\,i}$ remains above 0.5 over a wider interval of p, majority aggregation amplifies acceptance rather than rejection. In such cases, the rows-first OC curve may rise above the ideal benchmark. This is unusual, but it is fully consistent with the CJT: majority simply strengthens whichever alternative is individually more likely. The most important implication is that increasing the number of inspectors is not intrinsically beneficial. Its effect depends on the side of the 0.5 threshold on which the aggregated probabilities lie. Under columns-first aggregation, the probabilities being aggregated are usually $1 - a_i$ and $1 - b_i$, both typically larger than 0.5; hence a larger m stabilizes and improves the design. Under rows-first aggregation, the probabilities being aggregated are the local lot-acceptance probabilities $P_{a,\,i}$, which can erroneously cross below 0.5. In that case, a larger m amplifies the wrong tendency. This structural difference explains the overall superiority of the columns-first sequence in the presence of inspection error (Condorcet, 1785; Boland, 1989; Dietrich and Spiekermann, 2019).

V. Conclusions

This study has shown that, in attribute sampling plans with multiple inspectors, the aggregation sequence is not a secondary implementation detail; it is a design factor capable of substantially changing the severity and discriminatory power of the resulting OC curve. The ideal single-inspector, zero-error OC curve was used as the natural benchmark because it is the reference typically shared by supplier and customer when a sampling plan is designed. If a multiple-inspector extension is adopted to mitigate inspection error, one would ideally like the resulting OC curve to remain close to that benchmark. In practice, however, the individual error rates a_i and b_i are not always known and may vary across products, over time and across operating conditions. This makes it especially important to identify multiple-inspector configurations that are robust even when those parameters are only approximately known.

The results obtained here indicate that columns-first aggregation combined with majority voting is generally the most robust option (cf. *RQ1*). Its OC curves stay close to the ideal benchmark under mild error rates and remain reasonably stable even under very adverse conditions. In addition, the stabilizing effect tends to increase with m (Burke et al., 1995; Franceschini et al., 2025). Rows-first aggregation behaves very differently. Its deviations from the benchmark are already visible under low error rates and become pronounced under high error rates, particularly for severe plans with large n and small c (cf. *RQ2*). In those settings, false-defective classifications dominate at low defectiveness and drive local acceptance probabilities downward. Majority aggregation then amplifies rejection, yielding an OC curve that is often much more severe than intended. This mechanism also explains why increasing m may sometimes worsen performance rather than improve it.

CJT provides a unified explanation for these results. Majority voting is beneficial only when the probability being aggregated is individually more likely than not to be correct (cf. *RQ3*). Under columns-first aggregation, majority acts on unit-level assessments, where individual correctness is typically above 0.5; therefore, the theorem works in favour of the design and redundancy tends to stabilize the OC curve. Under rows-first aggregation, majority acts instead on local LDDs, whose acceptance probabilities depend jointly on defectiveness, inspection error and plan severity and may fall below the 0.5 threshold. In that regime, majority voting amplifies rejection rather than correcting random mistakes, producing the instability and excessive severity observed in the OC curves (Condorcet, 1785; Grofman et al., 1983; Boland, 1989).

This distinction is particularly relevant for safety-critical and high-value sectors, such as aerospace, nuclear, medical-device or precision manufacturing, where the consequences of an erroneous LDD may be disproportionate. For instance, in the acceptance inspection of a lot of turbine components, electronic modules or safety-related mechanical parts, redundant inspection may be introduced to reduce the risk of overlooking defective units. The present results show, however, that redundancy alone is not sufficient: the sequence by which individual judgements are aggregated can substantially affect the final OC behaviour. In this perspective, columns-first aggregation is preferable because it first consolidates inspectors' judgements at the item level, limiting the propagation of local classification errors to the final LDD. Conversely, rows-first aggregation may lead to overly severe or unstable decisions, with unnecessary re-inspection, scrap, delay or rejection costs.

Accordingly, when redundant inspection is justified and inspection-error rates are difficult to estimate, multiple-inspector SSPs should preferably rely on columns-first aggregation with majority voting. Even $m = 3$ inspectors can provide a meaningful stabilizing effect, whereas rows-first designs should be adopted with caution and preferably only after verifying their behaviour under realistic error-rate and defectiveness scenarios (Maisano et al., 2025).

The study has some limitations that also point to future work. Firstly, only the simplest class of attribute plans, namely SSPs, was considered. Secondly, only Type-B OC curves were analysed, whereas Type-A curves would be needed in the case of isolated lots where the condition $n \ll N$ is not met (Montgomery, 2019). Thirdly, only majority voting was studied in the aggregation step, because unanimity had already been shown to be excessively punitive under realistic error conditions. Fourthly, the analysis assumed independent inspectors; correlated errors remain an important extension. Future studies could therefore extend the comparison to more complex plans, include Type-A curves and integrate economic considerations such as inspector cost, the cost of erroneous LDDs and the trade-off between robustness and inspection effort (Kenett et al., 2023; Maisano et al., 2025).

References

Aurum, A., Petersson, H., Wohlin, C. (2002). State-of-the-art: software inspections after 25 years. Software Testing, Verification and Reliability, 12(3), 133–154. DOI: 10.1002/stvr.243

Boland, P. J. (1989). Majority systems and the Condorcet jury theorem. The Statistician, 38(3), 181-189.

Burke, R. J., Davis, R. D., Kaminsky, F. C., Roberts, A. E. P. (1995). The effect of inspector errors on the true fraction non-conforming: an industrial experiment. Quality Engineering, 7(3), 543–550. DOI: 10.1080/08982119508918802

Chandra, J., Schall, S. (1998). The use of repeated measurements to reduce the effect of measurement errors. IIE Transactions, 20(1), 83–87.

Chun, Y. H. (2009). Improving product quality by multiple inspections: Prior and posterior planning of serial inspection procedures. IIE Transactions, 41(9), 831–842. DOI: 10.1080/07408170802389324

Condorcet, J.-A.-N. de Caritat, Marquis de (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Paris. gallica.bnf.fr

Cymek, D.H. (2024) Effects of Blinded and Nonblinded Sequential Human Redundancy on Inspection Effort and Inspection Outcome in Low Prevalence Visual Search, Scientific Reports 14 (2024): 23003.

Deming, W. E. (2018). Out of the Crisis (reissue edition). MIT Press, Cambridge, MA.

Dietrich, F., Spiekermann, K. (2019). Jury theorems. In: Fricker, M. et al. (Eds.), The Routledge Companion to Social Epistemology, pp. 386–396. Routledge.

Duffuaa, S. O., El-Ga'aly, A. (2015). Impact of inspection errors on the formulation of a multi-objective optimization process target model under inspection sampling plan. Computers & Industrial Engineering, 80, 254–260.

Ferrell Jr, W. G., Chhoker, A. (2002). Design of economically optimal acceptance sampling plans with inspection error. Computers & Operations Research, 29(10), 1283–1300.

Fey, M. (2003). A note on the Condorcet Jury Theorem with supermajority voting rules. Social Choice and Welfare, 20, 27-32.

Franceschini, F., Maisano, D.A., Mastrogiacomo, L. (2022). Rankings and decisions in engineering. Cham, Switzerland: Springer International Publishing.

Franceschini, F., Maisano, D., Mastrogiacomo, L. (2025). Acceptance sampling with multiple inspectors: critical aspects in aggregating individual assessments. Quality Engineering, 37(2), 318–329.

Gadrich, T., Bashkansky, E. (2016). A Bayesian approach to evaluating uncertainty of inaccurate categorical measurements. Measurement, 91, 186–193.

Grofman, B., Owen, G., Feld, S. L. (1983). Thirteen theorems in search of the truth. Theory and Decision, 15, 261-278.

ISO/DIS 2859-1:2025 (2025). Sampling procedures for inspection by attributes – Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection. Draft International Standard, ISO, Geneva.

Jamkhaneh, E. B., Sadeghpour-Gildeh, B., Yari, G. (2011). Inspection error and its effects on single sampling plans with fuzzy parameters. Structural and Multidisciplinary Optimization, 43(4), 555–560.

Jaraiedi M., Iskander W.H. (1989). Posterior analysis of inspector accuracy. International Journal of Quality & Reliability Management, 6(2).

Kenett, R. S., Zacks, S., Gedeck, P. (2023). Sampling Plans for Batch and Sequential Inspection. In Industrial Statistics: A Computer-Based Approach with Python (pp. 397–442). Springer, Cham.

Knight, J. C., Myers, E. A. (1993). An improved inspection technique. Communications of the ACM, 36(11), 51–61.

Ladha, K.K. (1992). The Condorcet Jury Theorem, Free Speech and Correlated Votes. American Journal of Political Science, 36(3), 617-634. DOI: 10.2307/2111584.

Ladha, K.K. (1995). Information Pooling through Majority-Rule Voting: Condorcet's Jury Theorem with Correlated Votes. Journal of Economic Behavior & Organization, 26(3), 353-372. DOI: 10.1016/0167-2681(94)00068-P.

List, C., Goodin, R.E. (2001). Epistemic democracy: Generalizing the Condorcet jury theorem. Journal of Political Philosophy, 9(3), 277-306.

Maisano D., Ferrara L., Franceschini F. (2025). Characteristic Curves for Multiple-Inspector Sampling Plans with Inspection Errors. Quality and Reliability Engineering International (Early View).

Montgomery D.C. (2019). Introduction to Statistical Quality Control (8th ed.). Wiley, New York.

Schilling, E. G., Neubauer, D. V. (2009). Acceptance Sampling in Quality Control (2nd ed.). CRC Press, Boca Raton, FL.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

A Data-Driven Workforce Dimensioning Tool for Grocery Retail: A Kaizen Case Study

João Ferreira [1], André Carvalho [2,3] and Radu Godina [2,3,*]

[1] Kaizen Institute; joaoferreira@kaizen.com
[2] UNIDEMI-Department of Mechanical and Industrial Engineering, Faculty of Science and Technology (FCT), Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal; ame.carvalho@fct.unl.pt, r.godina@fct.unl.pt
[3] Laboratório Associado de Sistemas Inteligentes (LASI), 4800-058 Guimarães, Portugal

***Abstract*- Grocery retail operations face a persistent imbalance between workforce capacity and actual operational demand, with staffing decisions historically driven by experience and habit rather than data. This paper presents the development and implementation of a structured workforce dimensioning tool, built in Microsoft Excel during a professional internship in a Portuguese supermarket conducted in partnership with the Kaizen Institute. Grounded in Standard Work principles and structured within the Plan-Do-Check-Act improvement cycle, the tool calculates staffing requirements from three quantified inputs: normalised task times derived from a prior process standardisation phase, a customer demand profile constructed from historical POS sales records, and daily pallet replenishment workload data from delivery manifests. A two-week controlled pilot validated the model before full deployment as the primary scheduling instrument for section managers. Results against the equivalent prior-year period show a 17% reduction in hours worked, a 20% increase in labour productivity expressed in euros per hour worked, and a net reduction of 13 operators, while sales grew 6%. The study demonstrates that rigorous, data-grounded workforce planning, when built on prior task standardisation and supported by Daily Kaizen routines, generates significant and sustainable gains in retail labour efficiency without capital investment.**

Keywords – Workforce dimensioning, Standard Work, Kaizen, Labour productivity, Grocery retail

I. INTRODUCTION

Labour is one of the largest controllable costs in grocery retail, typically representing between 8% and 20% of total sales revenue (Talarico and Maya Duque, 2015). Variations within this range can significantly affect store profitability. Despite its importance, workforce planning in many supermarkets is still carried out informally. Schedules are often defined by section managers based on experience, employee availability, and established routines rather than on a structured assessment of operational needs. As a result, stores frequently have excessive staffing during periods of low demand and insufficient staffing during peak periods, increasing labour costs while negatively affecting service levels.

The issue is therefore not the availability of labour itself, but how it is allocated throughout the operation. A store that schedules too many associates early in the morning and too few during the busiest hours of the day wastes resources while simultaneously failing to meet customer demand effectively. Addressing this problem requires, at minimum, three types of information: a clear understanding of the tasks performed in store and the time required to execute them consistently, an accurate picture of customer demand fluctuations throughout the week, and a quantified assessment of replenishment workload by day. These data are generally available in most retail operations through point-of-sale systems, delivery records, and direct observation of operational activities. However, they are rarely organised and used in a structured way to support workforce planning decisions.

This paper presents the development and deployment of a workforce dimensioning tool designed to address this gap. The tool was developed during a five-month professional internship in a medium-sized Portuguese supermarket in partnership with the Kaizen Institute. Its development followed the principles of Standard Work, ensuring that task execution became measurable and consistent before workforce allocation decisions were made. The paper focuses specifically on the dimensioning tool, including its theoretical foundations, the standardisation work required for its implementation, its three-module structure, the pilot validation process, and the results obtained after full deployment.

This study contributes to the retail operations literature by proposing and validating a practical workforce dimensioning approach that integrates Standard Work, customer demand variability, and replenishment workload into a single planning model. In doing so, it demonstrates how operational data already available in retail environments can be transformed into a structured decision-support approach for workforce allocation.

The remainder of the paper is structured as follows. Section 2 reviews the literature related to retail labour planning, Standard Work, demand variability, PDCA and Daily Kaizen approaches that supported implementation continuity. Section 3 presents the research methodology. Section 4 describes the current situation analysis and diagnostic findings. Section 5 details the design and deployment of the workforce dimensioning tool. Section 6 presents the results obtained after implementation. Finally, Sections 7 and 8 discuss the main implications of the study and present the conclusions.

II. Literature Review

A. *Labour Cost and Workforce Planning in Grocery Retail*

Labour represents the dominant controllable cost in grocery retail. Talarico and Maya Duque (2015) place staff costs at between 8% and 20% of revenue and note that the sector's thin margins mean even modest inefficiency in workforce deployment carries disproportionate consequences for profitability. The difficulty is not only financial. Customer flow in a supermarket fluctuates across hours, days, and seasons in patterns that are partially predictable but never fully stable, and replenishment workload shifts with delivery schedules and seasonal product mix. Matching staffing to this two-dimensional variability without a structured instrument is, in practice, impossible to do consistently.

The academic literature consistently identifies a dual failure mode in retail workforce management. Chapados, Joliveau, and Rousseau (2011), in a scheduling optimisation study across a retail chain, found that operations without objective staffing criteria simultaneously overstaff during slow periods and understaff during peaks. Their modelling demonstrated that formalising the scheduling logic, even with a relatively simple demand-based calculation, reduced total labour cost substantially. Talarico and Maya Duque (2015) reached the same conclusion through a different optimisation approach, showing that replacing experiential scheduling with a structured model generated consistent reductions in labour cost as a proportion of revenue. Both studies share a fundamental premise: the inefficiency is not in the quantum of labour available but in the discipline with which it is allocated.

High employee turnover amplifies the problem. Evans (2020) reports annual turnover rates exceeding 50% in several European retail markets, driven by low wages, unpredictable schedules, and physically demanding conditions. High turnover creates a continuous instability in team composition that makes experience-based scheduling even less reliable: a section head who calibrated team size based on knowing their people personally must repeatedly recalibrate as team membership changes. It also generates direct costs in recruitment and onboarding that compound the operational inefficiency. A dimensioning tool that calculates staffing requirements from task standards and demand data rather than from familiarity with specific individuals is, in this context, inherently more robust to turnover.

The service quality dimension of workforce planning is captured by Caruelle, Lervik-Olsen, and Gustafsson (2023), who show that customer satisfaction is particularly sensitive to deviations between expected and actual waiting time. The negative effect of a longer-than-expected wait is asymmetrically larger than the positive effect of a shorter-than-expected one, which means that the cost of understaffing at peak periods is not merely additive but disproportionate. A checkout queue that builds because a shift is understaffed does not merely inconvenience customers in the moment; it creates a service failure that is directly traceable to a scheduling decision made earlier that day, when peak demand was predictable and the appropriate staffing level was calculable. Marques, Jorge, and Reis (2022) and Marques, Carvalho, and Santos (2022) both document Portuguese retail cases in which Lean-driven improvements to labour planning and process standardisation generated labour productivity gains in the range of 20–25%, confirming that the opportunity identified in the optimisation literature translates to real-world outcomes.

B. *Standard Work as a Prerequisite for Dimensioning*

Any workforce dimensioning model produces outputs that are only as reliable as the task time data it receives as input. This is the dependency that makes Standard Work a prerequisite rather than a parallel initiative. Standard Work, in the Kaizen framework, involves defining the optimal sequence, method, and duration for each recurring task, documenting those definitions clearly enough that any trained associate can execute them consistently, and using the resulting data as the operational baseline for planning decisions (Lodgaard et al., 2016). In a grocery retail context, this means knowing, with measurement-based precision, how long a pallet of ambient grocery takes to replenish from arrival at the dock to completion on the shelf, how many minutes a butchery opening routine requires under a standardised method, and how service time varies by customer type and section across a typical trading day.

Without that knowledge, a staffing calculation must rely on assumptions, and assumptions in a variable environment produce plans that are systematically wrong. When task durations are not standardised, they vary significantly from associate to associate and from day to day. Any staffing model that attempts to plan around this variability must include wide safety margins to absorb the uncertainty, which systematically inflates headcount and defeats the purpose of the exercise. When task times are measured and standardised, those margins can be tightened to the actual operating requirements of the task, and the resulting plans are precise enough to be operationally meaningful.

Marques, Jorge, and Reis (2022) describe this sequence explicitly in their Portuguese retail case: the creation of a standards manual preceded and enabled the subsequent restructuring of replenishment schedules, which in turn produced the efficiency improvements they report. The manual is not the improvement itself; it is the instrument that makes structured improvement possible. In the present study, the same logic applied directly: the Standards Manual created during the standardisation phase of the project became the input layer for the dimensioning tool, supplying the task time and frequency data from which staffing calculations were derived. Without it, the tool could not have been built on the basis of measured reality rather than managerial estimate.

C. *Demand-Side Inputs: Customer Flow and Replenishment Workload*

Standard Work addresses the supply side of the dimensioning problem: how long tasks take. The demand side requires understanding when and how much work arises. In grocery retail, operational demand comes from two primary sources that behave quite differently and must be modelled separately. The first is customer arrivals, which generate service requirements at checkouts and attended counters and whose intensity varies across the day and week in ways that are traceable in sales records. The second is goods replenishment, which generates physical

handling workload in the sales floor and warehouse and whose volume depends on delivery schedules and product mix.

Customer demand follows recognisable patterns but with meaningful within-week variation. Caruelle, Lervik-Olsen, and Gustafsson (2023) document the service perception consequences of demand-supply mismatch at checkout and counter positions: the asymmetric sensitivity to wait time means that even moderate understaffing at peak hours creates disproportionate customer dissatisfaction. Accurate mapping of arrival distributions by half-hour and day of week is therefore not a refinement to a scheduling model but a foundational input. Sales records from the POS system provide this data directly; the analytical step is to structure it into a demand curve and link it to service time estimates that allow the required staffing level to be calculated for each time window.

Replenishment workload is the other major driver of staffing need, and it is frequently the one that is most poorly accounted for in informal scheduling. Reiner, Teller, and Kotzab (2013) identify in-store logistics as one of the most time-consuming dimensions of retail operations, absorbing up to 40% of total associate hours. The key driver of variability is not just delivery volume but the interaction between volume and product category: the time required to process a pallet of ambient grocery differs substantially from a pallet of chilled products, beverages, or cleaning materials, because physical handling characteristics, storage proximity, and shelf presentation requirements vary significantly across categories. A planning model that treats all pallets as equivalent will produce systematic misestimates of replenishment workload, particularly on heavy-delivery days or in sections with complex presentation requirements. Capturing this variability, through measured replenishment times per category combined with historical delivery manifests, is what allows the model to produce day-specific and section-specific staffing estimates rather than weekly averages.

D. PDCA and Daily Kaizen as Implementation Structures

A dimensioning tool that is technically sound but organisationally unsupported will be used once, found imperfect, and abandoned. The PDCA cycle addresses this by structuring the implementation as an iterative process rather than a one-time deployment. In the Plan phase, the problem is diagnosed, the model is specified, and expected outcomes are defined. The Do phase pilots the tool under controlled conditions, allowing its parameters to be tested against operational reality before standardisation. The Check phase compares outcomes against the baseline and identifies the specific deviations - in task time estimates, demand patterns, or correction factors - that need adjustment. The Act phase institutionalises the validated model and initiates a new cycle for any gaps that remain (Nguyen et al., 2020; Queiroz et al., 2019). The value of this structure is not procedural tidiness; it is that it creates an explicit moment to catch errors before they propagate through the full deployment.

The pilot phase deserves particular emphasis because it is the step most undervalued in practice. A controlled test period is not merely procedural; it is the mechanism through which the gap between theoretical parameters and operational reality is identified and closed. Task time standards, however, carefully measured in the gemba, are derived under observation conditions that may not fully replicate the variability of live operation. A two-week pilot, run with actual team members following actual schedules derived from the tool, surfaces the discrepancies that static measurement cannot capture. In the present study, the pilot revealed three tasks whose standardised durations underestimated actual execution time in live conditions, small errors individually but consequential in aggregate. Correcting them before full deployment was only possible because the PDCA structure created a dedicated test phase with structured observation and feedback.

Daily Kaizen, structured as brief daily stand-up meetings at both the team level and the management level, provides the continuous feedback mechanism that keeps a dimensioning tool accurate over time. As Bastos and Santos (2020) describe, Daily Kaizen creates a regular cadence of indicator review, deviation identification, and corrective action, operating on a timescale short enough to catch problems before they compound. In the context of workforce planning, this means that when task durations drift from standards, when customer flow patterns shift seasonally, or when product mix changes alter replenishment workload, those changes are surfaced in the daily routine and reflected in a tool update within the same weekly planning cycle. Without this mechanism, a dimensioning tool becomes progressively obsolete as the operational reality it was calibrated to describe diverges from the current state of the operation.

Lodgaard et al. (2016) identify behavioural inertia as one of the most persistent barriers to Lean adoption in retail. A section manager with fifteen years of experience building schedules from memory does not switch to a spreadsheet-based model because the spreadsheet becomes available; they switch when the daily routine makes using it the path of least resistance. Daily Kaizen creates that routine. The brief daily review of actuals against plan, done consistently, makes the tool's outputs the reference point for the conversation rather than an optional supplement to it. In this project, embedding the dimensioning tool into the Daily Kaizen cycle was not an add-on feature designed to improve adoption rates; it was the mechanism through which the tool became the store's actual planning instrument rather than a consulting deliverable that would have been archived within a month.

III. METHODOLOGY

This study follows a single-case design, conducted in a medium-sized Portuguese supermarket operating under a national retail chain and referred to as Company X. The intervention was carried out during a five-month professional internship with the Kaizen Institute. The case study format was selected because the research problem is fundamentally context-dependent: the tool's design, calibration, and results cannot be meaningfully separated from the specific operational conditions, team structure, delivery schedules, and demand patterns of the store in which it was built

(Wohlin, 2021; Martinsuo and Huemann, 2021). This embeddedness is a feature rather than a limitation: it is precisely the kind of context-specificity that the case study format is designed to document and analyse (Bell and Warren, 2023).

To address the research question, "How can the application of Kaizen methodologies and the development of support tools contribute to increasing productivity in a food retail store?", the units of analysis, the methodologies applied, and the unit of observation were identified according to the following framework, as shown in Table I.

TABLE I. CASE STUDY DETAILS AND RESEARCH QUESTION

Research Question	***Units of Analysis***	***Data Collection Methodology***	***Unit of Observation***
How can the application of Kaizen methodologies and the development of support tools contribute to increasing productivity in a food retail store?	*Butchery* *Fish Counter* *Checkout Counters* *Store*	*Unstructured Interviews*	*Employees and Store Managers*
		Operational Records	*Operational Documents and Data*
		Operational Observation	*Employees*

The methodological structure followed the PDCA cycle. The Plan phase encompassed the current situation diagnosis through direct gemba observation across all sections, structured interviews with section managers and associates, and systematic analysis of delivery manifests, POS sales data, and payroll records. A dedicated Excel tool was developed to standardise task timing data collection, registering start and end times for each activity and computing section-level and associate-level productivity automatically. This phase produced the Standards Manual and the task time database that formed the input layer for the dimensioning model. The Do phase covered tool development, a two-week controlled pilot under live conditions, and full deployment across all store sections. The Check phase compared results against the equivalent prior-year period rather than against a pre-intervention baseline, a deliberate choice to control for the strong seasonal demand patterns in grocery retail that make simple before-after comparisons unreliable. The Act phase defined the governance model for ongoing tool revision and embedded it into the store's Daily Kaizen cycle.

The primary performance indicator was labour productivity expressed as euros of revenue per hour worked, derived from the store's accounting and payroll systems. This metric was chosen over alternatives such as cases per hour or cost per hour because it integrates both the revenue and labour cost dimensions of the operation, is directly actionable for management decisions, and can be compared across periods without requiring normalisation for product mix or pricing changes. Additional indicators tracked included total hours worked, operator headcount, and total sales value. All comparisons used the same period of the prior year as the reference, rather than the immediately preceding period, to control for the seasonal patterns that make week-on-week or month-on-month comparisons in grocery retail misleading.

IV. CURRENT SITUATION DIAGNOSIS

The diagnostic phase confirmed the dual failure mode described in the literature. Staffing decisions across Company X were made empirically by section heads with no systematic reference to demand data or workload calculations. Schedules reflected historical habit and perceived availability, producing recurring oversized teams during low-demand windows and insufficient coverage at peak periods in checkouts and fresh service counters. The extent of this misalignment became clear during the measurement phase.

Productivity varied significantly across sections. In the self-service area, values ranged from approximately 39 cases per hour per associate in ambient grocery to over 100 cases per hour in the fruit and vegetable section. These differences were not attributable solely to inherent task complexity. A substantial part of the variability reflected inconsistency in execution method, absence of standardised work sequences, and staffing mismatches that left some sections idle and others overloaded at the same hour. Task times were highly variable because each associate performed daily activities according to personal habit. There were no documented standards for any recurring task, which meant that every calculation of workload, implicit or explicit, rested on individual experience rather than measured fact. This was particularly visible in the fresh service sections.

Opening and closing routines in the butchery and fishmonger sections were particularly inefficient. Measured durations were substantially above what a structured process would require, and the excess time was traceable to undocumented and inconsistent task sequencing. There was also no formal task allocation mechanism at the start of each shift, which meant that workloads were distributed informally and unevenly, visible only to whoever happened to be managing that specific day.

The same absence of structure extended to the warehouse. Products were stored without a clear organisation logic, pallet movements generated redundant travel within the store, and there were no visual indicators to signal replenishment priorities or sequences. These conditions meant that the time to process a pallet of merchandise was not only long but highly variable, making it impossible to predict replenishment workload accurately without first addressing the underlying process. The diagnostic conclusion was unambiguous. A dimensioning tool built without prior task standardisation would have been unreliable, because its inputs would have reflected an inefficient and variable process rather than a normalised and measurable one. The Standards Manual had to come first.

Taken together, these conditions are reflected in the baseline productivity data. Figure 1 shows the values recorded across sections prior to any intervention, measured in €/h worked - the primary metric used to assess the impact of the

dimensioning tool. The values recorded were significantly below the benchmarks observed in similar Kaizen projects in grocery retail, indicating that teams were systematically over dimensioned relative to the actual workload generated. This gap between staffed hours and productive output was not isolated to one area. It was consistent across sections, pointing to a structural problem in how staffing decisions were made rather than isolated performance issues. The baseline therefore does not represent a floor - it represents the starting point from which a correctly dimensioned workforce, supported by standardised tasks and demand-driven scheduling, is expected to deliver measurable gains in €/h worked.

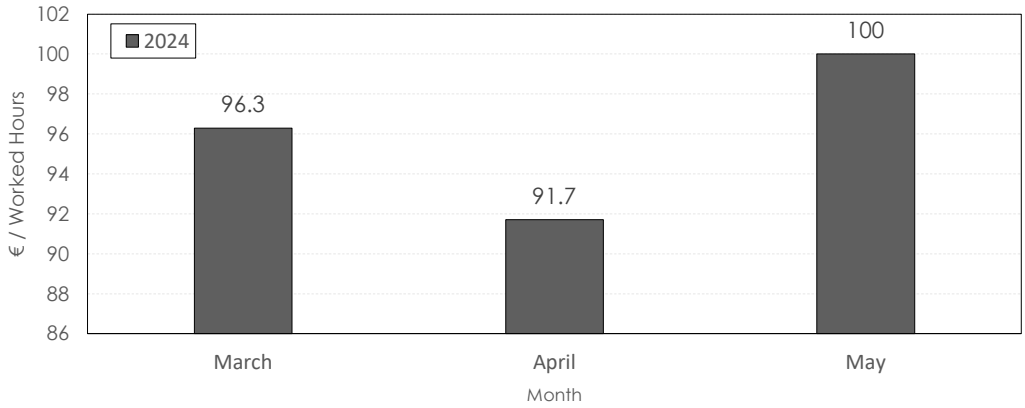


Figure 1. Baseline productivity per section

V. Tool Development and Implementation

A. *Prerequisites: Standards Manual and Task Normalisation*

The first phase of the project was the creation of a Standards Manual covering every recurring task in the store. This was not a parallel initiative to the dimensioning tool; it was its enabling condition. The manual was built through a combination of direct gemba observation, daily interviews with section managers and associates, and iterative testing of task sequences in live conditions. For each task, a standard execution time was established through repeated measurement and validated with the most experienced team members to ensure it reflected achievable rather than idealised performance. The frequency with which each task is performed across the week was also recorded, since a task performed twice a week generates a very different workload contribution from one performed on every shift.

The manual covered the full breadth of the store's operations: shelf replenishment, order management, traceability procedures, opening and closing routines for each section, customer service standards, and administrative tasks. For fresh sections, it included production sequences and presentation planograms. The manual was physically distributed to all sections and conceived as a living document subject to periodic revision through the Daily Kaizen cycle. Beyond its role as input to the dimensioning tool, it also served as the training reference for new associates, which in a sector with 50%+ annual turnover represents a substantial operational benefit in its own right.

B. *Tool Architecture: Three Calculation Modules*

The workforce dimensioning tool was built in Microsoft Excel and structured as three functional modules, each addressing a distinct category of operational demand. The modular structure was deliberate: it allows inputs to be updated independently as conditions change, without requiring a full rebuild of the model. Each module outputs a workload figure in hours, which the tool aggregates into a total staffing requirement expressed in Full-Time Equivalents (FTEs) per shift and day.

The first module covers normalised daily tasks, as seen in Figure 2. It lists every standard activity for a given section as catalogued in the Standards Manual, assigns the measured execution time and the weekly frequency of each activity, and calculates the total task-based workload in hours per day of the week.

	Tarefas	Slot	Minutos	TU	Segunda-feira	Terça-feira	Quarta-feira	Quinta-feira	Sexta-feira	Sábado	Domingo
LS	Reposição Seletiva		60	1,00	10	2	6	2	6	2	6
LS	Reposição Diretos		35	0,58	1	1	3	3	3	1	1
LS	Receção Diretos		10	0,17	3	6	7	8	8	4	2
LS	Folheto		60	1,00				5			
LS	Afichets		60	1,00				7			
LS	Rastriabilidade		7	0,12	5	5	5	5	5	5	5
LS	Reposição de Fruta		30	0,50	16	10	10	10	10	10	10
LS	Ajuste de Frentes		30	0,50	4	2	4	2	4	2	4
LS	Colocação de Preços		30	0,50	1		1		1		
LS	Implantações		1	0,02			4		4		5
LS	Validades		30	0,50	1	1	1	1	1	1	1
LS	Vasilhame		30	0,50	1	1	1	1	1	1	1

Figure 2. Task normalisation module: input structure with time and frequency

This module captures the fixed workload component of each section: the work that must be done regardless of customer flow, including cleaning, traceability checks, stock rotation, opening and closing procedures, and standard replenishment tasks. The sum of these values constitutes the minimum staffing need for each section before any demand-driven adjustment is applied.

The second module addresses customer service demand, as it can be seen in Figure 3. For checkout positions and attended counters in the butchery, fishmonger, and delicatessen sections, a demand profile was constructed from historical POS data for the equivalent prior-year period, showing the number of customers per half-hour interval across each day of the week. Weekend and Friday afternoon peaks were particularly pronounced and had previously been systematically undercovered. This distribution was combined with the average service time per customer, extracted from the POS system and verified through direct gemba observation, to produce an hourly staffing requirement for each customer-facing position. The model also incorporated the variability of service time across sections: counter service in the butchery differs meaningfully in duration from a standard checkout transaction, and the model treated them separately.

Average of SOMA_Column Labels														
Row Labels	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	14	50	75	90	93	74	62	66	75	93	108	98	60	7
2	16	58	81	94	87	71	62	68	76	91	110	96	61	7
3	16	56	77	88	85	67	59	66	70	90	109	99	63	7
4	17	58	81	92	88	70	59	64	70	89	106	96	60	7
5	17	60	88	96	94	75	66	68	81	92	109	100	62	8
6	19	67	99	121	112	85	71	76	77	83	95	89	63	9
7	17	68	102	125	118	84	69	72	74	83	90	81	56	6

	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2ª feira	1,1	2,0	3,0	3,6	3,7	3,0	2,5	2,6	3,0	3,7	4,3	3,9	2,4	0,6
3ª feira	1,3	2,3	3,2	3,8	3,5	2,8	2,5	2,7	3,0	3,6	4,4	3,8	2,5	0,6
4ª feira	1,3	2,2	3,1	3,5	3,4	2,7	2,4	2,6	2,8	3,6	4,3	4,0	2,5	0,6
5ª feira	1,3	2,3	3,3	3,7	3,5	2,8	2,3	2,6	2,8	3,5	4,2	3,8	2,4	0,5
6ª feira	1,4	2,4	3,5	3,9	3,7	3,0	2,7	2,7	3,2	3,7	4,3	4,0	2,5	0,6
Sab	1,5	2,7	4,0	4,9	4,5	3,4	2,9	3,0	3,1	3,3	3,8	3,5	2,5	0,7
Dom	1,3	2,7	4,1	5,0	4,7	3,4	2,8	2,9	2,9	3,3	3,6	3,3	2,2	0,5

TMA	2,4
Tempo FTE	60
Horas FTE	60

Figure 3. Demand module: customer arrival profile and service time matrix

The third module covers pallet replenishment workload in the self-service area, as represented in Figure 4. Historical delivery manifests provided the average number of pallets received per product category per day of the week, reflecting the store's actual delivery schedule. These figures were multiplied by the measured replenishment time per pallet type. Replenishment time varied significantly across categories: ambient grocery required substantially more handling time per pallet than beverages or personal care products, because of higher unit counts per pallet, more complex shelf presentation, and greater physical weight. The module calculated total replenishment hours per section per day, capturing the day-of-week variability in delivery volumes that had previously made it impossible to schedule replenishment teams accurately without data.

	Paletes							
Áreas	p. MC	p. DP	p. Merce	p. Charcu	p. Fruta Legum	p. Congela	p. Lactic	p. Líqui
Tempo médio de reposição (h)	1,1	0,7	1,4	0,2	0,7	1,2	1,2	1,4
Segunda-feira	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Terça-feira	0,0	2,0	3,0	0,0	4,0	2,0	1,0	2,0

Figure 4. Replenishment workload module: pallets by category and processing time

C. Aggregation, Correction Factors, and Output

Once the three modules were populated, the tool aggregated their outputs into a total workload figure per section and shift. Before converting this into a staffing requirement, four operational correction factors were applied. First, statutory rest breaks per shift were deducted from each associate's productive time. Second, the store's observed average absenteeism rate was incorporated as a buffer, ensuring that scheduled headcount reflected expected rather than theoretical availability. Third, time allocated to the daily stand-up meeting was treated as a fixed deduction from productive time per shift. Fourth, unavoidable intra-store travel for associates covering multiple areas was estimated and added to total workload. Physical workspace constraints were also incorporated: in certain sections,

maximum team size is bounded by available working space regardless of calculated demand, and these caps were applied as hard constraints on the output. Minimum staffing floors were defined for safety-critical positions, including at least one open checkout at all times and one associate per attended counter during trading hours.

The tool's final output is a weekly coverage matrix (Figure 5) showing the required number of associates per section per shift, split between morning and afternoon periods, for each day of the week. The output sheet is structured to be read directly by section managers with no intermediate calculation: each cell contains the required FTE for a given section, period, and day, and the manager constructs next week's schedule from it each Thursday. The choice of Thursday as the scheduling day was itself a change from prior practice, which had been ad hoc; anchoring the schedule to a fixed day created a predictable planning rhythm and ensured that the tool's outputs were used within the same weekly cycle in which they were produced.

Segunda-Feira	Terça-Feira	Quarta-Feira	Quinta-Feira	Sexta-Feira	Sábado	Domingo
10	11	11	12	11	12	10
Pessoas	Pessoas	Pessoas	Pessoas	Pessoas	Pessoas	Pessoas

Figure 5. Dimensioning tool output: weekly staffing matrix by day and period

D. Adaptation for Fresh and Counter Sections

The base model was designed for the self-service area, where the combination of task workload, customer service demand, and pallet replenishment covers the full range of staffing drivers. For fresh sections, a two-module version was developed that retains the normalised task and counter service modules but replaces the pallet replenishment component with a production activity intensity profile specific to each section's daily output requirements. In the butchery and fishmonger sections, for example, the primary workload driver is production volume, which follows a daily curve tied to sales patterns rather than a delivery schedule. Minimum staffing floors were defined per section and per period to always ensure service coverage regardless of what the demand calculation generated for any individual slot.

E. Pilot and Full Deployment

The tool was implemented in two stages following the PDCA framework. A controlled two-week pilot was run with schedules derived strictly from the tool's outputs, with no manual override by section managers. Associates from sections where the tool calculated lower workload than current staffing were redeployed to cover requirements elsewhere; a small number were placed on planned leave to bring actual headcount into alignment with the model's recommendations. The pilot required daily coordination between the consulting team and store management. In addition to monitoring operational outcomes, the consulting team observed task execution in each section to verify that associates were following the standardised methods on which the time estimates had been based - since a schedule calibrated to a standard time is only valid if the standard is being followed.

The pilot revealed three tasks whose standardised times in the manual slightly underestimated actual execution time under live conditions. These were corrected in the Standards Manual before full deployment, with the corresponding update propagated to the dimensioning tool's input sheets. Following this recalibration, the tool was formally adopted as the primary scheduling instrument for all sections of the store. Its integration into the Daily Kaizen routine meant that any future deviation from standards, or any change in delivery volumes or customer patterns, was surfaced in the daily stand-up and reflected in the tool within the same weekly planning cycle.

VI. Results

Performance was assessed by comparing results against the equivalent prior-year period, a baseline chosen specifically to control for the strong seasonal demand patterns that characterise grocery retail and that would otherwise confound a simple before-after comparison. All figures reported below use this same-period prior-year comparison unless otherwise indicated.

Total hours worked fell 17% compared to the same period in the previous year. This reduction was a direct consequence of eliminating the structural surplus of oversized shifts in low-demand windows, which the tool made visible for the first time by replacing scheduling intuition with a quantified workload model. Critically, this reduction was not achieved by cutting service capacity: peak-period coverage was maintained throughout, because the reallocation of hours was guided by the demand profiles embedded in the model rather than by a blunt headcount-reduction target.

Labour productivity, expressed as euros of revenue per hour worked, increased 20%. Figure 6 integrates both the reduction in total hours and the concurrent trend in sales performance. A 20% productivity improvement achieved without any change in store format, product range, or technology points to the correction of a planning problem rather than the addition of an operational capability. The original staffing model was not generating competitive output per hour because it was not designed to match effort to demand; once that alignment was introduced through the dimensioning tool, the productivity gain followed structurally.

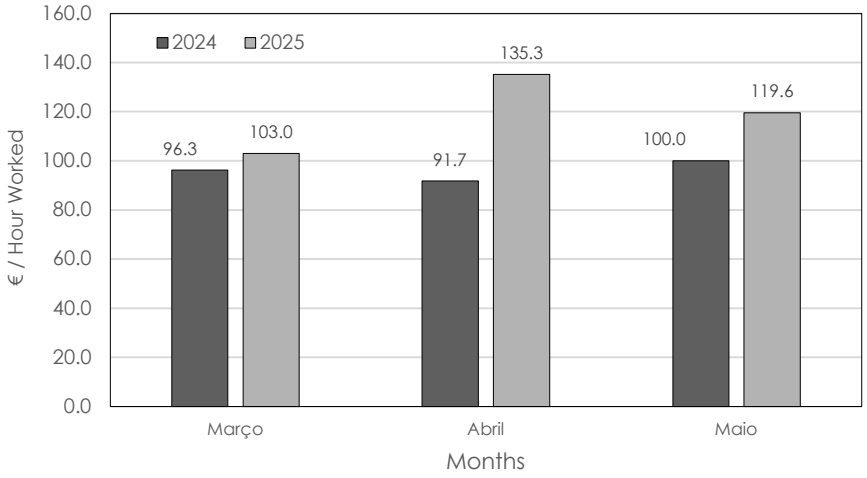


Figure 6. Evolution of the productivity

Team size was reduced by 13 operators over the implementation period. This reduction was facilitated by a period of above-average voluntary turnover, which allowed the store to absorb the adjustment through natural attrition rather than active workforce reduction. Sales grew 6% in parallel, confirming that the headcount reduction did not compromise service quality or shelf availability. Table II summarises the key outcomes.

TABLE II. KEY PERFORMANCE OUTCOMES

Indicator	***Baseline***	***After Implementation***
Labour productivity (euros / hour worked)	96 €/Hh	119,3€/Hh
Hours worked vs. prior-year period	-	-17%
Number of operators	68 operators	55 operators
Total sales value	-	+6%

At section level, the most pronounced reductions in overstaffing occurred in the self-service area and at checkout positions, both of which had been most affected by the pattern of scheduling without demand data. The butchery and fishmonger sections benefited from the removal of excess morning-shift staffing that had previously been a consequence of routine rather than measured need. Both sections had been included in the Standards Manual and opening model revisions during the preceding standardisation phase, which is what allowed their workload estimates to be reliable enough to support the dimensioning calculation.

The results obtained in this project are aligned with the existing literature on food retail operations. In the field of inventory management and on-shelf product availability, the stabilisation of replenishment processes and the use of simple signals to adjust orders resulted in a reduction in stockouts and greater predictability of flow. This behaviour is consistent with evidence showing that leaner inventories, synchronised with demand, lead to reductions in tied-up capital and improvements in effective product availability, particularly in contexts characterised by lower demand uncertainty. In parallel, the improvement in in-store stock accuracy contributed to mitigating the problem described in the literature, whereby a significant proportion of items presented inaccurate inventory records, resulting in service failures and lost sales (Dombrowski & Mielke, 2013).

Regarding labour and productivity, the combination of workforce sizing with task standardisation and daily management reduced working hours and performance variability, while increasing productivity. These effects are consistent with the literature, which quantifies the weight of labour in the total cost structure of retail and highlights scheduling inefficiencies with impacts on both efficiency and service quality. The results also support the idea that visual management practices, workload balancing, and workforce polyvalence reduce the need for reactive staffing reinforcement and stabilise operations, with indirect positive effects on waiting times (Caruelle et al., 2023) and customer satisfaction, as suggested by studies linking flow management to service perceptions (Patel & Deshpande, 2017; Bastos, 2020).

VII. Discussion

The results confirm the argument developed in the literature review: the primary cause of labour inefficiency in this store was not a shortage of analytical capability or operational knowledge, but the absence of the practice of using the available data systematically. The information required to build the dimensioning tool, task times, customer arrival distributions, pallet counts, was present in the store's own systems and collectible through structured gemba observation. The tool required no specialist software. What was absent was the discipline to collect that information, structure it, and act on it through a planning instrument rather than through managerial estimate.

The 20% labour productivity gain is consistent with outcomes reported in comparable Portuguese retail interventions (Marques, Jorge and Reis, 2022; Marques, Carvalho and Santos, 2022). This case contributes a more granular account of the mechanism: the gain derived specifically from replacing experience-based scheduling with a workload-anchored calculation that correctly allocated hours across shifts and sections. The 17% reduction in total hours was not achieved by reducing service capacity but by eliminating the systematic surplus that informal scheduling generates in low-demand periods, a surplus that is invisible without a planning model but obvious once one is applied. This distinction matters for practitioners who may interpret workforce efficiency initiatives as headcount reduction programmes rather than as planning improvement exercises.

The dependency between the Standards Manual and the dimensioning tool proved to be the most important structural feature of the intervention. The argument for investing in task standardisation before building a planning model is not always intuitive, particularly in an operational environment where managers are accustomed to managing by experience and where documenting task times can seem like an administrative distraction from the daily running of the store. The justification is straightforward: a dimensioning model built on measured data produces plans that are operationally trustworthy and that managers will use. A model built on estimates produces plans that are systematically wrong in ways that are hard to diagnose, and that managers will eventually abandon in favour of the informal methods they already know.

The pilot phase produced findings that a purely analytical approach would not have captured. The three task time corrections were small in isolation - a few minutes per activity - but applied systematically across all sections and all days of the week, they would have introduced a consistent directional bias into the staffing outputs. More importantly, two of the three discrepancies arose from the difference between task times measured under observation and task times under normal operating conditions, a measurement artefact that is well documented in industrial engineering but easy to overlook in a field setting. Catching and correcting these before full deployment was possible only because the PDCA structure created a dedicated test period with structured observation, rather than moving directly from development to deployment. The management team's confidence in the tool after the pilot was higher than it would have been without one, because they had seen the errors found and corrected rather than having them emerge post-deployment.

Three limitations bear acknowledgement. First, the study covers a single store, which restricts the direct transferability of the quantitative results to settings with different store formats, team compositions, or delivery structures. The tool's logic is generalisable; the calibrated parameters are not, and any replication would require its own measurement and pilot phase. Second, the headcount reduction was partially facilitated by above-average voluntary turnover during the implementation period. In a more stable labour environment, achieving the same reduction would require deliberate redeployment planning across sections or shifts, which carries different managerial and HR implications. Third, the dimensioning tool was implemented concurrently with other improvements at the same store. While the labour productivity and hours metrics are most directly attributable to scheduling decisions, the possibility of partial interaction effects with replenishment route standardisation and fresh department production planning cannot be entirely ruled out.

VIII. Conclusions

This paper has presented the development, deployment, and quantified results of a workforce dimensioning tool built on Standard Work principles and structured within the PDCA improvement cycle. The tool addresses the core failure of retail workforce management identified in the literature: the absence of an objective, data-grounded criterion for scheduling decisions. By converting three categories of measurable operational input into a weekly staffing matrix by section and shift, it replaces experience-based scheduling with a structured planning instrument that is accessible to store managers without specialist expertise, requires no dedicated software infrastructure, and can be maintained and updated through existing management routines.

The results demonstrate what the literature predicts but rarely documents at this level of operational detail. A 17% reduction in hours worked, a 20% increase in labour productivity, and a reduction of 13 operators, achieved while sales grew 6%, are the direct consequence of replacing allocation by habit with allocation by calculation. The instrument that made this possible was not technically complex. What made it effective was the rigour applied upstream of it: the Standards Manual that ensured its inputs were measured rather than assumed, the controlled pilot that ensured its parameters were calibrated rather than theoretical, and the Daily Kaizen routines that ensured it remained accurate rather than obsolescent.

For practitioners, the practical implication is clear. The barrier to workforce dimensioning of this kind is not technical. The data exists in every store that operates a POS system and receives goods under delivery manifests. The analytical logic requires a spreadsheet. What the intervention demands is the organisational discipline to measure task times before modelling from them, and the patience to run a pilot before standardising the output. These are investments in process rigour that the results here show pay for themselves quickly and without capital expenditure.

As for the research question, "How can the application of Kaizen methodologies and the development of support tools contribute to increasing productivity in a food retail store?", the data confirm a positive and consistent response. The reduction in times and losses, the increase in productivity, and the improvement in service perception demonstrate that the pillars of Standards, Daily Management, and Data are determinant factors for stabilising, improving, and sustaining performance. The results achieved are aligned with the orders of magnitude reported in the literature, and are in some cases higher, particularly regarding labour and productivity. In cases where the results are lower, such as replenishment time and stockouts, this can be explained by the specific operational context, while clear pathways are identified to achieve the desired outcomes without the need for substantial investment.

Future research could usefully address the conditions under which a dimensioning model developed in one store can be transferred and recalibrated to another, which is the natural next step for organisations that want to replicate this approach across a portfolio of stores. The behaviour of the model under high-volatility conditions, promotional events, seasonal peaks, and supply disruptions, also merits investigation, since these are the periods when planning errors carry the highest cost and the gap between scheduled and required staffing tends to widen most sharply.

Acknowledgment

André Carvalho and Radu Godina acknowledges the FCT, I.P.-Fundação para a Ciência e a Tecnologia, in the scope of the project UID/00667/2025 (https://doi.org/10.54499/UID/00667/2025) (UNIDEMI).

References

Bastos, A. and Santos, S. (2020), Strat to Action: The KAIZEN™ Method for Turning Strategy into Action, 2nd ed., McGraw Hill Interamericana, Madrid.

Bell, R. and Warren, V. (2023), 'Illuminating a methodological pathway for DBA researchers: Utilizing case studies and mixed methods for applied research', Social Sciences and Humanities Open, Vol. 7, No. 1, p. 100391.

Caruelle, D., Lervik-Olsen, L. and Gustafsson, A. (2023), 'The clock is ticking - or is it? Customer satisfaction response to waiting shorter vs. longer than expected', Journal of Retailing, Vol. 99, No. 2, pp. 247–264.

Chapados, N., Joliveau, M. and Rousseau, L.-M. (2011), 'Retail store workforce scheduling by expected operating income maximization', in Proceedings of the International Conference on AI and OR Techniques in Constraint Programming, Springer, Berlin, pp. 42–57.

Dombrowski, U. and Mielke, T. (2013), "Lean Leadership - Fundamental Principles and their Application", *Procedia CIRP*, Vol. 7, pp. 569–574.

Evans, E. (2020), 'Retailers must not forget their staff are their superpower', International Longevity Centre UK, available at: https://ilcuk.org.uk (accessed May 2026).

Lodgaard, E., Ingvaldsen, J.A., Gamme, I. and Aschehoug, S. (2016), 'Barriers to lean implementation: Perceptions of top managers, middle managers and workers', Procedia CIRP, Vol. 57, pp. 595–600.

Marques, P.A., Carvalho, A.M. and Santos, J.O. (2022), 'Improving operational and sustainability performance in a retail fresh food market using Lean: A Portuguese case study', Sustainability, Vol. 14, No. 1, p. 403.

Marques, P.A., Jorge, D. and Reis, J. (2022), 'Using Lean to improve operational performance in a retail store and e-commerce service: A Portuguese case study', Sustainability, Vol. 14, No. 10, p. 5913.

Martinsuo, M. and Huemann, M. (2021), 'Reporting case studies for making an impact', International Journal of Project Management, Vol. 39, No. 8, pp. 827–833.

Nguyen, V., Nguyen, N., Schumacher, B. and Tran, T. (2020), 'Practical application of Plan-Do-Check-Act cycle for quality improvement of sustainable packaging: A case study', Applied Sciences, Vol. 10, No. 18, p. 6332.

Queiroz, E., Alencar, D., Costa, A. and Santos, M. (2019), 'Application of PDCA cycle methodology in management of continuous improvement in receiving and movement logistics', International Journal of Innovation Education and Research, Vol. 7, No. 11, pp. 980–991.

Patel, P. and Deshpande, V. (2017), "Application Of Plan-Do-Check-Act Cycle For Quality And Productivity Improvement-A Review", International Journal of Research in Applied Science and Engineering Technology, Vol. 5, pp. 197–201.

Reiner, G., Teller, C. and Kotzab, H. (2013), 'Analyzing the efficient execution of in-store logistics processes in grocery retailing - the case of dairy products', Production and Operations Management, Vol. 22, No. 4, pp. 924–939.

Talarico, L. and Maya Duque, P.A. (2015), 'An optimization algorithm for the workforce management in a retail chain', Computers and Industrial Engineering, Vol. 82, pp. 65–77.

Wohlin, C. (2021), 'Case study research in software engineering - it is a case, and it is a study, but is it a case study?', Information and Software Technology, Vol. 133, p. 106514.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

The New Construction Products Regulation (EU) 2024/3110 and the challenges of digitalisation in Quality 4.0

Ana Amaral [1]

1 CTCV – Centro Tecnológico da Cerâmica e Vidro; anasofia@ctcv.pt

The aim of this article is to clarify the changes resulting from the entry into force of the new Construction Products Regulation and the changes that organisations will need to implement to ensure they remain in compliance with the law. For preparing this work, it was carried out a detailed reading and review of the articles of the new regulation, with particular emphasis on the new provisions (e.g. the Digital Passport for Construction Products) and paying special attention to the illustrative annexes setting out the main changes that must be implemented. This paper ia a contribute to help organisations better understand the changes they need to make in order to keep their products on the market.

The new Regulation consolidates harmonised conditions for the marketing of construction products and seeks to eliminate barriers to trade among Member States. It provides standardised information on product characteristics and performance, while also introducing a broader and more explicit focus on sustainability. The principal changes introduced by this Regulation concern digitisation and environmental sustainability, most notably through the inclusion of mandatory environmental requirements for construction products. Whereas the previous regulatory framework was primarily focused on technical performance, the new Construction Products Regulation (CPR) requires manufacturers to assess and declare the environmental, functional, and safety requirements applicable to their products.

Consistent with the principles of the circular economy, the environmental sustainability dimension of the new Regulation is intended to foster the transition towards more sustainable production processes and a more efficient use of natural resources. This approach is grounded in a life-cycle perspective, with particular emphasis on reuse, remanufacturing, and the reduction of waste throughout the sector. CPR 2024 is also aligned with the United Nations Sustainable Development Goals (SDGs), particularly SDG 11 - Sustainable Cities and Communities - and SDG 12 - Responsible Consumption and Production, by seeking to reduce the environmental impact of construction products, promote circularity and energy efficiency, and support the design and construction of more sustainable buildings.

Furthermore, the new Regulation is expected to accelerate the digital transformation of the construction sector and enhance its competitiveness. Its objective is to ensure the availability of structured product information based on open standards and developed in an interoperable format. One of the central pillars of this digital transition is the Digital Product Passport (DPP), conceived as a digital identity tool that facilitates access to and share of product information. This includes data on technical performance, material composition and origin, instructions for use and safety, repair and maintenance activities, recyclability, and environmental impacts across the product life cycle. DPP is one of the upcoming challenges of Quality 4.0.

Keywords – regulation, construction, digital, sustainability

I. Introduction

The new Construction Products Regulation (CPR) – Regulation (EU) 2024/3110 – was published in the Official Journal of the European Union on 18 December 2024. It entered into force in part on 7 January 2025 and, from 8 January 2026, the new Regulation will be in force. However, given the changes it introduces, the period of coexistence with Regulation 305/2011 may be extended, at the latest, until 2032.

An analysis and implementation of Regulation 305/2011 revealed a number of shortcomings, namely:

- A standardisation process with poor performance,

- Unclear provisions and inconsistencies with other EU legislation,

- Weak market surveillance,

- An incomplete approach to the regulatory needs of Member States,

- A lack of a consistent approach to product safety.

To address these shortcomings and introduce emerging issues that the European Union wishes to highlight, Regulation 2024/3110 was drawn up; in essence, it aims to steer the economy towards an increasingly green and circular model.

The new regulation maintains the aim of harmonising rules for the marketing of construction products within the EU, facilitating the free movement of products within a single market through standardised information and characteristics, whilst also addressing the sustainability performance of construction products.

The key changes introduced by the regulation include digitalisation and sustainability, with the introduction of mandatory environmental requirements for construction products. Whilst the previous regulation focused primarily on technical performance, the new CPR requires manufacturers to assess and declare the environmental, functional and safety requirements applicable to construction products.

Of particular note is the introduction of the 3+ System (Control of environmental sustainability assessment by the notified body) into the assessment and verification systems. The 3+ System will apply across the board to all products, regardless of their function on the construction site. All products, in accordance with the requirements of their role on the construction site, already had an assessment and verification system (System 1, 1+, 2+, 3 and 4), which will now coexist in parallel with the 3+ System for the assessment of performance/environmental sustainability. However, the assumptions, the values entered and the modelling associated with the 3+ system will have to be validated with the involvement of a notified body.

In line with the circular economy, the focus on environmental sustainability in the new regulation aims to promote the transition to more environmentally friendly production processes and a more efficient use of natural resources, based on the life cycle of products, with an emphasis on reuse and remanufacturing, and by encouraging waste reduction in the sector. The aim is therefore to direct actions towards the overarching objectives of reducing the environmental impact of construction products, promoting circularity and energy efficiency.

The European Union's new Construction Products Regulation is aligned with the Sustainable Development Goals (SDGs), as it promotes more responsible and environmentally friendly practices in the construction sector. These goals, set by the United Nations, aim to ensure a balance between economic development, environmental protection and the well-being of society.

Under the new rules, construction products must now meet more stringent sustainability criteria. This means they must be more durable, consume fewer natural resources and have a lower environmental impact throughout their life cycle. Furthermore, the regulation encourages the reuse and recycling of materials, contributing to a circular economy and the reduction of waste. Another important aspect is improving the energy efficiency of buildings. By using more efficient materials, it becomes possible to reduce energy consumption, which helps to lower greenhouse gas emissions. In this way, the regulation contributes to the fight against climate change. Finally, greater transparency in product information enables consumers and businesses to make more informed and sustainable choices.

In summary, this regulation is in line with the SDGs because it promotes more sustainable construction, reduces environmental impact and contributes to the creation of more efficient, safe and balanced cities.

The new regulation is set to drive the digital transition in the construction sector, with a view to boosting the sector's competitiveness. But the changes must also include the producers. Several digital technologies [6] are specifically applied in quality control with the goal of reduction in the 'Cost of Non-Quality' (COQ): traditionally, companies spend a great deal of money inspecting products, reworking defective parts or managing customer returns. Quality 4.0 shifts the focus from detection to prevention, drastically reducing raw material waste and rework time. The use of Internet of Things (IoT) and sensors applied to monitor machine variables (temperature, vibration, pressure) in real time, enables deviations in the process to be identified before they result in defective products. Big data and analytics allow mass collection of industrial data that enables the application of machine learning algorithms to predict tool failures and anticipate deviations from quality standards (Predictive Maintenance). Also Cloud Computing reveals essential for centralising quality data from different factories or suppliers, enabling teams anywhere in the world to collaborate and make decisions simultaneously. The application of new digital technologies lead to positive Impacts observed across industries. Decision-making based on real-time data: In opposite to the traditional model, where quality reports are analysed at the end of the shift or week, with Quality 4.0, if a machine starts to go out of calibration, the system issues an immediate alert or self-adjusts, preventing the production of entire batches of defective products. Improved traceability, with the use of technologies such as RFID tags or integrated systems allows the exact history of a product to be traced (from the supplier's raw materials to the end customer), facilitating audits and the rapid resolution of warranty issues.

Despite the benefits, there are some barriers and challenges to implementation [6], such us the lack of standardisation and structure. Many industries attempt to implement isolated technologies (e.g. installing sensors on a machine),but lack a methodology or organisational culture that links this data to the company's overall quality strategy. Another barrier is the 'skills gap'. Traditional quality professionals are proficient in statistics and classic tools (such as Lean Six Sigma), but often lack skills in data analysis, programming or digital literacy. Similarly, data engineers are unfamiliar with manufacturing processes.

Most of the times, there are high initial costs. It is necessary an investment in technological infrastructure, cybersecurity and staff training, especially for Small and Medium-sized Enterprises (SMEs).

There is also the aim for information to be structured, based on open standards and developed in an interoperable format – that is, transferable via an open data exchange system that is vendor-neutral – so that communication allows information from all sources to be read.

The acquisition, collection, processing and analysis of data will require companies to restructure, particularly as the digitisation of records has not yet taken place in many cases.

The transition to Industry 4.0 must go hand in hand with Quality 4.0, to bring all the benefits of operational improvement into the company, making organisations more efficient and effective, and enabling and encouraging Data-Driven Decision-Making. For example, this involves introducing improvements linked to the use of AI (Artificial Intelligence) for fault prediction, the use of IoT (Internet of Things) to connect devices that allow continuous monitoring, and the implementation of Augmented and Virtual Reality for product inspection. Moving from a basic level of digitalisation to exploiting its full potential, as outlined in Quality 4.0, is the challenge that companies need to address.

II. Digital Product Passport

One of the central pillars of digitalisation is the creation of the Digital Product Passport (DPP), a digital identity tool designed to make product information available and facilitate its sharing (e.g. technical performance, materials and their origin, repair activities, recyclability, and environmental impacts throughout the product's life cycle). Article 76 of RPC 2024/3110 further states that the DPP must include what is described in Fig.1: the declaration of performance and conformity, information on the environmental impact of the product's life cycle (climate change, eutrophication, acidification, etc., and labelling regarding environmental sustainability) and technical details on safety (instructions for use, safety, hygiene and health information, etc.).

Figure 1 – Content of Digital Product Passport

The DPP aims to promote transparency for manufacturers and consumers, improve the management of complex information and facilitate product traceability throughout the value chain. The structure of the DPP has not yet been finalised; details are pending the adoption of a delegated act, which is expected to be published in mid-2026. Eighteen months after the publication of this delegated act, the DPP will become mandatory. The introduction of the DPP will act as a catalyst for the sector's digitalisation and will be the driving force behind a broader digital transition, in line with other EU initiatives to digitise documentation. The introduction of the DPP will require the development and updating of digital solutions within companies, so as to adapt their data storage and sharing management to ensure compliance with information availability requirements. At the same time, digitisation will open up new opportunities to improve operational efficiency.

The introduction of the Digital Product Passport will require companies to develop and update their digital solutions in order to adapt their data storage and sharing processes, thereby ensuring compliance with the requirements for providing information. Digitisation will form the basis for the creation of a comprehensive database, which is intended to serve as a repository for all the information contained in the Product Data Sheets (PDS) and to enable the technical and environmental performance of construction projects using these products to be assessed.

The new Regulation presents companies with a range of challenges and opportunities. On the one hand, there is a need to adapt to new environmental requirements and other requirements that may arise in harmonised product standards. Digitalisation will lead to the development of digital infrastructure and investment in monitoring systems, which will support the creation and maintenance of the Digital Product Passport.

The introduction of these changes will also serve as a catalyst for improving process efficiency and reviewing the production control system, providing an opportunity for innovation and differentiation, as well as a boost to training and adaptation to the new reality.

The existence of Digital Product Passport for construction products will lead to the creation of a comprehensive database that will store a large amount of information contained in the many DPP and will permit the calculation of technical and environmental performance of buildings using these products.

The location, management and operation of this database raise issues regarding security, access rights, storage, and ensuring the reliability and integrity of the data, amongst others. For this reason, standards are being developed for the aforementioned issues, relating to the management of the database.

Currently, emerging digital methodologies in the construction sector, such as Building Information Modelling (BIM), widely used by specifiers and builders, are already facilitating full digital processing, the benefits of interoperability and a path towards realising the transformation paradigm.

The convergence between the DPP and BIM represents one of the most significant steps towards the digitalisation and sustainability of modern construction.

What may at first glance appear to be competing technologies should, in fact, be viewed as symbiotic partners. Whilst BIM organises the geometry and data of the building as a whole, the DPP acts as the detailed 'identity document' for each component that goes into that structure. The role of the BIM-DPP combination is very important. The two tools are interrelated and transform the life cycle of a project.

In BIM ecosystem, we work with digital objects that represent real-world elements (a heat pump, a steel beam, a window). Traditionally, properties are entered manually into these objects (manufacturer, power rating, dimensions). With the DPP, this relationship shifts to a smart hyperlink model. Each BIM element now contains the DPP's unique identifier (for example, a QR code or digital ID). Instead of overloading the BIM model with gigabytes of PDF technical data sheets, BIM links directly to the manufacturer's DPP, which is hosted in the cloud and constantly updated. The European Union has been strongly promoting the DPP to ensure sustainability.

A major advantage of integrating BIM with the DPP is the traceability and information regarding the product's circularity (cradle-to-cradle), which enables a dynamic inventory of the materials that make up a building. By analysing the different stages of construction, it is possible to appreciate the resulting benefits. During the Design and Specification stage, the designer can select a BIM object whose DPP ensures that the carbon footprint (as stated in the product's environmental product declaration – EPD) meets the project's targets. During Construction and Installation, the builder installs the physical component, scans the QR code on site and automatically verifies in the BIM model that the installed product matches the specification exactly.

In Operation and Maintenance (Asset Management) phase, the maintenance team accesses the BIM model and, via the DPP, can instantly view the repair history, compatible replacement parts and the manufacturer's safety instructions. At End of Life / Deconstruction, the building becomes a 'materials bank'. Before demolition, the owner extracts a list of all steel beams or concrete panels from the BIM. The DPP specifies exactly how they were manufactured, whether they contain hazardous substances, and how they can be recycled or reused.

The BIM-DPP integration is of enormous use in Real Data-based Asset Management. In Facility Management, the BIM-DPP link saves hundreds of hours of work. For example, if a light fitting breaks down, the technician does not need to search for information in dusty files. They click on the light fitting in the BIM model, open the associated DPP and have instant access to electrical diagrams updated by the manufacturer, safe dismantling manuals and energy efficiency certifications.

For this relationship to work in practice, engineering relies on open standards (Open BIM). The IFC (Industry Foundation Classes) format of BIM is extended to include properties that link the model to DPP data standards (such as structures based on ISO and GS1 standards for unique identification). It can be concluded that BIM provides the spatial and operational context (where the product is and how it relates to the rest of the building), whilst the Digital Product Passport provides the content and history (what the product is, where it came from and how it can be reused). Together, they transform buildings from static structures into circular data ecosystems.

III. KEY CHANGES IN THE NEW CONSTRUCTION PRODUCTS REGULATION

Regulation (EU) 2024/3110 introduces changes to the declarations of performance, which were originally established by Regulation 305/2011; these are now referred to as declarations of performance and conformity and must be drawn up in accordance with the model set out in Annex V of the new regulation. In addition to information on origin, manufacturer, intended use and technical performance, the new declaration of performance and conformity will now include "the product's environmental sustainability performance throughout its life cycle with regard to the environmental characteristics set out in Annex II" – Article 15(2).

The introduction of the essential environmental characteristics will take place in phases, in accordance with the timetable set out in article 15 and Fig. 2, with the main deadlines set for 2026, 2030 and 2032. However, this forecast may be subject to change, as it is contingent upon the publication of harmonised product standards.

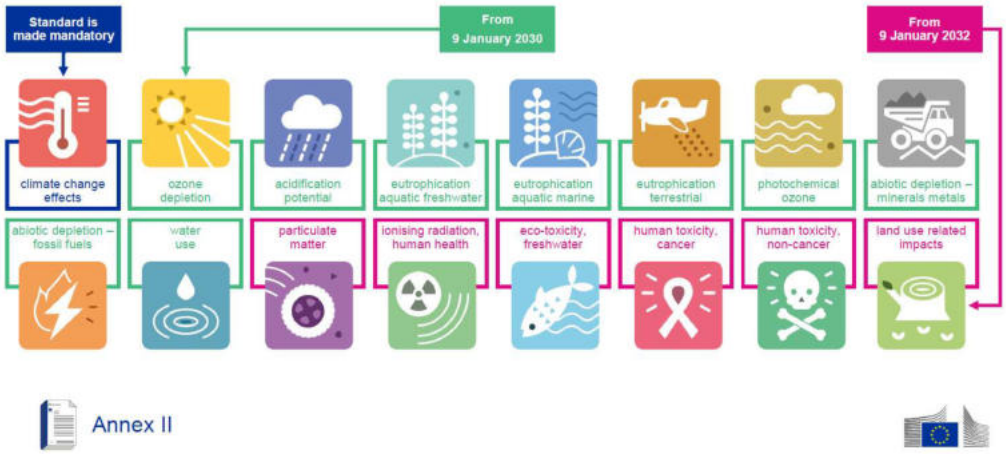


Figure 2 – Environmental sustainability essential characteristics

IV. FORECAST CATEGORIES OF ENVIRONMENTAL IMPACT

The predefined categories of environmental impacts, relating to the life-cycle assessment of a product, are set out in Annex II to RPC 2024/3110 and are as follows:

a) Climate change impacts — total;

b) Climate change impacts — fossil fuels;

c) Climate change impacts — biogenic;

d) Climate change impacts — land use and land-use change;

e) Ozone depletion;

f) Acidification potential;

g) Freshwater eutrophication;

h) Marine eutrophication;

i) Terrestrial eutrophication;

j) Photochemical ozone;

k) Abiotic depletion — minerals, metals;

l) Abiotic depletion — fossil fuels;

m) Water use;

n) Suspended particulates;

o) Ionising radiation — human health;

p) Ecotoxicity — freshwater;

q) Human toxicity — carcinogenic;

r) Human toxicity — non-carcinogenic;

s) Impacts related to land use.

These categories of environmental impact already correspond to those set out in the Product Category Rules specified in EN 15804+A2+ AC/2021, which form the basis for drawing up the Environmental Product Declaration (EPD), thereby ensuring transparency in the assessment of products' environmental performance throughout their life cycle (Figure 3). This EPD tool has been widely sought after for the life-cycle assessment of a product, as well as used to quantify its integrated performance within the respective building or construction project. Furthermore, the EPD includes a range of indicators for renewable and non-renewable energy, waste and incorporated secondary materials, amongst others, in order to quantify these indicators, thereby promoting and enhancing the circular and more sustainable economy.

V. SUSTAINABILITY CRITERIA IN SUPPLIER SELECTION

Given the financial scale of public procurement in Member States, the RPC stipulates that public procurement procedures must take into account and incorporate sustainability criteria when selecting suppliers. The Regulation refers to green public procurement (Article 83) and aims to influence public procurement decisions based on

mandatory minimum environmental sustainability requirements for construction products. These are set out in delegated acts applicable to public contracts, which may take the form of technical specifications, selection criteria, contract performance conditions or award criteria. The new regulation thus promotes more sustainable practices in public projects and influences the entire construction market.

VI. Market surveillance

Another key area covered by the new regulation is market surveillance. Regulation 2024/3110 will enable stricter product surveillance (in conjunction with Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and product compliance). In addition, it establishes enhanced powers for competent authorities, stricter penalties for non-compliant products, better consumer protection (with the creation of a complaints portal) and greater cooperation between EU Member States, notably through the ADCO Groups.

Finally, the publication of the new regulation is in line with the European Union's strategic goals and objectives. Regulation 2024/3110 is consistent and appears to be integrated with other European Union legislation and policies, notably the Ecodesign Regulation for Sustainable Products, the European Green Deal, the Energy Efficiency Directive and the Circular Economy Action Plan, amongst others.

VII. General implementation timetable for the new RPC

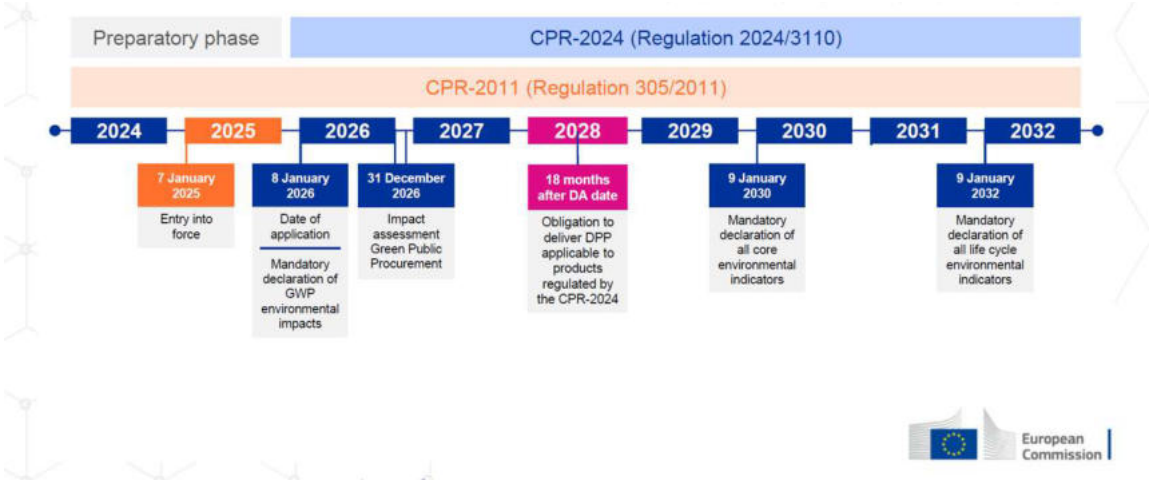


Figure 3 – Schedule for the transition from the previous to the New Construction Products Regulation

A review of the harmonised standards is planned to ensure they are aligned with the new regulation over an extended period (see Figure 2), and their update should include:

- Alignment with technological advances and best available techniques,

- Greater adaptability to market needs,

- A more active role for the European Commission,

- Introduction of environmental characteristics (consistent with EN 15804 – Environmental Product Declarations and cPCR – complementary Product Category Rules).

The Commission has set up several Acquis groups to carry out the work of revising harmonised standards. The aim is also to define a technical structure for the documentation that ensures consistency across products and materials. At the same time, the list of product families set out in Annex VII was analysed to prioritise them, according to the products' function in the construction works and their interactions. The purpose was to establish prioritisation for the revision and updating of the respective harmonised standards, in accordance with its role in construction. At the moment, only the standards for the 10 priority products are already under review, with the remaining 26 families awaiting their turn.

REFERENCES

Almeida, M. (2019). Environmental performance of products in the ceramics sector in Portugal. Tese de Doutoramento. Universidade de Aveiro.

Almeida, M.I., Dias, A.C., Demertzi, M. e Arroja, L. (2016). Environmental profile of ceramic tiles and their potential for improvement. Journal of Cleaner Production, 131, pp. 583–593. Disponível em: https://doi.org/10.1016/j.jclepro.2016.04.131.

CEN (European Committee for Standardization) (2019). EN 15804:2012+A2:2019+AC: Sustainability of construction works - Environmental product declarations - Core rules for the product category of construction products. Bruxelas: CEN.

Environmental product declarations (EPD) - Key Sustainability Factor (2021). [Vídeo online]. Disponível em: https://www.youtube.com/watch?v=hpysho6nj68 [Acedido em: 21 Mai. 2026].

ISO (International Organization for Standardization) (2006). EN ISO 14025: Environmental labels and declarations — Type III environmental declarations — Principles and procedures. Genebra: ISO.

ISO (International Organization for Standardization) (2018). ISO 19650-1: Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling - Part 1: Concepts and principles . Genebra: ISO.

ISO (International Organization for Standardization) (2018). ISO 19650-2: Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling - Part 2: Delivery phase of assets . Genebra: ISO.

ISO (International Organization for Standardization) (2020). ISO 19650-3: Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling - Part 3: Operational phase of assets. Genebra: ISO.

Magodi, A., Mashamba, T. e Mokgohloa, K. (ano de publicação pendente). Impact of Quality 4.0 Implementation in Industries: A Systematic Literature Review. The South African Journal of Industrial Engineering [Aguardando detalhes de Volume/Fascículo/Páginas].

União Europeia (2024a). Regulamento (UE) 2024/1781 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, que estabelece um quadro para a definição dos requisitos de conceção ecológica aplicáveis aos produtos sustentáveis, que altera a Diretiva (UE) 2020/1828 e o Regulamento (UE) 2023/1542 e que revoga a Diretiva 2009/125/CE. Jornal Oficial da União Europeia, L/2024/1781.

União Europeia (2024b). Regulamento (UE) 2024/3110 do Parlamento Europeu e do Conselho, de 27 de novembro de 2024, que estabelece regras harmonizadas para a comercialização dos produtos de construção e que revoga o Regulamento (UE) n.º 305/2011. Jornal Oficial da União Europeia, L/2024/3110 [Capítulos I, II, III, VIII, X, XII; Anexos II, III, V, VII, IX].

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Leveraging Interdeparture Times for Bottleneck Detection and Process Improvement

Pedro Costa [1], Ana Moura [2], Galina Robertsone [3] and Tatjana Tambovceva [4]

[1] University of Aveiro - Center for Research & Development in Mathematics and Applications, Portugal; pedroccosta@ua.pt
[2] University of Aveiro - Center for Research & Development in Mathematics and Applications, Portugal; ana.moura@ua.pt
[3] Faculty of Engineering Economics and Management, Riga Technical University, Latvia; galina.robertsone@rtu.lv
[4] Faculty of Engineering Economics and Management, Riga Technical University, Latvia; tatjana.tambovceva@rtu.lv

***Abstract*— This study aims to enhance production throughput and operational stability in an assembly line within a Portuguese manufacturing company. To achieve this, the research focuses on identifying and mitigating the primary throughput bottleneck, guided by the Theory of Constraints. The Interdeparture Time Variance Method, a data-driven bottleneck detection approach, was applied using production data to locate the constraint. Recognizing the limitations of variance-based metrics over short observation windows, the quantitative analysis was supplemented with direct observation and time studies to diagnose the root causes of the bottleneck. The evaluation revealed that workload imbalances, insufficient buffer capacities, and non-standardized tasks significantly constrained production flow. Consequently, targeted improvement measures were proposed, including a redesigned line layout, rebalanced operator tasks, revised operating standards, and regulated buffer sizes. Implementing these modifications is projected to reduce the bottleneck's cycle time, thereby increasing throughput from 35 to 40 units per hour. Ultimately, this research demonstrates that while data-driven bottleneck detection methods provide valuable quantitative insights, integrating them with qualitative workflow assessments is essential for capturing the full operational complexity of dynamic manufacturing environments. This study contributes with practical application of bottleneck detection and exploitation methodologies to help organizations improve complex assembly lines, mitigate the shifting nature of bottlenecks, and establish a more predictable production flow.**

Keywords – bottleneck detection, theory of constraints, interdeparture time variance method, throughput bottleneck

I. INTRODUCTION

The term "bottleneck" originates from the narrowest part of a bottle, where the restricted opening slows the flow of liquid compared to the rest of the container. In manufacturing, a throughput bottleneck is the point in a process that constrains the overall production rate. Within industrial settings, the concept of a bottleneck has been defined in various ways. Different researchers offer distinct interpretations, and even the same author may adopt multiple definitions across different applications and works (Kuo et al. (1996). This variation often arises because the criteria for identifying a bottleneck can vary depending on the bottleneck detection method (BDM) used. Despite these differing viewpoints, there is consensus that a bottleneck is the element within a process that restricts overall throughput.

For the purposes of this study, a bottleneck is defined as:

The resource that, most of the time, limits the process' throughput.

Conversely, a resource can be understood as an agent assigned to a specific set of activities, defined by its interfaces with other agents, and capable of independently processing a unit. When several resources share the same set of assigned activities, they are treated as a single resource group for analytical purposes. In such cases, the corresponding cycle time (CT) is divided by the number of identical resources to ensure clarity and accuracy in analysis. This definition also accounts for the complexity introduced by shifting bottlenecks, a topic discussed in greater detail later.

The Theory of Constraints (ToC) broadens the traditional notion of a bottleneck by framing it as a constraint, i.e., any factor that limits the overall capacity of a production system (Goldratt and Cox, 2016). According to ToC, the throughput of an entire manufacturing operation is governed by its most restrictive constraint. Therefore, to boost productivity, it is crucial to identify and address this limiting factor. It is important to recognize that a production system will always have at least one constraint, as when one is mitigated, another inevitably emerges to take its place. The key, then, is to continuously identify and alleviate the current constraint, thereby initiating a cycle of continuous improvement.

This study focuses on improving production throughput in the L35 assembly line of a manufacturing company located in Portugal. The L35 line, responsible for assembling boiler control panels, operates under high product variety and complex interdependencies among operators. These characteristics hinder the identification of bottlenecks and the achievement of balanced workloads, thereby limiting overall system performance.

While competent applications of BDMs are common, the primary contribution of this work lies in bridging the gap between purely quantitative data-driven bottleneck detection and the qualitative realities of highly variable, operator-dependent manufacturing environments. This study evaluates the Interdeparture Time Variance Method (ITVM) in a context characterized by shifting bottlenecks and complex operator interdependencies. To achieve this, the paper addresses the following research questions: (1) How can ITVM be effectively applied to identify throughput bottlenecks in assembly lines with limited data granularity? and (2) How do operator autonomy and non-standardized tasks influence the formation and shifting of these bottlenecks? By addressing these questions, this research generates transferable knowledge regarding the limitations of

variance-based metrics and demonstrates that quantitative BDM outputs must be coupled with qualitative workflow assessments to reliably guide process improvement in non-static systems.

This article is structured as follows: first, the research context and objectives are presented. Next, relevant concepts and methods are reviewed. This is followed by a description of the research methodology, an analysis of the results, and a discussion of their implications. Finally, conclusions are drawn, and directions for future research are suggested.

II. Literature Review

Constraints can generally be classified into three categories (Urban and Rogowska, 2020):

- Resource: These include internal limitations such as a machine with a lower capacity than others or an operator responsible for the most time-consuming tasks. Resource constraints can also be external, such as a marketing strategy that generates more demand than the system can fulfil.
- Material: These arise when raw materials are unavailable, causing the production process to slow down or halt entirely to match the supply rate.
- Policy: These are the result of management decisions or organizational policies. For example, the insistence on perfectly balancing production with demand can sometimes hinder the achievement of broader business goals.

For the purposes of this article, the focus will be narrowed to resource-based constraints, specifically throughput bottlenecks, to maintain a clear and manageable scope.

In theory, the process bottleneck should remain constant, determined by the stage with the longest processing time. However, this assumption holds true only in static systems: environments free from variation or unexpected disturbances, which are rare in real-world manufacturing (Roser and Nakano, 2015). In dynamic production settings, the location of the bottleneck can shift due to factors such as fluctuating process times, particularly in manual operations, or unpredictable disruptions, such as equipment failures and other random events (West et al., 2022). The more evenly balanced a system is, the more susceptible it becomes to these shifts, as the lack of buffers between activities increases the system's sensitivity to irregularities.

Bottlenecks can be classified as either short-term or long-term. A short-term bottleneck temporarily constrains the process, with its duration ranging from a single production shift to a week (Chang et al., 2007). Such bottlenecks are often caused by transient issues, like a machine breakdown that lasts for one shift. In contrast, a long-term bottleneck consistently limits system throughput over extended periods, typically exceeding a month (Chang et al., 2007). Long-term constraints usually stem from resources with significantly higher CTs than others, making them the most common bottleneck in the long run.

Both types of bottlenecks offer valuable insights. Long-term bottlenecks are crucial for the planning and design of manufacturing systems, while short-term constraints help identify special causes of disruption and enable rapid response to emerging issues (Chang et al., 2007).

To accomplish the first main objective, the identification of the throughput bottleneck, it is necessary to select an appropriate BDM. After the ToC was introduced by Goldratt and Cox in 1984, various articles began to appear, each proposing different techniques for detecting bottlenecks.

Tang et al. (2024) classify BDMs into three main categories:

- BDMs based on static models;
- BDMs based on simulations;
- BDMs based on data-driven approaches.

As mentioned earlier, truly static systems are rarely found in practice. Nevertheless, static model-based BDMs can still be valuable for identifying long-term bottlenecks, which can be important for system planning and design. Many companies are still in the process of transitioning to Industry 4.0, meaning that not all operational data are yet integrated into information systems. For these organizations, static model-based BDMs remain the most practical option for pinpointing throughput bottlenecks, even though they do not provide short-term analysis capabilities. At this stage of industry maturity, the most frequently used methods are the Process Time Method, which identifies the bottleneck as the resource with the highest activity CT (Roser et al., 2015), and the Utilization Method, which selects the resource with the highest ratio of effective process time to total production time (Betterton and Silver, 2012).

Simulation-based BDMs are helpful for detecting throughput bottlenecks while accounting for their variability due to process changes and other influencing factors. Although these methods tend to be more precise than static model-based approaches (Tang et al., 2024), they require considerable time to develop (Subramaniyan et al., 2016) and reflect only the factors explicitly included in the simulation model (Tang et al., 2024).

Among the three main categories, data-driven approaches are considered the most effective for bottleneck detection (Tang et al., 2024). By leveraging real-time data from manufacturing information systems, these methods can accurately capture system dynamics and identify bottlenecks, whether short- or long-term (Lai et al., 2021). Compared to simulation methods, data-driven approaches typically require less setup time and offer higher precision (Lai et al., 2021). However, their effectiveness depends on the availability of technology to collect real-time shop-floor data, which is still lacking in many companies (Tang et al., 2024).

When real-time data is limited, data-driven methods can be adapted to suit the available information (Brochado et al., 2022). Over the past decade, numerous data-driven methods have been developed, often as adaptations or enhancements of existing techniques. According to several reviews, the most frequently cited data-driven BDMs include:

- Active Period Method: Monitors process status over time, categorizing each as active or inactive. The process with the longest continuous active period is identified as the bottleneck (Roser et al., 2017).
- Arrow Method: Identifies bottlenecks based on the likelihood of blockage, when a resource cannot process until the next resource consumes buffer, and starvation , when a resource cannot process until it receives input from the previous resource (Shu-Yin Chiang et al., 2000).
- Bottleneck Walk Method: Involves real-time observation of process and inventory states along a production line. Data collected during this walk is analyzed to identify patterns and constraints, resulting in a prioritized list of bottlenecks (Roser et al., 2015).
- ITVM: Detects the bottleneck as the resource with the lowest variance in interdeparture times (IT), indicating it spends most of its time neither starved nor blocked (Betterton and Silver, 2012).
- Inventory-Based Methods: Use metrics such as total waiting time, average waiting time, and queue length between processes to identify bottlenecks (Roser and Nakano, 2015).
- Turning Point Method: Similar to the Arrow Method, this approach uses blockage and starvation times to locate the bottleneck. The bottleneck is identified where the difference between blockage and starvation times shifts from positive to negative, and the combined total of these times is less than those of adjacent processes (Li et al., 2009).

In summary, the literature underscores the importance of accurately identifying bottlenecks to enhance production efficiency. Over the years, a range of detection methods has been proposed, from static models to advanced data-driven techniques, each with their own strengths and limitations. While data-driven approaches offer the greatest accuracy, their dependence on real-time data collection technology makes them most suitable for highly mature, Industry 4.0-enabled environments. For settings with limited data, practical solutions like the Process Time Method remain valuable, allowing for structured bottleneck identification and process improvement without significant technological investment.

However, a significant gap remains in the literature regarding the empirical application of data-driven BDMs, specifically in environments that fall between low-tech manual assembly and fully mature Industry 4.0 systems. Previous studies have predominantly validated ITVM using simulations or finished goods data, leaving its operational robustness in real-world, station-level environments underexplored. This study addresses this limitation by applying the ITVM to an environment where shifting bottlenecks are heavily influenced by operator behavior rather than just machine capacity. ITVM was selected over alternative data-driven methods, as its variance-based logic theoretically requires only departure timestamps, making it uniquely suited for legacy systems lacking continuous real-time status monitoring. By positioning this study within the scarce empirical applications of ITVM, this paper links the theoretical promise of variance-based bottleneck detection to the practical necessity of qualitative validation by process observation and interpretation.

III. Bottleneck Identification

The methodology for this work is organized into two main chapters. The first chapter focuses on identifying the throughput bottleneck within the process, utilizing a data-driven approach. The second chapter addresses the exploitation of the identified bottleneck, detailing the targeted improvement actions designed to enhance overall process performance.

A. Process Description

The finished product, Heatronic 3, is a boiler control panel, for now on called product X. For the purposes of this study, only one model, model *435, will be considered, delivered as a fully assembled unit.

The assembly process of product X on the L35 production line is organized into a sequence of operations performed by two operators (Op1 and Op2) across multiple workstations.

The process begins with Op1, who loads the printed circuit board assembly (PCBA) into the in-circuit test (ICT) equipment. Upon completion of the test, Op1 removes the PCBA, attaches the display module, and places the unit into the first pressing station (MER) to remove excess PCB material. After this operation, the PCBA is transferred to the subsequent workstation.

At the FA1 station, Op1 initiates the mechanical assembly by positioning a housing unit in an assembly fixture. Op1 then installs two dials, a label, and two keypads. The partially assembled unit is subsequently placed in a second press, where a nut and a contact spring are assembled onto the housing. During the press cycle, Op1 performs a parallel task: preparing an earth terminal by fastening two screws to a metal component. This subassembly is then mounted and secured onto the housing. Following this, the PCBA is integrated into the housing, and the semi-finished product is transferred to the next station, where it is handed over to Op2.

Op2 begins at the FA2 station by assembling the back cover onto the unit and securing it with screws. At the FA3 station, Op2 prepares an additional back cover and positions it, together with six cable glands, in a third press, where the assembly is performed automatically. Subsequently, Op2 supplies one of two available functional test (FCT) stations, removing the previously tested unit and loading the new one for automatic testing. After completion of the FCT, a label is applied, and the unit is transferred to the high-voltage test (HVT) station, where the same replacement procedure is followed.

In the next stage, Op2 places the unit in an assembly fixture, where additional components are installed, including a label, a variant mask, and three insertion elements. The unit then proceeds to another assembly station, where a back lid is positioned,

two labels are applied, and two clamps are secured with screws. Subsequently, Op2 supplies one of two automatic optical inspection (AOI) stations by positioning the unit, along with the prepared back lid and a side lid, initiating the inspection cycle. After completing this sequence, Op2 returns to the initial workstation.

In parallel, Op1 continues the process by assembling the components that have successfully passed the AOI stage. The final product is then transported to the packaging area, where it is placed into a collective container. Op1 subsequently returns to the starting position, and the production cycle is repeated.

During the assembly process, stochastic tasks, such as supplying the line with components or transporting units for repair, are not governed by strict standardized work instructions. Instead, these responsibilities are assigned flexibly, negotiated in real-time by the operators based on their perceived available time. While this collaborative approach allows operators to self-organize, this lack of standardized allocation inherently introduces unpredictable process variability and dynamic shifts in workload distribution.

B. Selection of the BDM

The ITVM was chosen for the L35 assembly line primarily because of its high flexibility and adaptability to specific, unusual scenarios. The L35 line operates under high product variety and complex interdependencies among operators, resulting in unusual work dynamics. Because the ITVM is a data-driven approach designed for dynamic systems, it can effectively capture these complexities by statistically analyzing the available real-time data. This adaptability makes it significantly more adequate than other methods for accurately identifying the true throughput bottleneck in such a highly variable production environment.

The ITVM is a data-driven approach that relies on the statistical analysis of the intervals between WIP departures at different stations within a production line (Betterton and Silver, 2012). The underlying principle is that a bottleneck station, due to its consistently high utilization and minimal idle time, will display the lowest variance in ITs compared to other stations.

This behavior arises because the bottleneck operates at, or near full capacity, experiencing fewer occurrences of blocking, unable to pass work forward, and starvation, waiting for work from the previous station. In contrast, stations upstream of the bottleneck tend to accumulate WIP and are frequently blocked, while downstream stations are more likely to be starved and idle. As a result, non-bottleneck stations exhibit greater fluctuations in their ITs, whereas the bottleneck maintains a more stable outflow, reflected in a lower variance.

The ITVM identifies the bottleneck as the station with the smallest interdeparture time variance (ITV) over a chosen period. This aligns with the definition of the bottleneck as the resource that most limits throughput during that time frame. The method uses variance rather than standard deviation, which increases sensitivity to differences in ITs because it is measured in squared units.

Although ITVM is intuitive and straightforward to implement using time-stamped production data, there is limited evidence in the literature of its application in real-world manufacturing environments, with most studies focusing on finished goods rather than individual stations (Betterton and Silver, 2012). Therefore, this project also serves as a practical test of ITVM's effectiveness in an actual industrial setting, potentially validating it as a reliable and accessible tool for bottleneck detection.

C. Initial Process Evaluation

The initial step in applying the ITVM, as with other data-driven approaches, is data collection. For this method, the only required inputs are the departure times from each machine or workstation. The ease of this step varies depending on the company's digitalization maturity. In some cases, production lines are already equipped to collect clean, structured manufacturing data, making data gathering straightforward. However, in this case, the data recorded by the machines were originally intended for quality control, not for bottleneck analysis.

The only available data on the production line came from the quality control tests (ICT, FCTs, HVT, AOIs), the MER, and the packaging printer scanner (PAK). For each machine cycle, the last recorded timestamp was manually extracted from the manufacturing information system. The dataset was limited to the second shift across three continuous days. Data was collected under standard operational conditions, characterized by regular operator rotations and typical stochastic interruptions, to ensure the sample reflected the true operational consistency of the line. The granularity of the data was constrained to the completion timestamps generated at the end of specific automated testing cycles, meaning the time horizon captured the cumulative variance of all preceding manual tasks. This data collection process resulted in a discrete series of departure timestamps for each individual DRD.

Next, the ITs between each consecutive WIP departure were calculated for each machine. ITs that occurred immediately before and after planned stops, such as morning breaks and lunch, were excluded, as these would inflate the ITVs and would not contribute meaningful information to the analysis, since such stops are routine and expected.

With only three days of data, it was not possible to determine whether a machine was a long-term bottleneck due to frequent malfunctions or false negative test results. Therefore, the analysis was not conducted day by day to identify short-term bottlenecks. Instead, since the aim was to identify a long-term bottleneck and provide a sustainable solution for throughput improvement, the entire dataset was cleaned of outliers. This was done to obtain a representative sample of the assembly line's operation under normal conditions, excluding unplanned stops such as machine breakdowns or maintenance. Outliers were removed using the boxplot method, which involves calculating the interquartile range (IQR) by subtracting the first quartile (Q1) from the third quartile (Q3). This careful data preparation ensured that the analysis focused on typical process behavior, allowing for a more accurate identification of the true long-term bottleneck.

The datapoints were identified as outliers if they fell out of the interval $[Q1 - 1{,}5 \times IQR;\ Q3 + 1{,}5 \times IQR]$. A sample of 300 ITs was randomly picked from the outlier-free dataset, so the sample size for all models would be the same. The ITVs for this

analysis were calculated using the sample variance formula, since this dataset no longer represented the entirety of the population.

$$ITV_{sample} = \frac{\sum(x_i - \bar{x})^2}{n - 1} \quad (1)$$

Where ITV_{sample} is the sample variance; $x_1, x_2, \ldots, x_i$ are the observations; n is the size of the sample; and $\bar{x}$ is the population mean.

Ideally, each workstation on the line would be equipped with its own data-recording device (DRD) to enable detailed process monitoring. However, out of the 11 workstations, only 6 can record data (Fig. 1).

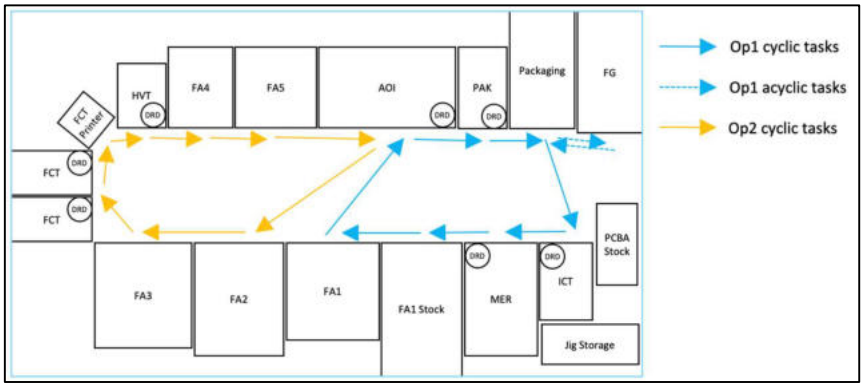


Fig. 1. Initial line layout

As a result, each ITV reflects not only the activities at its corresponding DRD workstation, but also all tasks performed since the previous DRD. For example, the ITV for the AOI station encompasses not only AOI supply activities but also those carried out in FA4 and FA5. Consequently, the analytical focus of this study must adapt to the available data. Because data collection relies on machine-level DRDs, the initial unit of analysis for the ITVM is the physical workstation. However, the ITVM does not pinpoint a single resource or specific manual activity as the bottleneck; rather, it highlights a constraint zone between DRDs. To move from bottleneck identification to actionable improvement, the unit of analysis must subsequently shift to the individual operator and their specific activities within that zone, utilizing direct observation and time studies to isolate the root cause.

The ITVM results are presented in Table 1.

TABLE I. ITVM RESULTS; THE IDENTIFIED BOTTLENECK IS HIGHLIGHTED IN RED

Resource	Op1	ICT	Op1	MER	Op1/2	FCT	Op2	HVT	Op2	AOI	Op1	PAK
ITV		841		1193		1424		1093		1760		2266

With the initial throughput bottleneck for model *435 identified, the next step is to analyze the underlying factors contributing to this constraint. This will provide the basis for suggesting targeted improvements to enhance the assembly line's performance.

IV. BOTTLENECK EXPLOITATION

A. *Factors Identification and Evaluation*

Having quantitatively established the bottleneck location using the ITVM, the analysis must now transition to a diagnosis of its root causes. As the sample indicates that the bottleneck for model *435 is most likely either Op1 or the ICT station, further investigation is required to untangle the operational behaviors driving this constraint.

In long-term bottleneck analysis, the key metric to consider is the CT. Over extended periods, the resource with the highest CT typically emerges as the true bottleneck, consistently limiting the system's maximum throughput. Unlike short-term analyses, which can be influenced by temporary disruptions or fluctuations, long-term evaluations average out these variations and highlight the underlying, persistent constraints of the process.

To accurately assess the CT of each resource involved in model *435's assembly, time studies were conducted for all relevant activities. The start and end points of each activity were clearly defined, and an initial sample of 15 time measurements was collected for each. Since the objective was to estimate the mean of a continuous variable within a single population, the minimum required sample size for each activity was determined using the approach described by (Abebe, 2022):

$$N = \left(\frac{Z_{a/2} \times s}{d}\right)^2 \quad (2)$$

Where N is the minimum sample size; $Z_{a/2}$ is the critical value for a given confidence level; s is the standard deviation of the initial sample; and d *is the* acceptable margin of error.

Given that the process under study is a production process, experts recommend using a confidence level of 95% to 99% and an acceptable margin of error of 3% to 5%. For this analysis, considering the available time frame, a 95% confidence level and a 5% margin of error were selected.

The minimum sample size for each activity was determined accordingly. However, the sampling process proved more complex than anticipated due to recurring deviations from the norm, which further highlighted the potential role of these deviations in causing process instability. Although the CTs of the tests exhibited minor fluctuations, these were not significant.

The average duration of each activity was summed for each operator. Subsequently, a 10% allowance, the company's standard, was added exclusively to the operator's CT. The CTs of all process resources were then visualized in a bar chart (Fig. 2).

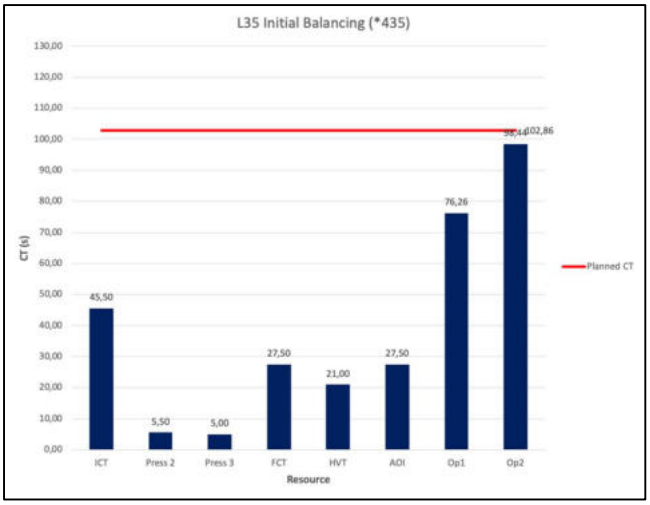


Fig. 2. Initial line balancing

The identified bottleneck does not correspond to the resource with the highest CT. This indicates that factors beyond process time are disrupting the production flow and limiting throughput.

Direct observation of the assembly process revealed several contributors to these disturbances. First, stochastic tasks, such as supplying the line, transporting units for repair, or changing finished goods pallets, introduce variability. The buffer between Op1 and Op2 for model *435 is limited to just one unit, which may be insufficient. When one operator is occupied with these additional tasks, the other can experience starvation or blocking and may end up performing the entire process alone. This situation can cause the CT to rise significantly, as the operator effectively becomes the bottleneck and works at nearly twice the usual CT.

Second, the inherent variability of certain activities also impacts performance. Mistakes during delicate operations, such as screwing components or applying labels, can significantly slow the process.

Third, a lack of standardization further disrupts workflow. Operators do not always follow standard procedures; for instance, Op1 may batch-produce initial stages before completing later steps, destabilizing the production flow and causing slowdowns.

Fourth, process dependencies add complexity. For model *435, the Op1 and Op2 interfaces occur at two key points: after FA1, when a unit is handed off to Op2, and after the AOI, when a unit is returned to Op1. These handoff points can become sources of starvation or blocking for both operators (Fig. 3), increasing process variability and unpredictability. This was confirmed during observation, as Op1 would frequently assist Op2 with the first or last task to prevent being blocked or starved.

These factors contribute to the discrepancy between the resource with the highest CT and the actual bottleneck, highlighting the importance of considering operational dynamics alongside process times in bottleneck analysis.

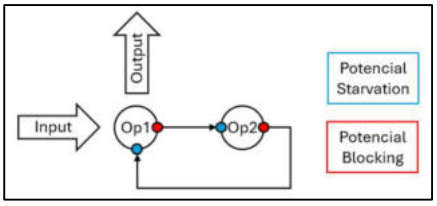


Fig. 3. Initial process dependencies between operators

Line balancing also plays a crucial role in the formation of bottlenecks. If there is a noticeable imbalance between the CTs of Op1 and Op2, the operator with the lower CT may end up performing non-production tasks when starved or blocked. Since the buffer capacity between Op1 and Op2 in model *435 is limited to 1 unit, Op1 frequently has to pause until Op2 processes more units. This situation was observed during production, with Op1 sometimes experiencing extended idle periods or engaging in activities unrelated to production. As a result, Op1 can unintentionally become the bottleneck, not because of process time but because of waiting and operational dynamics.

These findings highlight that, for model *435, bottlenecks are influenced not only by process times but also by operational variability, process design, and operator behavior.

B. *Suggestions for Improvements*

The previous analysis highlighted a range of factors contributing to production inefficiencies on L35, such as workload imbalances between operators, process variability, and the negative impact of non-standardized activities and buffer management. These insights emphasize the importance of adopting a comprehensive approach to process improvement, one that not only identifies the bottleneck but also aims to stabilize and enhance overall line performance.

The proposed improvements for the assembly line, include: redesigning the line layout and rebalancing operator tasks; introducing a revised operating standard for the MER workstation; implementing and regulating buffers between operators; and standardizing the allocation of stochastic activities.

The first improvement involves changing the line layout. This adjustment is intended to improve line balancing by redistributing activities among operators, thereby increasing throughput and minimizing points of starvation and blocking. Since model *435 is part of the full product family, which represents most of the planned production volume, the new layout is designed with these models in mind. The spare parts family, which accounts for a smaller share of production, is not prioritized for layout adaptation. The revised layout and operator flow for the full product family, including model *435, are shown in Fig. 4.

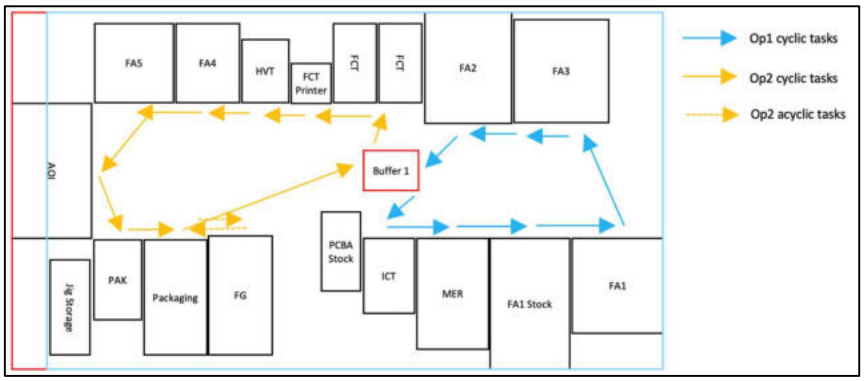


Fig. 4. Layout suggestion; Initial layout limits in blue, suggested layout limits in red

As previously discussed, this new layout simplifies process dependencies (Fig. 5), reducing process variability and creating a more stable, predictable production flow.

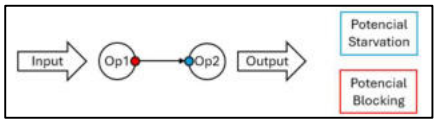


Fig. 5. Improved process dependencies between operators

With the new allocation of activities, the process flow for model *435 is further improved. To enhance throughput, a revised operating standard for the MER workstation was introduced. Under this new standard, Op1 will always leave one processed PCBA on the press. When reaching the step just before press activation, Op1 replaces the processed PCBA with the latest one from the ICT, activates the MER, and immediately proceeds to the next station. This approach eliminates the need for Op1 to wait for the press cycle to finish, effectively reducing non-value-added time and decreasing Op1's CT by 4,50s, the duration of the press operation.

The new line balancing for model *435, resulting from the updated layout and process adjustments, was estimated using Maynard's Operation Sequence Technique (MOST). MOST was chosen over the traditional Methods-Time Measurement (MTM) method because it is more straightforward to apply and does not require specialized MTM training (Viharos and Bán, 2020). Each operator's movements were measured individually and converted into standardized time values, after which the standard allowances were applied. The results are presented in Fig. 6.

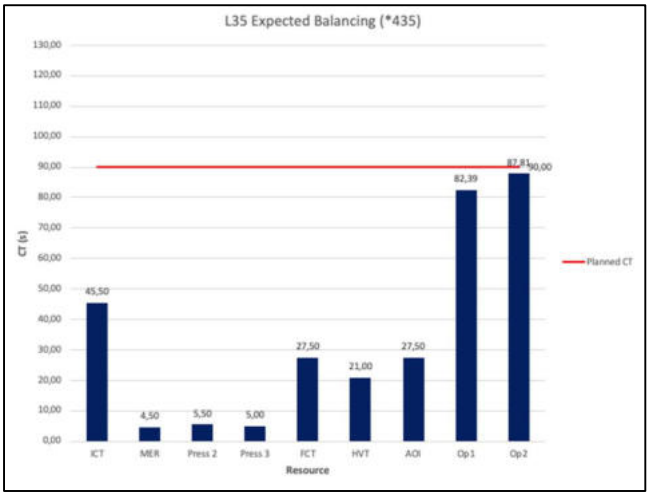


Fig. 6. Expected line balancing

Stochastic activities, such as delivering parts to repair operators and supplying the assembly line, are now systematically assigned to the operator with the lower CT. This approach ensures that the bottleneck resource remains focused on core production tasks, as any loss of productive time at the bottleneck directly affects the line's overall throughput.

The introduction of buffers between stations can significantly enhance both throughput and operational stability in assembly lines (Tiacci, 2024; Usubamatov et al., 2014). Buffers serve as a safeguard against the inherent variability and occasional disruptions that occur during production (Goldratt and Cox, 2016). Without adequate buffers, even minor delays at one station can quickly lead to upstream congestion and downstream starvation, ultimately reducing the efficiency of the entire process.

The buffer between Op1 and Op2 for model *435 was limited to just 1 unit. The frequent occurrences of starvation and blocking indicated that this buffer size was insufficient. To determine a more effective buffer size, it is recommended to base the calculation on the relationship between the bottleneck resource's hourly throughput and the average CT of the resources adjacent to the buffer. This ensures that the buffer is properly scaled to accommodate fluctuations and maintain a steady, continuous flow throughout the process.

$$Buffer\ i\ Ratio = \left| \frac{3600}{Max(x)} - \frac{3600}{\frac{x_i + x_{i+1}}{2}} \right| \tag{3}$$

Where x is the resource's CT; n is the number of resources in the process; and $i = 1,\ldots,n-1$.

It was important to determine an appropriate buffer size that suits the process's specific production characteristics. Instead of assigning a unique buffer to each model, which would result in unused buffers occupying valuable space, model *435 was grouped with other full-product models to form a product family. The buffer size for this group was then calculated by rounding up the family's average Buffer i Ratio, ensuring the buffer is both efficient and space-conscious.

$$Buffer\ Size = \left\lceil \frac{\sum Buffer\ i\ Ratio}{W} \right\rceil \tag{4}$$

Where W is the number of products in the family of models; n is the number of resources in the process; and $i = 1,\ldots,n-1$.

The calculated necessary buffer sizes were (Table 2):

TABLE II. CALCULATED BUFFER SIZES

Average Buffer i Ratio	Buffer Size
1,83	2

The combined impact of these improvements is a production system with reduced waiting and idle time, more consistent task execution, and a better-balanced workload across operators. These enhancements are mutually reinforcing, reducing non-value-added time and updating the line layout to improve both line balancing and process dependencies, while standardizing the allocation of stochastic activities further streamlines production flow. Together, these changes lead to higher throughput and greater predictability and stability in the assembly process.

V. EXPECTED RESULTS

Implementing these measures is expected to not only increase throughput but also stabilize process performance, reducing the likelihood of shifting bottlenecks and supporting more effective long-term planning and continuous improvement efforts.

With the new line balancing (Fig. 6), throughput for model *435 is projected to rise from 35 un/h to 40 un/h, corresponding to an improvement in the bottleneck CT from 98.44s to 87.81s. These changes are anticipated to make the process more stable

and predictable. As a result, the ITVM should consistently identify Op2 as the bottleneck for model *435, reflecting a more balanced and controlled production environment.

VI. Discussion

In the updated balancing charts (Fig. 6), the MER is considered an independent resource. In the previous configuration, its processing time was incorporated into the CT of Op1, as the operator remained idle while waiting for the completion of the MER operation. Following the proposed improvements, Op1 no longer waits for the MER, thereby establishing a clear operational decoupling between the two resources.

It is critical to distinguish between the empirically observed bottleneck diagnostics and the projected expected effects of the proposed interventions. Because the implementation phase coincided with a personnel change, specifically, the introduction of a new operator with a different skill level, the isolated impact of the proposed improvements could not be definitively observed in a controlled manner. Therefore, the analytical link between the collected evidence and the final process-improvement claims relies on theoretical projections using MOST, rather than post-implementation data. While the initial diagnostics strongly suggest that mitigating operational variability will stabilize the line, empirical validation is required before claiming these solutions definitively prevent bottleneck shifting.

Post-implementation observations indicate that Op1, responsible for the initial assembly stages of model *435, experienced an increase in physical workload. This effect is primarily due to the concentration of tasks that require an electric screwdriver. Future improvements should therefore prioritize ergonomic interventions, such as the integration of torque arms, to reduce physical strain and promote more stable and consistent throughput. Additionally, the consolidation of the FA1 and FA1 Stock workstations is recommended, as their integration may reduce operator movement, minimize material handling, and improve both efficiency and ergonomics.

Regarding the BDM, the ITVM presents limitations when applied over short time intervals. For instance, in the presence of temporary stoppages within a limited observation window, e.g., one hour, resources not responsible for the disruption may be incorrectly identified as bottlenecks due to reduced sample size and low variance. This suggests that variance-based metrics alone may be insufficient for short-term analysis, and that hybrid approaches combining multiple indicators could provide more reliable results for high-resolution monitoring. Notably, Betterton and Silver (2012) acknowledge that short observation periods, such as a single shift, may not be adequate for robust application of the method, thereby supporting the use of larger datasets in this study. This also aligns with West et al. (2022) observations, which state that the length of the selected variance interval has a significant impact on the quality of the detection. Furthermore, their findings indicate that immediate bottlenecks are subject to a slightly higher delay due to the average-based determination of the variance.

Furthermore, reliance solely on ITVM-based analysis is insufficient to fully explain the root causes of bottlenecks. Although the method provides valuable quantitative insights, it does not capture the full operational complexity of the production system. Consequently, ITVM results were complemented with direct observation and detailed workflow analysis. This combined quantitative–qualitative approach enabled the identification of additional influencing factors not evident from IT data alone, leading to a more comprehensive understanding of system constraints and more effective improvement actions. Additional recommendations include collaborating with the client to redesign product features, particularly in assembly operations involving threaded fasteners. The adoption of alternative fastening solutions, such as snap-fit mechanisms, can significantly reduce CTs at constraint operations and further enhance throughput.

Finally, a more robust, systematic data collection framework is recommended. Improved data availability and quality will support more accurate analyses, enable real-time process monitoring, and facilitate faster identification of inefficiencies, thereby strengthening continuous improvement efforts.

It is important to note that increases in manufacturing throughput do not necessarily translate directly into higher sales. However, even when production improvements do not fully resolve constraints at the value stream level, they can eliminate the manufacturing system as the primary bottleneck. This allows management to redirect focus toward other limiting factors within the broader system, thereby supporting overall performance improvement.

VII. Conclusion

This work presented a systematic analysis of the L35 assembly line, focusing on identifying process constraints and developing targeted improvement measures to enhance throughput and operational stability. The results demonstrate that applying a data-driven BDM, namely the ITVM, provides valuable quantitative support for identifying system constraints. However, the findings also highlight that such methods, when used in isolation, are insufficient to fully capture the complexity of real production environments. In particular, the study confirms that the ITVM's reliability is influenced by the size of the observation window, with reduced accuracy in short time frames due to a limited sample size and sensitivity to transient disturbances. Consequently, integrating quantitative analysis with direct observation and detailed workflow assessment proved essential for accurately diagnosing bottlenecks and understanding their underlying causes. This combined approach enabled a more comprehensive evaluation of the system and supported the definition of more effective and context-aware improvement actions. The proposed modifications to the production system, including layout reconfiguration and revised line balancing, are expected to improve throughput, reduce operator idle time, and enhance process decoupling between resources. Additionally, these changes enhance system robustness by mitigating the risk of bottleneck shifting and promoting a more stable, predictable production flow. Nevertheless, the evaluation of these improvements was limited to expected outcomes, as the implementation phase was affected by personnel changes, introducing variability that constrained the experimental validation of outcomes. From an operational perspective, the study also identified the importance of ergonomic considerations in sustaining performance improvements. Increased physical workload on specific operators, particularly due to repetitive tool usage, may negatively impact consistency and should be addressed through appropriate ergonomic

solutions. Furthermore, opportunities for process simplification and efficiency gains were identified, including workstation consolidation and potential product design improvements to reduce assembly time.

Beyond the specific confines of the L35 assembly line, this study generates broader implications for both practice and research. For practitioners, it demonstrates that in environments where operators retain a high degree of autonomy over stochastic tasks, bottlenecks are rarely static; they are dynamically driven by human behavioral variability rather than just CT imbalances. For researchers, this study highlights critical methodological implications regarding the empirical application of data-driven BDMs. First, it addresses the ITVM's sensitivity to data granularity and observation window size. In environments with limited DRDs, researchers must adapt ITVM to identify constraint zones rather than individual bottleneck resources. Furthermore, the study validates that applying variance-based metrics over short time intervals risks generating false positives, as temporary stoppages can be incorrectly identified as bottlenecks due to misleading low variance. Second, this case demonstrates that while ITVM is highly effective for constraint localization, it is fundamentally diagnostic rather than explanatory.

For future research, it is recommended to further develop bottleneck detection methodologies by incorporating hybrid performance metrics that combine variance-based indicators with additional parameters, enabling more reliable short-term analysis and real-time monitoring. Moreover, establishing more robust, systematic data collection systems is essential to support higher-resolution analysis and continuous improvement initiatives. Finally, it is important to recognize that improvements in manufacturing throughput do not inherently translate into increased market performance. However, by eliminating the production system as a primary constraint, organizations can shift their focus toward higher-level bottlenecks within the value stream, thereby enabling more effective global improvement.

ACKNOWLEDGMENTS

This work is supported by CIDMA (https://ror.org/05pm2mw36) under the Portuguese Foundation for Science and Technology (FCT, https://ror.org/00snfqn58), Grants UID/04106/2025 (https://doi.org/10.54499/UID/04106/2025) and UID/PRR/04106/2025 https://doi.org/10.54499/UID/PRR/04106/2025).

References

Abebe, H.T. (2022) "Determination of Sample Size and Errors," in O. O. Awe, E. A. Vance, and K. Love, Promoting Statistical Practice and Collaboration in Developing Countries. 1st ed. New York: Chapman and Hall/CRC, pp. 321–338. Available at: https://doi.org/10.1201/9781003261148-27.

Betterton, C.E. and Silver, S.J. (2012) "Detecting bottlenecks in serial production lines – a focus on interdeparture time variance," International Journal of Production Research, 50(15), pp. 4158–4174. Available at: https://doi.org/10.1080/00207543.2011.596847.

Brochado, A.F. et al. (2022) "A data-driven model with minimal information for bottleneck detection - application at Bosch thermotechnology," International Journal of Management Science and Engineering Management, 18(4), pp. 318–331. Available at: https://doi.org/10.1080/17509653.2022.2116121.

Chang, Q. et al. (2007) Short-Term System Bottleneck Analysis. TP07PUB3. Society of Manufacturing Engineers.

Goldratt, E. and Cox, J. (2016) The Goal: A Process of Ongoing Improvement. 3rd ed. Routledge.

Kuo, C.T., Lim, J.T. and Meerkov, S.M. (1996) "Bottleneck in serial production lines: A system-theoretical approach," 2, pp. 233–276.

Lai, X. et al. (2021) "Data-driven dynamic bottleneck detection in complex manufacturing systems," Journal of Manufacturing Systems, 60, pp. 662–675. Available at: https://doi.org/10.1016/j.jmsy.2021.07.016.

Li, L., Chang, Q. and Ni, J. (2009) "Data driven bottleneck detection of manufacturing systems," International Journal of Production Research, 47(18), pp. 5019–5036. Available at: https://doi.org/10.1080/00207540701881860.

Roser, C. et al. (2017) "Bottleneck Prediction Using the Active Period Method in Combination with Buffer Inventories," Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing. Edited by H. Lödding et al., 514, pp. 374–381. Available at: https://doi.org/10.1007/978-3-319-66926-7_43.

Roser, C., Lorentzen, K. and Deuse, J. (2015) "Reliable shop floor bottleneck detection for flow lines through process and inventory observations: the bottleneck walk," Logistics Research, 8(1), p. 7. Available at: https://doi.org/10.1007/s12159-015-0127-2.

Roser, C. and Nakano, M. (2015) "A Quantitative Comparison of Bottleneck Detection Methods in Manufacturing Systems with Particular Consideration for Shifting Bottlenecks," Advances in Production Management Systems: Innovative Production Management Towards Sustainable Growth. Edited by S. Umeda et al., 460, pp. 273–281. Available at: https://doi.org/10.1007/978-3-319-22759-7_32.

Shu-Yin Chiang, Chih-Tsung Kuo and Meerkov, S.M. (2000) "DT-bottlenecks in serial production lines: theory and application," IEEE Transactions on Robotics and Automation, 16(5), pp. 567–580. Available at: https://doi.org/10.1109/70.880806.

Tang, J. et al. (2024) "A Comprehensive Review of Theories, Methods, and Techniques for Bottleneck Identification and Management in Manufacturing Systems," Applied Sciences, 14(17), p. 7712. Available at: https://doi.org/10.3390/app14177712.

Tiacci, L. (2024) "Combining balancing, sequencing and buffer allocation decisions to improve the efficiency of mixed-model asynchronous assembly lines," Computers & Industrial Engineering, 194, p. 110357. Available at: https://doi.org/10.1016/j.cie.2024.110357.

Urban, W. and Rogowska, P. (2020) "Methodology for bottleneck identification in a production system when implementing TOC," Engineering Management in Production and Services, 12(2), pp. 74–82. Available at: https://doi.org/10.2478/emj-2020-0012.

Usubamatov, R. et al. (2014) "Analysis of buffered assembly line productivity," Assembly Automation, 34(1), pp. 34–40. Available at: https://doi.org/10.1108/AA-11-2012-086.

Viharos, Z.J. and Bán, B. (2020) "Comprehensive Comparison of MTM and BasicMOST, as the Most Widely Applied PMTS Analysis Methods." 17th IMEKO TC 10 and EUROLAB Virtual Conference.

West, N., Schwenken, J. and Deuse, J. (2022) "Comparative Study of Methods for the Real-Time Detection of Dynamic Bottlenecks in Serial Production Lines," Advances and Trends in Artificial Intelligence. Theory and Practices in Artificial Intelligence. Edited by H. Fujita et al., 13343, pp. 3–14. Available at: https://doi.org/10.1007/978-3-031-08530-7_1.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

A Decision Support Model for Problem-Solving Framework Selection in Industry 4.0 Environments

Sofia Mendes Pedro [1], Juliana Salvadorinho [2,3] and Ana Sofia Matos [2,3]

1 *Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal; sl.pedro@campus.fct.unl.pt*
2 *UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal; j.salvadorinho@fct.unl.pt; asvm@fct.unl.pt*
3 *Laboratório Associado de Sistemas Inteligentes, LASI, 4800-058 Guimarães, Portugal*

Abstract — The selection of problem-solving frameworks such as DMAIC, Lean, PDCA, A3, 8D, TRIZ, and DMADV remains largely experience-based, despite increasing complexity in Industry 4.0 environments. This study proposes a hybrid decision support model combining a three-level decision tree for constraint-based filtering with a TOPSIS multi-criteria ranking method. Based on a systematic literature review of 83 articles, the model incorporates key decision criteria, including data availability, organisational maturity, time horizon, problem scope, and digital maturity. The model filters infeasible frameworks and ranks suitable alternatives, generating interpretable recommendations. A Python-based prototype is being developed to demonstrate its practical applicability. The model provides a structured and replicable approach to framework selection, reducing subjectivity and supporting data-driven continuous improvement in modern industrial contexts.

Keywords – Decision Support Model, Problem-Solving frameworks, Industry 4.0, Continuous Improvement

I. Introduction and Literature Review

In an increasingly complex and digitalized industrial landscape, organizations face a critical and recurring challenge: selecting the most appropriate problem-solving framework to address operational, quality, and design-related issues. The proliferation of frameworks such as DMAIC (Define, Measure, Analyse, Improve, Control), Lean Manufacturing, PDCA (Plan, Do, Check, Act), A3, 8D (Eight Disciplines), TRIZ (Theory of Inventive Problem Solving), and DMADV (Define, Measure, Analyse, Design, Verify) has created a rich but fragmented toolbox. While each framework is effective under specific conditions, their improper selection can lead to inefficiencies, wasted resources, delayed problem resolution, and reduced customer satisfaction.

Despite the extensive application of these frameworks in the industry, the literature reveals a critical gap: the absence of a structured and systematic decision support mechanism to guide practitioners in selecting the most suitable framework based on problem characteristics and organizational context. Framework selection remains largely experience-based and subjective, particularly in organisations with lower continuous improvement maturity (Barsalou and Perkin, 2024). This issue is further enhanced by the increasing complexity of Lean 4.0 environments, where engineers must navigate both traditional frameworks and digitally enhanced solutions (Schumacher et al., 2024).

Recent studies emphasise this gap. Comparative analyses of Lean tools highlight the lack of clear selection criteria (Iuga and Rosca, 2017), while more recent research points to the growing difficulty of aligning problem-solving approaches with Industry 4.0 capabilities (Ganesan et al., 2024; Yang et al., 2025). Although hybrid approaches are increasingly adopted, their selection and integration remain unstructured and inconsistent.

To address this limitation, this research proposes the development of a structured decision support model (DSM) to guide the selection of problem-solving frameworks.

The central research question is:

"How should a decision support model be structured to enable organisations to select the most appropriate problem-solving framework, or combination of frameworks, tailored to the specific type of problem and its contextual conditions?"

This paper is structured as follows. Section II presents the goals and methodology, describing the systematic literature review protocol. Section III reports the results of the literature review, covering the most prevalent problem-solving frameworks, their characterization by inputs, outputs, and constraints, the levels of Industry 4.0 integration identified across the corpus, and a SWOT analysis of the frameworks. Section IV presents the proposed decision support model, including the three-level decision tree, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) multi-criteria ranking stage, the Python prototype under development, and the competitive advantages of the model for organizations, followed by limitations and future work.

A. Importance of Decision Support Models in Organizations within Industry 4.0 context

Decision Support Models (DSMs) play a critical role in structuring complex decision-making processes involving multiple criteria, uncertainty, and trade-offs. Rather than replacing human judgement, DSMs enhance it by providing transparency, consistency, and reproducibility.

The literature demonstrates their effectiveness across various industrial applications. For instance, DSMs have been used in supplier selection problems through multi-criteria optimisation approaches (Arunkumar et al., 2006), in Lean tool prioritisation using methods such as VIKOR (Variable Ideal Komprowise Ranking) (Anvari et al., 2014), and in cost estimation models for engineered products (Khan et al., 2023). Additionally, decision frameworks have been applied to equipment lifecycle management and end-of-life strategies (Bashkite et al., 2014; Karaulova and Bashkite, 2016).

Importantly, DSMs enable the integration of hard constraints (eliminating infeasible options) and soft criteria (ranking feasible alternatives), reflecting how decisions are made in real-world industrial contexts. Empirical studies also demonstrate their organisational impact: structured decision-making improves operational performance, enhances data-driven culture, and increases employee engagement in continuous improvement initiatives (Colombari and Neirotti, 2024; Naughton et al., 2024).

Despite these benefits, an important limitation remains. DSMs have not been systematically developed for selecting problem-solving frameworks. This represents a significant research and practical gap, particularly in environments where the choice of framework directly impacts operational performance.

This gap becomes even more critical in the context of Industry 4.0, which has fundamentally reshaped the context of problem-solving. Technologies such as Internet of Things (IoT) sensors, big data analytics, artificial intelligence (AI), and digital twins enable real-time monitoring, predictive analytics, and enhanced decision-making capabilities.

Recent studies illustrate how these technologies augment traditional frameworks. Digital twin-based frameworks enable predictive quality assurance and supply chain resilience (Amer et al., 2026), while data-driven systems enhance decision-making in Lean and smart manufacturing environments (Tripathi et al., 2022). Similarly, big data analytics has been shown to significantly extend Lean manufacturing capabilities by enabling real-time process optimization (Ganesan et al., 2024).

However, this transformation also introduces new challenges. The integration of digital technologies varies significantly across organizations and frameworks, ranging from basic digital support to fully autonomous systems (Schumacher et al., 2024). As a result, selecting an appropriate problem-solving framework now requires consideration not only of the problem itself but also of the organization's digital maturity.

B. Digitalisation of Lean and Six Sigma Tools

The digital transformation of problem-solving frameworks goes beyond embedding them in Industry 4.0 environments, it fundamentally changes the inputs each framework requires and the outputs it can deliver. This shift has direct implications for how a decision support model must be designed.

Within DMAIC and Six Sigma, data collection and statistical analysis have been augmented by Artificial Neural Networks (ANN) and machine learning, enabling the handling of complex, non-normal datasets that traditional DMAIC statistical tools cannot process (Sen, 2015; Fahey et al., 2020). (Kuthe and Tharakan, 2009) demonstrated that ANN-based approaches within DMADV outperform classical regression in nonlinear manufacturing processes, effectively expanding the analytical scope of these frameworks beyond their original design. In parallel, Vision Systems and automated inspection technologies now generate real-time defect data that feeds directly into the Measure and Control phases of DMAIC, compressing cycle times and reducing human measurement error (Palaniappan, 2024).

Within Lean, digitalization has transformed Value Stream Mapping (VSM) from a static, paper-based mapping tool into a dynamic, data-driven instrument. Digital VSM, supported by IoT sensors and MES (Manufacturing Execution Systems) integration, allows real-time value stream monitoring rather than point-in-time snapshots, significantly improving the accuracy of waste identification (Villalba-Diez et al., 2019; Yang et al., 2025). Digital Twin technology extends this further, enabling simulation of future-state VSM scenarios before physical implementation, replacing traditional Kaizen events with model-validated improvement proposals (Schumacher et al., 2024; Amer et al., 2026). Tools such as Kanban have equally evolved: digital Kanban boards integrated with Enterprise Resource Planning (ERP) and MES systems provide automated pull signals and real-time Work in Progress (WIP) visibility, removing the manual overhead that previously limited Lean to simpler production environments (Cantini et al., 2024).

For structured problem-solving tools A3, 8D, and PDCA, digitalization has primarily manifested through workflow automation and data integration platforms. (Naughton et al., 2024) documented a case where PDCA-based continuous improvement (CI) projects were managed through a Microsoft PowerApps digital stage-gate system, producing automated alerts, hybrid workforce tracking, and real-time progress dashboards. (Cantini et al., 2024) showed that the A3, when linked to digital data collection processes, improves the reliability of inputs feeding into analytics platforms, addressing a key weakness of traditional A3, which depended on manually gathered, often inconsistent data.

Critically, these developments mean that the digital maturity of an organization is no longer merely a contextual backdrop for framework selection, it actively determines which version of a framework an organization can deploy and what outputs it can realistically expect. An organization without IoT infrastructure cannot implement a digital VSM; an organization without

data science capabilities cannot leverage ANN-enhanced DMAIC. This is precisely why digital maturity is incorporated as an explicit criterion in the TOPSIS stage of the proposed model, alongside data availability, organizational maturity, time available, and problem scope.

II. Goals and Methodology

This study aims to develop a structured decision support model that guides practitioners in selecting the most appropriate problem-solving framework given the specific characteristics of a problem. The research follows a systematic literature review (SLR) to characterise existing frameworks and identify decision criteria.

The SLR was conducted in the Scopus database using two search strings applied in parallel. The primary search was *"problem solving" AND ("Lean" OR "Six Sigma") AND (method* OR framework) AND "manufacturing"* and it returned 299 records. A secondary search targeting decision support contexts was *"problem solving" AND ("decision support" OR "decision model") AND ("Lean" OR "Six Sigma")* and it returned 7 additional records. The combined corpus of 306 records was screened through four sequential stages:

- **Stage A** - Language screening (English and Portuguese): 305 records retained
- **Stage B** - Source type screening (journal articles, conference proceedings, book series): 279 records retained
- **Stage C** - Full-text accessibility screening: 106 records retained
- **Stage D** - Relevance screening (paper must present, apply, compare, or evaluate problem-solving models or decision support models in an organizational context): 83 records retained

These 83 articles formed the final corpus on which the present study focused.

III. Results

A. ***Most Widely Used Problem-Solving Frameworks***

The analysis of the reviewed literature reveals a clear dominance of structured, process-oriented problem-solving frameworks (Fig. 1), with DMAIC-based approaches emerging as the most widely applied framework across industrial contexts (Gentili et al., 2006; Sahay et al., 2011; Pramod et al., 2026). Their phased and data-driven logic, Define, Measure, Analyse, Improve, Control, makes them particularly effective for defect reduction, process optimisation, and quality improvement initiatives. This consistency across studies reinforces DMAIC's position as the benchmark framework for analytical problem-solving.

Lean-based approaches also demonstrate strong prevalence, particularly in studies focused on waste elimination, flow efficiency, and operational performance (Clark, 2016; Ishak et al., 2019; Rossini et al., 2021). Unlike DMAIC, Lean methodologies are applied both at operational and strategic levels, highlighting their flexibility and adaptability across different organizational contexts (Cvetkovic et al., 2017). Their integration into broader systems further reinforces their relevance in continuous improvement environments.

The PDCA cycle appears frequently as a complementary and iterative improvement framework (Arredondo-Soto et al., 2021). Its simplicity and cyclical nature make it particularly suitable for incremental improvements and continuous learning environments, especially where rapid experimentation is prioritised over statistical depth.

Innovation-driven frameworks, particularly TRIZ, plays a critical role in addressing complex engineering contradictions (Hsieh et al., 2015; Shao-hua et al., 2017). Unlike incremental improvement frameworks, TRIZ is primarily applied in situations requiring inventive or design-oriented solutions.

In addition, structured problem-solving frameworks such as 8D and A3 are widely used to support root cause analysis and standardized problem resolution (Kumar Phanden et al., 2022; Kurnia et al., 2023). These frameworks are especially relevant in manufacturing and regulated industries, where documentation, traceability, and team-based problem-solving are critical. Notably, 8D distinguishes itself through its emphasis on immediate containment actions, while A3 enhances communication and structured thinking.

Finally, design-focused frameworks such as DFSS (Design for Six Sigma)/ DMADV (Define, Measure, Analyse, Design, Verify) are increasingly adopted in product and process development contexts (Teli et al., 2024). These frameworks extend traditional Six Sigma approaches by embedding quality at the design stage, shifting the focus from reactive problem-solving to proactive defect prevention.

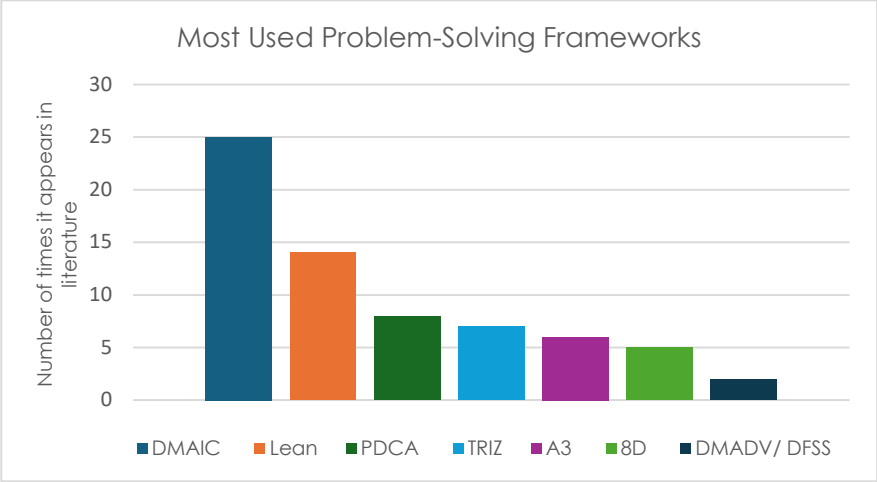


Fig. 1. Most Prevalent Frameworks in Literature

B. ***Framework Characterisation: Inputs, Outputs, and Constraints***

Table 1 summarizes the cross-article characterization of the seven problem-solving frameworks retained in the review, focusing on the dimensions most relevant for the decision support model: required inputs, expected outputs, and core application constraints. Rather than treating these frameworks as interchangeable improvement routines, the reviewed literature shows that each one is associated with a distinct problem context, evidence requirement, implementation logic, and organizational demand profile.

TABLE I. FRAMEWORK CHARACTERISATION: KEY INPUTS, OUTPUTS, AND CORE CONSTRAINTS

Table Head	Key required inputs	Primary Outputs	Core constraint
DMAIC	Historical data, Critical to Quality (CTQ) metrics, Measurement System Analysis (MSA), cross-functional team	Sigma level ↑, control plan, defect reduction	Process must exist; data required; 3–6 months; Green/Black Belt
Lean Mfg.	VSM, Gemba walk, cycle times, takt time, team buy-in	Lead time ↓, WIP ↓, future-state VSM	Repetitive process; management commitment; flow focus
PDCA	Improvement goal, baseline data, hypothesis, pilot resources	Validated solution, updated standard, next cycle input	Incremental problems; not for complex root causes
A3	Scoped problem, current-state data, Specific, Measurable, Achievable, Relevant, Time-bound (SMART) objectives, owner	A3 report, action plan, validated root cause	Limited scope; operational context; not for systemic issues
8D	Customer complaint/ Non-Conformance Report (NCR), defect data, team, containment feasibility	Containment (D3), permanent corrective action (D5–6), 8D report	Problem that already occurred; urgency required; automotive/regulated
TRIZ	Technical contradiction, system description, TRIZ facilitator	Inventive principles, resolved contradiction, innovative concept	Contradiction required; not for incremental fixes; high maturity
DMADV	Voice of the Customer (VOC), Quality Function Deployment *(QFD)*/ *House of Quality (HoQ)*, concept alternatives, benchmarks	Detailed design, Design Failure Mode and Effects Analysis (FMEA), customer sign-off	New product/process only; DFSS expertise; 3–9 months

C. ***SWOT Analysis***

While these frameworks remain central to structuring problem-solving activities, the findings highlight both their enduring relevance and their increasing limitations in data-intensive and dynamic environments. To synthesize these insights, a SWOT analysis (Fig. 2) was conducted.

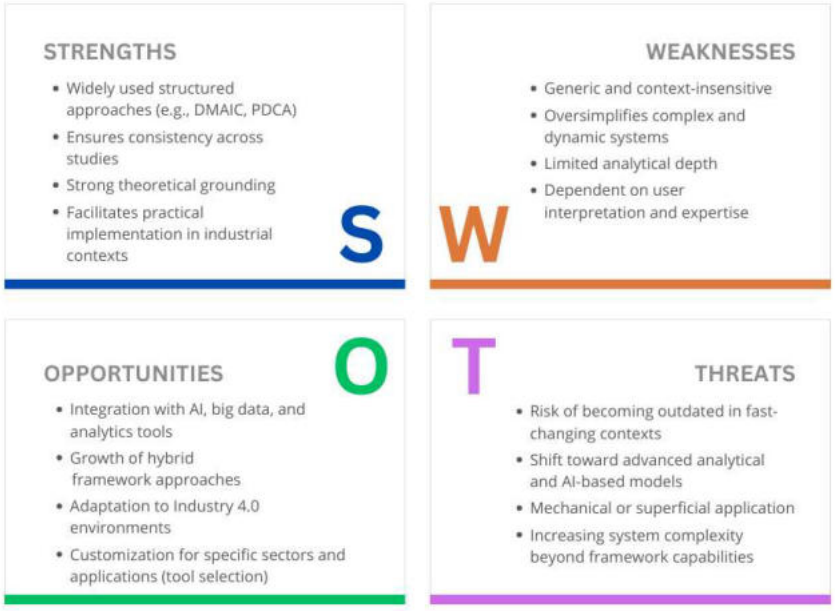


Fig. 2. SWOT Analysis of the literature

In terms of strengths, frameworks provide structured, systematic, and replicable approaches to process improvement, which facilitates their application across diverse industrial contexts (Rossini et al., 2021; Pramod et al., 2026). Their methodological clarity supports consistency and comparability, particularly in environments where standardization is critical. Widely adopted approaches such as DMAIC, PDCA, and Lean Six Sigma are grounded in robust theoretical foundations, reinforcing their credibility and long-term adoption (Arredondo-Soto et al., 2021; Byrne et al., 2021). Moreover, their ability to decompose complex problems into sequential and manageable steps enhances their practical usability and effectiveness in real-world implementations (Kurnia et al., 2023).

However, several weaknesses persist. Traditional frameworks are often characterized by a static and generic structure, limiting their adaptability to highly dynamic and context-specific environments (Iuga and Rosca, 2017). They also tend to simplify complex, multi-variable systems, which reduces their effectiveness in data-rich and uncertain conditions (Bertocci et al., 2021). Additionally, their application is frequently dependent on practitioner expertise, introducing subjectivity and variability in outcomes (Barsalou and Perkin, 2024). However, standalone frameworks often lack the analytical depth required to support advanced decision-making, particularly when compared to quantitative and data-driven models (Yang et al., 2025).

The identified opportunities reflect a clear shift toward integration and technological enhancement. The convergence of frameworks with decision-support models and digital technologies, such as artificial intelligence, machine learning, and simulation, enables the development of hybrid approaches that combine structural guidance with analytical capabilities (Fahey, Jeffers and Carroll, 2020). Furthermore, the incorporation of Industry 4.0 technologies, including Big Data and Digital Twins, enhances real-time monitoring, predictive analysis, and process optimization (Schumacher et al., 2024; Amer et al., 2026). This evolution not only increases the relevance of frameworks but also enables their adaptation to complex and highly digitalized industrial environments (Ganesan et al., 2024).

Finally, several threats may affect the continued relevance of traditional frameworks. The rapid advancement of digital technologies and AI-driven models introduces the risk of obsolescence for methodologies that remain purely conceptual (Sen, 2015; Yang et al., 2025). In parallel, the literature highlights the risk of rigid or mechanical application, where frameworks are used as prescriptive templates rather than adaptive tools (Clark, 2016). This can lead to superficial implementations and reduced effectiveness, particularly in complex operational settings.

D. ***The Proposed Decision Support Model***

The proposed model addresses the research question through a two-stage hybrid architecture. Stage 1 includes a three-level decision tree (Fig. 3) that performs hard constraint filtering, eliminating frameworks structurally incompatible with the problem's fundamental characteristics. Stage 2 consists of a TOPSIS approach that scores and ranks surviving candidates against four weighted criteria, producing a ranked recommendation with a numerical confidence score.

The selection of TOPSIS for Stage 2 is grounded in the SLR and builds explicitly on two precedents in the adjacent literature: Anvari et al. (2014) applied a modified VIKOR, a closely related MCDM (Multiple-Criteria Decision Making) method to Lean tool selection, demonstrating that structured multi-criteria methods outperform expert opinion alone. Alsyouf et al. (2011) combined QFD, VSM, and BSC (Balanced Scorecard) for Lean tool cost-effectiveness evaluation using comparable scoring logic. The present study's contribution lies not in the methods themselves, which are well-established, but in their novel integration and contextual application to the specific problem of problem-solving framework selection, a domain where no prior structured decision support model has been proposed.

D1. Stage 1: Three-Level Decision Tree

The three questions of the decision tree were selected for maximum discriminant power across the seven target frameworks:

Q1 - Does a process or product already exist? This question achieves the highest information gain of any possible first split, perfectly separating two fundamentally different domains: improvement of existing processes (DMAIC, Lean, PDCA, A3, 8D) from new-creation contexts (TRIZ, DMADV). No other question achieves this clean separation. If No → go to Q2b (design path). If Yes → go to Q2a (improvement path).

Q2a - Has the problem already occurred? / Q2b - Is there a technical contradiction? For existing processes (Q2a), the reactive/proactive distinction is critical because 8D includes an immediate (D3) containment step with no structural equivalent in proactive frameworks, a hard incompatibility that must be resolved before any quantitative scoring. For new designs (Q2b), the presence or absence of a technical contradiction precisely identifies TRIZ (contradiction present → TRIZ) versus DMADV (no contradiction → DMADV), closing the design path immediately.

Q3a - Is there urgency of containment? / Q3b — What is the primary focus? For reactive problems with confirmed urgency, 8D exits the tree directly without entering TOPSIS, no other framework provides (D3) containment, and this is a hard constraint, not a degree of preference. For proactive problems, Q3b separates quality/variability-focused problems (DMAIC, A3 as candidates) from flow/waste-focused problems (Lean, PDCA as candidates), informing TOPSIS weights without premature elimination.

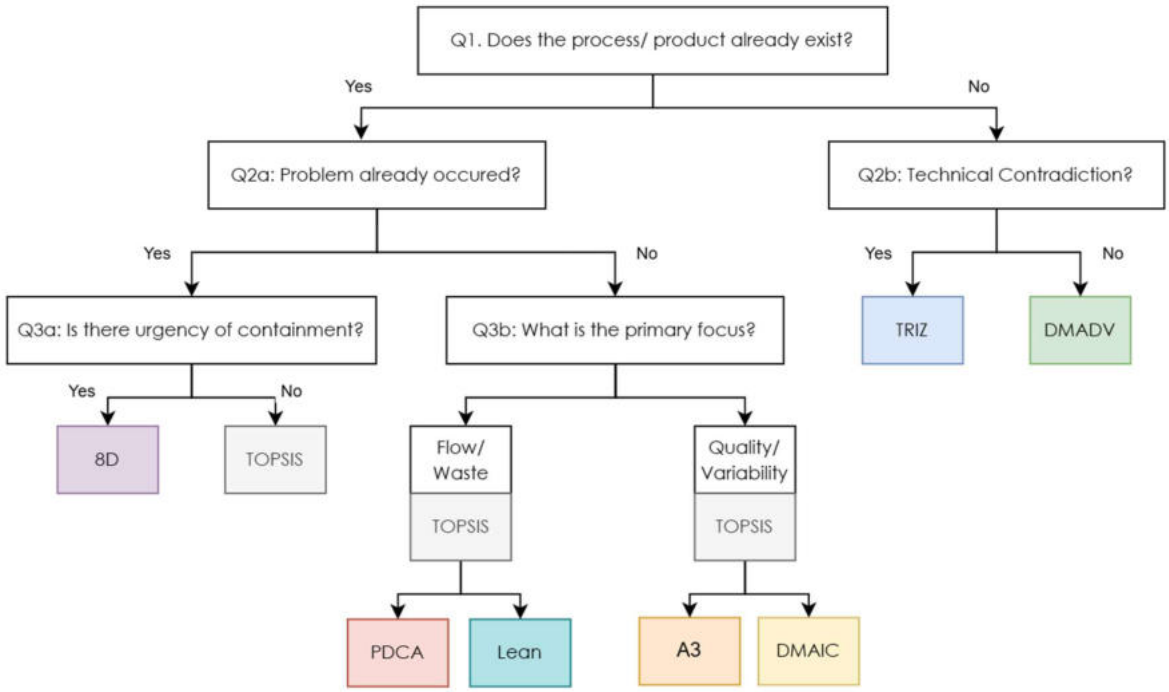


Fig. 3. Three-level decision tree

D2. Stage 2 - TOPSIS Multi-Criteria Ranking

After the decision tree removes the frameworks that are structurally incompatible with the problem, the remaining candidates are compared using TOPSIS. In simple terms, TOPSIS helps answer a practical question: which of the remaining frameworks fits this problem best overall?

The method works by comparing each candidate against an ideal profile and a worst-case profile. The ideal profile represents the framework that best matches the problem across all criteria, while the worst-case profile represents the least suitable one. The framework closest to the ideal and farthest from the worst-case receives the highest score, called the closeness coefficient *Ci*.

In this model, TOPSIS is used because the final choice is rarely based on only one factor. Even after the decision tree, several frameworks may still be feasible, but they may differ in data requirements, organizational maturity, time constraints, digital readiness, and problem scope. TOPSIS combines these factors into a single ranking, making the final recommendation more transparent and consistent than expert judgment alone.

The criteria were selected because they capture the main practical conditions that influence framework suitability (Table II): Data availability shows whether the organization has enough information to support data-driven methods such as DMAIC. Organizational maturity reflects whether the team has the skills and structural capacity needed for more advanced frameworks. Time available distinguishes fast-response methods from longer improvement cycles. Digital maturity captures whether the organization can support data-driven and Industry 4.0-aligned approaches. Problem scope helps separate localized problems from broader process-wide issues.

At this stage, no predefined weights are assigned to the criteria. Instead, all criteria are initially considered with equal importance to avoid introducing subjective bias into the model design. The relative importance of each criterion will be evaluated and validated in a later stage of research, using empirical data and/or expert input to ensure that the weighting reflects real-world decision-making conditions.

TABLE II. TOPSIS CRITERIA AND RATIONALE

Criterion	Rationale
Data Availability	Highest discriminant power: data-intensive frameworks such as DMAIC require measurable data, while PDCA and Lean can operate qualitatively
Organization Maturity	Advanced frameworks (e.g., TRIZ, DMAIC) require specialized competences unavailable in low-maturity organizations
Time Available	DMAIC requires 3–6 months; PDCA can produce results in days. A structural difference that cannot be traded off
Digital Maturity	Aligns framework selection with Industry 4.0 capabilities, enabling compatibility with data-driven and digitally integrated environments
Problem Scope	Distinguishes A3 (localized, single-step) from DMAIC and Lean (systemic, process-wide); least discriminant of the four

Each framework is scored against these criteria, and the weighted scores produce a final ranking. The top-ranked framework is recommended as the best fit. If the first two frameworks score very closely, the model can suggest a hybrid approach, since the problem may be suitable for more than one method.

IV. DISCUSSION AND FINAL REMARKS

The results of the systematic literature review provide a consistent and structured understanding of the current landscape of problem-solving frameworks and their application in industrial contexts. The analysis confirms the dominance of process-oriented and data-driven approaches such as DMAIC and Lean, while also highlighting the complementary role of iterative frameworks like PDCA and structured methodologies such as A3 and 8D. At the same time, the findings reveal significant variability in how these frameworks are selected and applied, with a strong dependence on practitioner experience and limited use of formal decision-support mechanisms. Furthermore, the review emphasizes the increasing influence of Industry 4.0 technologies, which are reshaping both the capabilities and requirements of traditional frameworks. These insights collectively validate the need for a structured, criteria-based support model capable of guiding framework selection in a consistent, transparent, and context-aware manner. Building on these findings, the following section discusses the practical implications, contributions, and limitations of the proposed model.

Deploying this decision support model is designed to generate advantages across three performance dimensions. In operational terms, the model is designed to reduce the time and cost associated with framework-selection mismatches, especially in cases where the wrong framework is applied, forcing abandonment, restart, or completion of a suboptimal improvement cycle. (Barsalou and Perkin, 2024) documented significant resource waste arising from problem-solving team selection errors in global manufacturing, and the proposed model wants to address this failure mode by making selection structured and evidence based. By replacing practitioner intuition with consistent decision logic, the model also is designed to reduce variability across continuous improvement projects and mitigates one of the main sources of inconsistency in framework deployment.

From a capability-development perspective, the model functions not only as a decision aid but also as a learning tool. Because the questions are explicit, the criteria are visible, and the reasoning is auditable, less experienced practitioners can gradually develop a clearer intuition about framework selection, building tacit knowledge that more experienced continuous improvement professionals usually acquire over years of practice. (Naughton et al., 2024) showed that structured continuous improvement guidance tools increase employee participation precisely because they reduce the perceived complexity of framework selection and make improvement more accessible to a wider group of practitioners.

In strategic terms, the model also supports organizational positioning under Industry 4.0. Its TOPSIS criteria can be extended to include digital readiness as organizations evolve toward Lean 4.0, and this flexibility is embedded in the Python architecture. The model provides the routing logic needed to direct organizations toward advanced hybrid approaches at the appropriate

stage of digital transformation, rather than prematurely applying high-maturity frameworks in low-maturity contexts or, conversely, using incremental PDCA when a problem requires the full statistical power of DMAIC. (Ganesan et al., 2024) showed that combining Lean with Big Data produces qualitatively different, not merely quantitatively better, improvement outcomes, and the proposed decision model offers the selection logic needed to reach that level when the organization is ready.

This study has some limitations. First, the proposed model is grounded in a systematic literature review, so its structure reflects the evidence available in the selected corpus rather than direct validation in multiple industrial settings. Second, the model was designed for seven specific problem-solving frameworks, which means that its decision logic and criteria may need adaptation if additional frameworks, sectors, or organizational conditions are introduced.

Future work will focus on validating the model with SMEs through a structured questionnaire to verify whether the inputs, outputs, and decision logic of each step are appropriate and reflect real practitioner decisions. This validation will also include a real-world example in the pharmaceutical industry to test the model's practical applicability in a regulated manufacturing context with specific quality and compliance requirements.

Based on the feedback from SMEs and results from the pharmaceutical case study, the model can then be refined and further tested. This iterative process will make the prototype more robust and better aligned with organizational reality, while supporting the possible inclusion of new criteria such as digital readiness or sustainability, depending on the application context.

V. Acknowledgment

The authors acknowledge Fundação para a Ciência e a Tecnologia I.P. for its financial support via the project UID/00667: Unidade de Investigação e Desenvolvimento em Engenharia Mecânica e Industrial.

References

Alsyouf, I. et al. (2011) "A framework for assessing the cost effectiveness of lean tools," European J. of Industrial Engineering, 5(2), p. 170. Available at: https://doi.org/10.1504/EJIE.2011.039871.

Amer, Y., Soufali, A. and Zaghwan, A. (2026) "A digital twin-based framework for predictive quality assurance and supply chain resilience in the automotive industry," Advanced Engineering Informatics, 69, p. 103969. Available at: https://doi.org/10.1016/j.aei.2025.103969.

Anvari, A., Zulkifli, N. and Arghish, O. (2014) "Application of a modified VIKOR method for decision-making problems in lean tool selection," The International Journal of Advanced Manufacturing Technology, 71(5–8), pp. 829–841. Available at: https://doi.org/10.1007/s00170-013-5520-x.

Arredondo-Soto, K.C. et al. (2021) "A Plan-Do-Check-Act Based Process Improvement Intervention for Quality Improvement," IEEE Access, 9, pp. 132779–132790. Available at: https://doi.org/10.1109/ACCESS.2021.3112948.

Arunkumar, N. et al. (2006) "Linear approach for solving a piecewise linear vendor selection problem of quantity discounts using lexicographic method," The International Journal of Advanced Manufacturing Technology, 28(11–12), pp. 1254–1260. Available at: https://doi.org/10.1007/s00170-004-2474-z.

Barsalou, M. and Perkin, R. (2024) "Statistical problem-solving teams: A case study in a global manufacturing organization in the automotive industry," Quality and Reliability Engineering International, 40(1), pp. 513–523. Available at: https://doi.org/10.1002/qre.3404.

Bashkite, V., Karaulova, T. and Starodubtseva, O. (2014) "Framework for Innovation-oriented Product End-of-life Strategies Development," Procedia Engineering, 69, pp. 526–535. Available at: https://doi.org/10.1016/j.proeng.2014.03.022.

Bertocci, F. et al. (2021) "A Guideline for Implementing a Robust Optimization of a Complex Multi-Stage Manufacturing Process," Applied Sciences, 11(4), p. 1418. Available at: https://doi.org/10.3390/app11041418.

Byrne, B., McDermott, O. and Noonan, J. (2021) "Applying Lean Six Sigma Methodology to a Pharmaceutical Manufacturing Facility: A Case Study," Processes, 9(3), p. 550. Available at: https://doi.org/10.3390/pr9030550.

Cantini, A., Costa, F. and Portioli-Staudacher, A. (2024) "Driving corporate digitization with reliable data through the A3 approach: an Italian case study," International Journal of Lean Six Sigma, 15(8), pp. 143–170. Available at: https://doi.org/10.1108/IJLSS-03-2024-0055.

Clark, D. (2016) "Quality improvement in basic histotechnology: the lean approach," Virchows Archiv, 468(1), pp. 5–17. Available at: https://doi.org/10.1007/s00428-015-1838-0.

Colombari, R. and Neirotti, P. (2024) "Leveraging Frontline Employees' Knowledge for Operational Data-Driven Decision-Making: A Multilevel Perspective," IEEE Transactions on Engineering Management, 71, pp. 13840–13851. Available at: https://doi.org/10.1109/TEM.2023.3291272.

Converso, G. et al. (2025) "Lean Service Waste Classification and Methodological Application in a Case Study," Journal of Manufacturing and Materials Processing, 9(4), p. 121. Available at: https://doi.org/10.3390/jmmp9040121.

Cvetkovic, N. et al. (2017) "Enhancing the Agility and Performances of a Project with Lean Manufacturing Practices," in B. Katalinic (ed.) DAAAM Proceedings. 1st ed. DAAAM International Vienna, pp. 0661–0670. Available at: https://doi.org/10.2507/28th.daaam.proceedings.093.

Fahey, W., Jeffers, P. and Carroll, P. (2020) "A business analytics approach to augment six sigma problem solving: A biopharmaceutical manufacturing case study," Computers in Industry, 116, p. 103153. Available at: https://doi.org/10.1016/j.compind.2019.103153.

Ganesan, A. et al. (2024) "Integration of Big Data and Business Analytics in Lean Manufacturing: A Strategic Approach," 2024 4th International Conference on Computer, Communication, Control & Information Technology (C3IT). 2024 4th International Conference on Computer, Communication, Control & Information Technology (C3IT), Hooghly, India: IEEE, pp. 1–6. Available at: https://doi.org/10.1109/C3IT60531.2024.10829419.

Gentili, E., Aggogeri, F. and Mazzola, M. (2006) "The Improvement of a Manufacturing Stream Using the DMAIC Method," Manufacturing Engineering and Textile Engineering. ASME 2006 International Mechanical Engineering Congress and Exposition, Chicago, Illinois, USA: ASMEDC, pp. 127–133. Available at: https://doi.org/10.1115/IMECE2006-14469.

Hsieh, H.-N., Chen, J.-F. and Do, Q. (2015) "Applying TRIZ and Fuzzy AHP Based on Lean Production to Develop an Innovative Design of a New Shape for Machine Tools," Information, 6(1), pp. 89–110. Available at: https://doi.org/10.3390/info6010089.

Ishak, A., Ginting, R. and Chandra, V. (2019) "The application of lean manufacturing to minimize waste in Crude Palm Oil (CPO) production process at PT. XYZ," IOP Conference Series: Materials Science and Engineering, 505(1), p. 012143. Available at: https://doi.org/10.1088/1757-899X/505/1/012143.

Iuga, M.V. and Rosca, L.I. (2017) "Comparison of problem solving tools in lean organizations," MATEC Web of Conferences. Edited by I. Bondrea, C. Simion, and M. Ință, 121, p. 02004. Available at: https://doi.org/10.1051/matecconf/201712102004.

Karaulova, T. and Bashkite, V. (2016) "Decision-making Framework for Used Industrial Equipment," Engineering Economics, 27(1), pp. 23–31. Available at: https://doi.org/10.5755/j01.ee.27.1.8618.

Khan, S.S. et al. (2023) "A Lean Knowledge-Based Decision Support System for Cost Estimation in the Foundry Industry," Brazilian Archives of Biology and Technology, 66, p. e20220658. Available at: https://doi.org/10.1590/1678-4324-2023220658.

Kumar Phanden, R. et al. (2022) "8Ds method of problem solving within automotive industry: Tools used and comparison with DMAIC," Materials Today: Proceedings, 65, pp. 3266–3272. Available at: https://doi.org/10.1016/j.matpr.2022.05.383.

Kurnia, H. et al. (2023) "Combination of lean thinking and A3 problem-solving methods to reduce the cost of purchasing cleaning agents in a paint manufacturer in Indonesia," SINERGI, 28(1), p. 103. Available at: https://doi.org/10.22441/sinergi.2024.1.011.

Kuthe, A.M. and Tharakan, B.D. (2009) "Application of ANN in Six Sigma DMADV and its comparison with regression analysis in view of a case study in a leading steel industry," International Journal of Six Sigma and Competitive Advantage, 5(1), p. 59. Available at: https://doi.org/10.1504/IJSSCA.2009.024215.

Naughton, E. et al. (2024) "A structured model for continuous improvement methodology deployment and sustainment: A case study," Heliyon, 10(21), p. e40034. Available at: https://doi.org/10.1016/j.heliyon.2024.e40034.

Palaniappan, M. (2024) "Illuminating Precision: Enhancing Clip Attach Vision Accuracy in Semiconductor Manufacturing," 2024 IEEE 40th International Electronics Manufacturing Technology (IEMT). 2024 IEEE 40th International Electronics Manufacturing Technology (IEMT), Penang, Malaysia: IEEE, pp. 1–5. Available at: https://doi.org/10.1109/IEMT61324.2024.10875203.

Powell, D. and Oliveira, M. (2021) "Insights from a Digital Lean Startup: Co-creating Digital Tools for Cognitive Augmentation of the Worker," Procedia CIRP, 104, pp. 1384–1388. Available at: https://doi.org/10.1016/j.procir.2021.11.233.

Pramod, V.G. et al. (2026) "Reducing rework in plastic extrusion process in pipe manufacturing using DMAIC," The International Journal of Advanced Manufacturing Technology, 142(7–8), pp. 4053–4063. Available at: https://doi.org/10.1007/s00170-025-17078-2.

Rossini, M., Kassem, B. and Portioli-Staudacher, A. (2021) "Lean Warehousing: Enhancing Productivity Through Lean," in D.J. Powell et al. (eds.) Learning in the Digital Era. Cham: Springer International Publishing (IFIP Advances in Information and Communication Technology), pp. 339–347. Available at: https://doi.org/10.1007/978-3-030-92934-3_35.

Sahay, C., Ghosh, S. and Bheemarthi, P.K. (2011) "Process Improvement of Brake Lever Production Using DMAIC (+)," Volume 3: Design and Manufacturing. ASME 2011 International Mechanical Engineering Congress and Exposition, Denver, Colorado, USA: ASMEDC, pp. 801–826. Available at: https://doi.org/10.1115/IMECE2011-63813.

Schumacher, S. et al. (2024) "Process Model For The Problem-Based Identification Of Solutions In Lean Production Systems 4.0." Available at: https://doi.org/10.15488/17750.

Sen, P. (2015) "Application of ANN in Six Sigma for CO modelling and energy efficiency of blast furnace: a case study of an Indian pig iron manufacturing organisation," International Journal of Six Sigma and Competitive Advantage, 9(2/3/4), p. 109. Available at: https://doi.org/10.1504/IJSSCA.2015.074957.

Shao-hua, C. et al. (2017) "Based on the Theory of TRIZ Solving the Problem of 18650 Battery Electrolyte Filling," IOP Conference Series: Earth and Environmental Science, 104, p. 012018. Available at: https://doi.org/10.1088/1755-1315/104/1/012018.

Teli, S., Surange, V. and Gaikwad, L. (2024) "Failure Analysis of Automotive Drive System: A Six Sigma DMADV Approach," Journal of Failure Analysis and Prevention, 24(5), pp. 2292–2315. Available at: https://doi.org/10.1007/s11668-024-02003-8.

Tripathi, V. et al. (2022) "Development of a Data-Driven Decision-Making System Using Lean and Smart Manufacturing Concept in Industry 4.0: A Case Study," Mathematical Problems in Engineering. Edited by K.-H. Chang, 2022, pp. 1–20. Available at: https://doi.org/10.1155/2022/3012215.

Villalba-Diez, J. et al. (2019) "Characterization of Industry 4.0 Lean Management Problem-Solving Behavioral Patterns Using EEG Sensors and Deep Learning," Sensors, 19(13), p. 2841. Available at: https://doi.org/10.3390/s19132841.

Yang, Q., Wang, X. and Wu, H. (2025) "Study on lean production management of new energy vehicle body painting based on the dual perspectives of digital transformation and VSM," PLOS ONE. Edited by M.R.N. Qureshi, 20(2), p. e0318253. Available at: https://doi.org/10.1371/journal.pone.0318253.

Yang, X. et al. (2025) "Recent Advances in Lean Techniques for Discrete Manufacturing Companies: A Comprehensive Review," Machines, 13(4), p. 280. Available at: https://doi.org/10.3390/machines13040280.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Edge Vision for Monitoring Manual Assembly Motions: Robust Calibration and Sensitivity Analysis

Sandro Dobovičeк [1], Elvis Krulčić [1] and Duško Pavletić [1]

[1] University of Rijeka (UNIRI) – Faculty of Engineering (Croatia); sandro.dobovicek@riteh.uniri.hr

Monitoring manual assembly in Quality 4.0 environments remains challenging, particularly when sensing solutions must be unobtrusive, repeatable, and robust to variations in operator behaviour. This paper considers an edge-based monocular pose estimation pipeline as a component of a measurement system for detecting quality-relevant manual assembly events, rather than as a black-box classifier. A standardised calibration protocol involving five operators generates known target counts for six pick bins and one deposit bin, enabling a calibration procedure that selects spatial regions of interest (ROIs) and temporal persistence parameters from recorded wrist trajectories. Parameter values are determined using plateau analysis to identify robust operating regions, and a controlled gap-bridging mechanism is introduced to mitigate the impact of brief detection dropouts. The results demonstrate zero event-counting error at the calibrated operating point on the protocol-conforming calibration data, with stable counting for the pick bins across a broad parameter range and a narrower deposit-bin plateau caused by short dwell times and incidental flyovers. ROI perturbation sensitivity maps further quantify how placement errors affect performance and identify the most sensitive boundaries. The proposed method provides a transparent calibration approach for repeatable in-process event monitoring, with additional potential for cycle time and ergonomics analysis.

Keywords – Quality 4.0, manual assembly monitoring, human pose estimation, edge vision

I. Introduction

Despite increasing automation, many manufacturing environments still rely on manual and mixed human-machine assembly, particularly in high-mix, low-volume production where flexibility and rapid changeovers are essential. In these settings, quality losses often arise from deviations in standardised work execution, such as skipped steps, unintended part selections, rework, and micro events like interruptions or hesitations. Consequently, there is growing interest in in-process quality assurance methods that monitor assembly execution and provide objective evidence of step conformity without adding intrusive instrumentation or disrupting operator ergonomics. This trend aligns with the Quality 4.0 perspective, which integrates quality management with real-time data streams, analytics, and feedback to support proactive, risk-oriented decision making (Mahin et al., 2024; Godina and Matias, 2019). In this paper, the authors adopt this perspective by treating operator motion as a process-relevant signal to support event-level monitoring during manual tasks.

A practical way to obtain such motion signals is through human pose estimation, which predicts anatomical keypoints (e.g., wrists, elbows, eyes) from video frames and outputs a structured representation as a skeleton-based model (Chen et al., 2022). Modern deep learning methods enable multi-person, real-time pose extraction in practical camera setups (Cao et al., 2021), and pose is widely used as an intermediate representation for downstream tasks such as tracking and action recognition (Gao et al., 2025, Liu et al., 2025, Menanno et al., 2024). This structured output provides quantitative information about human motion and supports mapping motion to process-relevant regions to derive event indicators for process monitoring.

However, pose outputs are continuous and noisy, while assembly verification typically requires discrete, auditable events (e.g., "step started" or "step completed"). This translation requires both spatial mapping through regions of interest (ROIs) and temporal decision logic that is robust to brief misdetections and tracking errors. In this work, the pose estimation model is therefore treated as one component of a measurement system: it produces wrist-coordinate observations, while calibrated ROI and temporal decision rules convert those observations into attribute-like event counts that can serve as process-quality evidence. This interpretation makes calibration and robustness checks essential.

Building on this perspective, the study presents a calibration method for a spatiotemporal event detector based on pose estimation outputs. The method selects spatial parameters, specifically region boundaries, and temporal parameters, including persistence and short-gap bridging, while emphasising robustness through plateau-based parameter selection across repeated calibration runs. The approach is demonstrated using a monocular RGB vision setup for manual assembly monitoring, but the underlying measurement-system perspective applies to other pose-based sensing implementations.

II. Related Work

Existing vision-based approaches to manual assembly monitoring rely on recurring design choices: defining process-relevant spatial context, stabilising predictions over time, and handling sensing imperfections such as occlusions and brief misdetections. The following review summarizes these approaches and highlights why parameter selection and calibration remain critical for reliable deployment.

A common approach operationalises spatial context through region (or zone) based interaction logic, mapping tracked hands or wrists to predefined workspace regions (such as parts or assembly areas) and combining this with temporal decision rules (such as confidence gating and state tracking) to robustly infer task progression. In AR-supported assembly, Zogopoulos et al. combine visual tracking with 2D image-based state tracking to provide step-level guidance, illustrating how region and state-based logic can drive instruction flow (Zogopoulos et al., 2021). Similarly, Mokuwe et al. present a low-cost, vision-based QA setup that uses hand tracking and rule-based logic to evaluate whether an assembly follows the expected steps or order, reinforcing the role of explicit decision logic in translating hand motion into step verification (Mokuwe et al., 2022).

Complementary 2D approaches infer progress from hand-part relations derived through object or hand detection rather than predefined zones. Andrianakos et al. present machine learning-based monitoring of human assembly operations using a webcam in 2D space, which can inform the operator that they are about to perform an incorrect task before it is completed (Andrianakos et al., 2019). Recent work by Ay and Emel proposes real-time task validation by combining deep object detection with hand-joint trajectory classification, using an estimated hand-skeleton representation from 2D video as a motion descriptor for fine-grained task recognition (Ay and Emel, 2025).

Beyond rule-based triggering, many approaches treat progress monitoring as a temporal inference problem, using sequential models or workflow constraints to enforce plausible step transitions. Berg et al. present an action recognition approach for assembly tasks using Hidden Markov Models (HMMs), leveraging temporal structure for consistent interpretations of task progression with a Microsoft Kinect® V2 RGB-D-based setup (Berg et al., 2018). Oyekan et al. also used a Kinect® setup with an HMM-inspired state-based workflow representation to track manual assembly progress from computer vision and RGB-D observations (Oyekan et al., 2020).

Another set of studies uses depth/RGB-D or multi-sensor setups to infer progress from hand-object-workspace interactions. Kaczmarek et al. demonstrate progress monitoring in manual assembly with 3D-image sensors and gesture-based interaction, showing how spatial relations in the workspace can be converted into monitoring signals (Keczmarek et al., 2015). Faccio et al. propose real-time operator assistance where a depth camera captures motion relative to the workstation and provides visual feedback during sequential tasks, illustrating an assistance architecture that relies on robust perception of operator-environment relations (Faccio et al., 2019).

Complementary approaches derive progress signals by modelling task progress through object interactions and verifying process states through object features such as positions, relative positions, and movements. Mühlbauer et al. propose a configurable progress-monitoring approach based on camera recordings and object-interaction features (Mühlbauer et al., 2022). Related work also learns assembly operations directly from visual sensing and classifies them using machine learning, providing a higher-level alternative to region-touch event detection (Rückert et al., 2020). Additionally, reviews of machine-vision-based prediction and monitoring of assembly and disassembly in make-to-order and SME contexts emphasizes that real-world performance is strongly affected by visibility and sensing setup. Princz et al. summarise recurring issues such as hand occlusions, sensitivity to camera configuration and placement, and the need to test under variable conditions, such as lighting and setup differences (Princz et al., 2023).

Taken together, the related work demonstrates that vision-based manual assembly monitoring can be implemented using region-based interaction logic, temporal models, and learned recognition approaches. However, when the required output is a count of predefined workstation interaction events, a key unresolved issue is how to configure and verify the decision layer that links noisy visual observations to discrete event counts for robustness. Existing studies offer limited guidance on selecting spatial ROI boundaries and temporal persistence thresholds, determining the operating margin these settings provide, or reporting sensitivity to parameter perturbations. This work addresses this gap by introducing a transparent calibration and sensitivity analysis procedure for pose-derived event counting under a standardised protocol. Its contribution is not a new pose estimator or a general action-recognition model, but a reproducible method for converting continuous wrist-coordinate observations into repeatable and auditable bin-interaction counts for manual assembly monitoring.

III. System Description and Calibration Method

A. Experimental setup

A simple process monitoring system was implemented at a manual assembly workstation. At this station, the operator assembles a product from components separated into six input bins. Two bins on the operator's left contain material supplied from the preceding station. The remaining four input bins in front of the operator contain parts supplied by an external supplier. After completing the assembly process, the operator deposits the finished assembly into a designated "End" bin on the right.

The sensing approach was selected based on the defined measurand. The objective is to count bin-interaction events within predefined workstation regions, rather than estimate metric three-dimensional hand trajectories. The event detector therefore operates on image-plane wrist keypoints using calibrated spatial and temporal decision parameters, for which a compact monocular RGB setup is sufficient. RGB-D sensing may be advantageous when depth-based separation of interactions or explicit three-dimensional motion information is required. In addition, the selected monocular edge-vision architecture offers a small physical footprint, flexible mounting, and straightforward replication across workstations.

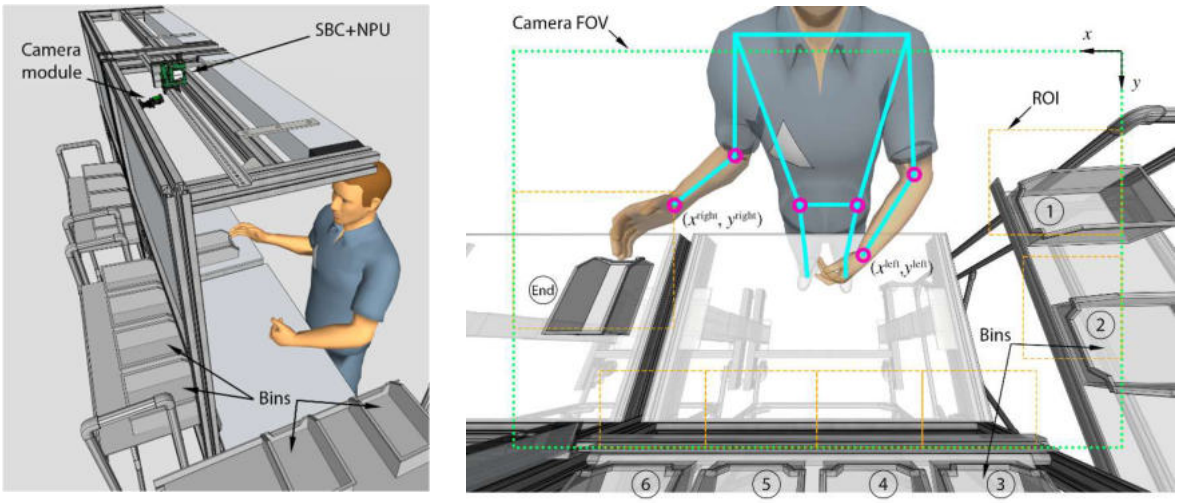
The monitoring system includes an edge computing unit mounted on a DIN rail attached to the station frame structure and a monocular camera mounted on a separate adjustable bracket (Fig. 1.a). The edge unit consists of a single-board computer (SBC, Raspberry Pi 5 with 8 GB RAM) paired with a neural processing unit (NPU) based on the

Hailo-8 accelerator. The accelerator delivers up to 26 tera operations per second (TOPS) with low power consumption, enabling real-time processing at approximately 30 frames per second (FPS). The selected monocular camera is the Camera Module 3 Wide, which features a 12-megapixel Sony IMX708 sensor, powered autofocus (non-NoIR variant), and a 120° diagonal field of view (FOV). The video stream is processed locally using a YOLO (You Only Look Once)-based pose estimation model (YOLO-v8m) to obtain wrist keypoint locations in image coordinates. These detections are mapped to predefined regions of interest (ROIs) on the image, corresponding to a workspace zone relevant to the monitored assembly task (Fig. 1.b). The resulting event stream is logged for calibration locally on the SBC, consuming approximately 9 W of power.

B. Calibration protocol and data collection

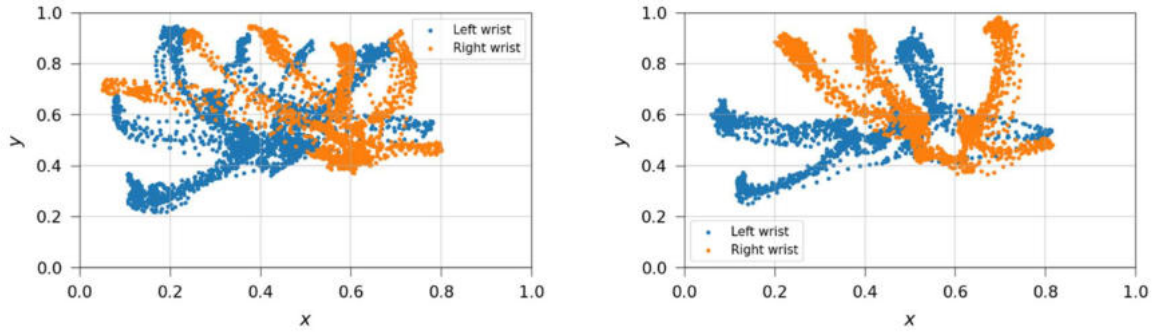
Calibration data should be collected using a standardised protocol designed to provide a known target. The chosen protocol target is touch count per bin, T_b. Each calibration run follows a fixed bin-interaction sequence: the operator picks T_b parts from Bin1, then T_b parts from Bin2, continuing sequentially through Bin6. After each pick, the part is placed at a designated location on the workstation table. Finally, the operator deposits T_b parts into the End bin. To account for kinematics-driven variability, calibration should be repeated under different postures (e.g. standing and seated), which alter wrist trajectories and occlusion patterns, thereby providing a more realistic calibration set.

For each calibration run, the system logs one record per frame containing a timestamp, frame index, and normalised keypoint coordinates for both wrists, $(x, y) \in [0,1]^2$ stored in .csv file. Fig. 2 shows wrist-keypoint trajectories from two representative calibration runs, first operator (a) and second operator (b). These illustrate operator-dependent hand use strategies and motion patterns, despite an identical prescribed sequence.



(a) (b)

Fig. 1. Experimental setup and vision pipeline. (a) Manual assembly workstation with material storage bins, overhead camera module, and edge-computing unit (SBC+NPU). (b) Processed video frame illustrating the camera field of view, lens distortion, regions of interest (ROIs), and wrist keypoints estimated by the pose estimation model, which provides normalised image coordinates.



(a) (b)

Fig. 2. Example wrist trajectories from two representative test runs. Scatter plots show normalised image coordinates of left and right wrist keypoints for two representative runs: (a) operator 1; (b) operator 2. The plots illustrate operator-dependent hand use strategies and motion patterns, despite an identical prescribed sequence.

C. Mathematical model

A touch event is defined as the wrist being located within a bin region of interest (ROI) and remaining within that ROI for at least a persistence threshold of m consecutive frames. Calibration aims to select an ROI configuration $\mathcal{R}$ and

persistence threshold m so that detected touch counts match a known calibration protocol across multiple runs. To increase robustness against brief tracking dropouts, the model optionally bridges short "no-bin" gaps of up to g frames between identical bin labels. Calibration assumes that the protocol target T_b is valid ground truth. Accordingly, runs in which the executed bin-interaction counts visibly deviate from the prescribed protocol are excluded from parameter fitting because T_b no longer represents their actual reference count. Such runs may be retained separately as checks of detector behaviour under atypical executions.

Within the proposed measurement-system perspective, the measurand is the count of predefined bin-interaction events occurring during the monitored assembly process. The pose estimator provides the observable signal as time-ordered wrist-keypoint coordinates in the image plane. ROI assignment, together with temporal persistence, constitutes the measurement rule by which these observations are converted into event counts. The standardised protocol defines the reference counts, while calibration identifies spatial and temporal parameter combinations for which the measured counts remain consistent with those references across repeated runs. The resulting OK plateau is therefore interpreted as the operating margin of the measurement system: a wider plateau indicates that correct event counting is preserved over a broader range of parameter values, reflecting lower sensitivity to threshold changes and greater practical robustness.

1) Input sets and indices

Let $\mathcal{S}$ be the set of calibration runs, indexed by r. Run r contains frames $t \in \{1, \dots, N_r\}$. Let $\mathcal{W} = \{\text{left}, \text{right}\}$ denote the wrist side, indexed by w. Let $\mathcal{B}$ denote the set of bins (regions), indexed by b, and let $\mathcal{B}^* \subseteq \mathcal{B}$ denote the target bins used for scoring. Let $\mathcal{M}$ denote the finite set of candidate persistence thresholds m evaluated during calibration. For each run r, frame t, and wrist w, the pose estimator provides normalised wrist coordinates $(x_{r,t}^w, y_{r,t}^w) \in [0,1]^2$.

2) ROI definition and candidate generation

Each bin $b \in \mathcal{B}$ has a rectangular ROI defined as

$$\text{ROI}_b = \left[x_b^{\min}, x_b^{\max}\right] \times \left[y_b^{\min}, y_b^{\max}\right]. \tag{1}$$

An ROI set is $\mathcal{R} = \{\text{ROI}_b\}_{b \in \mathcal{B}}$. Calibration selects $\mathcal{R}$ from a finite candidate set $\mathcal{C} = \{\mathcal{R}_c\}$, where c indexes candidate ROI sets generated by applying structured boundary perturbations of a base ROI set $\mathcal{R}_0$ (expansions, shrinkages and per-edge moves of ROIs, with coordinates clipped to [0,1]). $\mathcal{R}_0$ reflects the nominal bin locations under a fixed camera setup and serves as a physically meaningful starting point.

3) Frame-wise spatial assignment

An inside-ROI indicator is defined for bin b as

$$\text{I}_b(x, y) = 1\left(x_b^{min} \le x \le x_b^{max} \wedge y_b^{min} \le y \le y_b^{max}\right). \tag{2}$$

Because ROIs may be close or slightly overlapping, frame-wise labels use an ordered "first match-wins" rule over a predefined bin order $\langle b_1, \dots, b_{|\mathcal{B}|} \rangle$. The symbol 0 denotes the no-bin label, assigned when the wrist keypoints lies outside all ROIs. The frame-wise bin label is therefore defined as

$$\ell_{r,t}^w(\mathcal{R}) = \begin{cases} b_j & \text{for the smallest } j \text{ such that } I_{b_j}\left(x_{r,t}^w, y_{r,t}^w\right) = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

4) Temporal smoothing by gap bridging

Let $g \in \mathbb{Z}_{\ge 0}$ denote the maximum bridged gap length between identical bin labels, introduced to suppress short label interruptions caused by measurement noise near ROI boundaries. Starting from the frame-wise sequence $\ell_{r,t}^w(\mathcal{R})$, the bridged gap sequence $\tilde{\ell}_{r,t}^w(\mathcal{R}, g)$ is obtained as follows. If

$$\ell_{r,t}^w = b, \quad \ell_{r,t+1}^w = \cdots = \ell_{r,t+k}^w = 0, \quad \ell_{r,t+k+1}^w = b, \tag{4}$$

with $k \le g$, then

$$\tilde{\ell}_{r,t+1}^w = \cdots = \tilde{\ell}_{r,t+k}^w = b \tag{5}$$

Only gaps of the no-bin label 0 are bridged; excursions into different bin labels are not bridged.

5) Touch-event definition and counting

For each run r and wrist w, a touch event for bin b is defined as a maximal contiguous segment of equal non-empty labels in $\tilde{\ell}_{r,t}^w(\mathcal{R}, g)$. A segment (s, e, b) with start frame s, end frame e, and label $b \ne 0$, is valid if $(e - s + 1) \ge m$. Let $\hat{T}_{r,b}(\mathcal{R}, m, g)$ denote the number of valid touch-event segments assigned to bin b for run r and wrist w. The total detected number of touches for run r and bin b is then aggregated over both wrists as

$$\hat{T}_{r,b}(\mathcal{R}, m, g) = \sum_{w \in \mathcal{W}} \hat{T}_{r,b}^w(\mathcal{R}, m, g). \tag{6}$$

D. Calibration metrics

The model defines the total absolute touch-count error as

$$E(\mathcal{R}, m, g) = \sum_{r \in \mathcal{S}} \sum_{b \in \mathcal{B}^*} \left|\hat{T}_{r,b}(\mathcal{R}, m, g) - T_b\right|, \tag{7}$$

which measures the magnitude of deviations. It then defines the failed pair count as

$$F(\mathcal{R}, m, g) = \sum_{r \in \mathcal{S}} \sum_{b \in \mathcal{B}^*} 1\left(\hat{T}_{r,b}(\mathcal{R}, m, g) \ne T_b\right), \tag{8}$$

which measures the frequency of deviations regardless of size. The pass rate is given by

$$P(\mathcal{R}, m, g) = 1 - \frac{F(\mathcal{R},m,g)}{|\mathcal{S}||\mathcal{B}^*|}. \tag{9}$$

Perfect agreement is indicated by

$$OK(\mathcal{R}, m, g) \Leftrightarrow 1(F(\mathcal{R}, m, g) = 0). \tag{10}$$

For fixed $(\mathcal{R}, g)$ the OK threshold set is defined as the subset of candidate thresholds that produce perfect agreement:

$$\mathcal{M}_{OK}(\mathcal{R}, g) = \{m \in \mathcal{M} : OK(\mathcal{R}, m, g) = 1\}, \tag{11}$$

with plateau width $W(\mathcal{R}, g) = |\mathcal{M}_{OK}(\mathcal{R}, g)|$.

E. Calibration algorithm

The calibration procedure uses the recorded wrist trajectories from the protocol runs, together with a base ROI set, and evaluates a finite set of ROI candidates and persistence thresholds to identify a robust operating point. The base ROI set $\mathcal{R}_0$ is defined from the fixed-camera workstation view by manually placing rectangular ROIs over the nominal image plane locations of the six input bins and the End bin. Specifically, the procedure uses the set of calibration runs $\mathcal{S}$, represented by recorded wrist trajectories stored as .csv files, together with $\mathcal{R}_0$, the bin order, target bins $\mathcal{B}^*$, protocol target T_b, candidate thresholds $\mathcal{M}$, gap parameter g, and the ROI perturbation scheme.

Because rectangular ROI boundaries form a continuous design space, the objective is not to identify a globally optimal ROI geometry across all possible rectangle configurations. Instead, the procedure calibrates the detector within a finite, physically meaningful perturbation neighbourhood of the manually defined base configuration $\mathcal{R}_0$. Among the evaluated candidates, preference is given to ROI sets that preserve correct event counting over the broadest range of persistence thresholds. The algorithm steps are:

1) Generate ROI candidates

Construct $\mathcal{C} = \{\mathcal{R}_c\}$ by applying structured perturbations to $\mathcal{R}_0$ (expansions, shrinkages and per-edge moves of ROIs, with coordinates clipped to $[0,1]$).

2) Evaluate candidates across thresholds

For each candidate $\mathcal{R}_c \in \mathcal{C}$ and each $m \in \mathcal{M}$:

a) For each run $r \in \mathcal{S}$ and wrist $w \in \mathcal{W}$, assign labels $\ell_{r,t}^{w}(\mathcal{R}_c)$ using first match-wins;

b) Apply gap bridging up to g to obtain $\tilde{\ell}_{r,t}^{w}(\mathcal{R}_c, g)$;

c) Compute detected touch counts $\hat{T}_{r,b}(\mathcal{R}_c, m, g)$ for all $r \in \mathcal{S}$ and $b \in \mathcal{B}^*$;

d) Compute calibration metrics $E(\mathcal{R}_c, m, g)$, $F(\mathcal{R}_c, m, g)$, pass rate $P(\mathcal{R}_c, m, g)$, and the perfect-agreement indicator $OK(\mathcal{R}_c, m, g)$.

3) Compute OK plateaus

For each $\mathcal{R}_c$, compute the OK threshold set $\mathcal{M}_{OK}(\mathcal{R}_c, g)$ and its plateau width $W(\mathcal{R}_c, g)$.

4) Select the best ROI candidate

Choose $\mathcal{R}^* \in \mathcal{C}$ that maximizes $W(\mathcal{R}_c, g)$. If ties occur, break ties by selecting the candidate with smallest best achievable error, $\min_{m \in \mathcal{M}} E(\mathcal{R}_c, m, g)$.

5) Select the final threshold

If $W(\mathcal{R}^*, g) > 0$, set m^* to the midpoint of the OK plateau $\mathcal{M}_{OK}(\mathcal{R}^*, g)$. Otherwise, set $m^* = \arg\min_{m \in \mathcal{M}} E(\mathcal{R}^*, m, g)$. The final calibrated ROI set $\mathcal{R}^*$ is stored as the calibrated ROI configuration.

In the present study, calibration uses a single-round selection from the finite candidate set $\mathcal{C}(\mathcal{R}_0)$. The procedure can optionally be refined by re-centring candidate generation at the selected solution (set $\mathcal{R}_0 \leftarrow \mathcal{R}^*$) and repeating the evaluation within a narrower perturbation neighbourhood. Such refinement is considered worthwhile only if it results in a wider feasible range of m (a wider OK plateau).

Taken together, these steps define a calibration procedure that links the pose-derived coordinate signal to the targeted measurand (event counting) through explicitly selected spatial and temporal parameters. The objective is not merely to identify a parameter setting that reproduces the reference counts once, but to select a configuration that remains correct across repeated protocol runs and maintains a measurable operating margin. In this sense, the calibrated detector is treated as a configured measurement procedure whose counting behaviour can be inspected, challenged, and recalibrated when deployment conditions change.

Consistent with this measurement system interpretation, the event decision layer remains transparent and rule based. A neural network classifier or sequence model could be trained for wrist trajectory-based action recognition, but this would require a larger labelled dataset of correct and incorrect executions, would be less directly auditable, and could overfit to operator-specific or station-specific motion styles. In the proposed approach, machine learning is applied upstream in the pose estimator, which provides coordinate observations, while the conversion of these observations into event counts remains deterministic, inspectable, and recalibratable. This design supports the

objective of establishing traceable parameter settings and operating margins, rather than maximising black-box classification performance on a limited data set.

IV. Results and Analysis

Five operators participated in the study. For each operator, three protocol-conforming calibration runs were retained for parameter fitting and for the counting-performance analysis reported below, yielding a total of 15 runs. These runs were collected at different times during the operators' shifts to capture within-shift variability. During each retained run, the standardised protocol required $T_b = 10$ touches for $b \in \{\text{Bin1}, \ldots, \text{Bin6}\}$ and $T_{End} = 10$ for the End bin. Calibration runs included both standing and seated postures and were recorded using the same workstation layout and camera placement. Frame-wise wrist coordinates were logged at an approximately stable frame rate.

Table I summarises the basic run statistics for these 15 protocol-conforming calibrations runs, including the number of runs, average FPS, and average run duration. The system maintained a stable frame rate, while average run durations varied moderately among the five operators. This variation reflects natural differences in working pace and exposes the algorithm to realistic temporal patterns and the distinct motion styles of individual operators. These protocol-conforming runs were used both to determine the calibrated ROI and temporal parameter settings and to characterise the resulting event-counting behaviour reported in this section.

TABLE I. Overview of Protocol-Conforming Calibration Run Parameters

Operator	Calibration run parameters		
	No. of protocol-conforming runs	*Average FPS*	*Average run time (mm:ss)*
1	3	30.33	03:14
2	3	30.31	03:05
3	3	30.42	03:36
4	3	30.32	02:31
5	3	30.34	02:51

No separate protocol-conforming validation dataset was collected for formal out-of-sample performance estimation. During data collection, two additional runs were identified as protocol-deviating because the operators visibly missed one prescribed pick from a bin, meaning that the nominal target count $T_{End} = 10$ did not represent the actual executed count. These two runs were not included in Table I, were set aside before calibration, and were not used for parameter fitting. After calibration, they were processed separately as checks of atypical behaviour. In both cases, the operators picked nine rather than ten parts from one bin, and the calibrated detector likewise reported nine corresponding bin-touch events.

Across the runs, interaction patterns were observed for all seven bins, and it was concluded that interactions with the End bin tended to have shorter dwell times. This effect was particularly pronounced for some operators, who deposited finished parts with faster movements and minimal hand presence above the End bin. In addition, Bin4 contained small parts that were occasionally difficult to separate, which sometimes resulted in brief corrections, such as lifting two parts and returning one. These within-protocol corrections were retained to ensure that calibration remains robust under realistic manual handling.

For the calibration algorithm, the gap-bridging length was set to $g = 5$ frames (approximately 0.17 s at 30 FPS) to suppress brief "label 0" interruptions caused by intermittent wrist keypoint dropouts, while keeping the smoothing time window short to avoid masking genuine ROI exits. Calibration begins from an initial ROI configuration $\mathcal{R}_0$ obtained by manually placing rectangular regions around the workstation zones in normalised image coordinates. To identify a robust ROI configuration, the candidate set $\mathcal{C} = \{\mathcal{R}_c\}$ is generated by applying structured boundary perturbations to $\mathcal{R}_0$. Each candidate set was produced by expanding or shrinking an ROI and/or shifting individual edges by predefined step sizes of 0.01, 0.02 and 0.05, with all coordinates clipped to $[0, 1]$.

The resulting calibration-selected ROI configuration $\mathcal{R}^*$ is shown in Fig. 3, overlaid with aggregated wrist trajectories from all calibration runs. Fig. 3 also provides a spatial interpretation of the Bin4 handling behaviour. Compared with the neighbouring top-row bins, the Bin4 ROI extends to a slightly lower y range, consistent with brief upward and downward wrist excursions during part separation and verification.

To examine the distinct temporal behaviour of the two event classes, the calibration metrics were evaluated separately for the pick-bin set $\mathcal{B}^*_{\text{pick}} = \{\text{Bin1}, \ldots, \text{Bin6}\}$ and for the deposit bin $\mathcal{B}^*_{\text{deposit}} = \{\text{End}\}$. Using the calibrated ROI set $\mathcal{R}^*$ and the gap-bridging parameter, the total absolute touch-count error E was evaluated over candidate persistence thresholds $m \in \mathcal{M}$. For the pick bins, the overall calibration curve $E(\mathcal{R}^*, m, g)$ shows a contiguous OK plateau, indicating that perfect agreement with the protocol targets is achieved for a range of persistence thresholds m (Fig. 4.a). This plateau supports robust parameter selection in the middle of the plateau, with $m^*_{\text{pick}} = 11$.

In contrast, the End bin exhibits a narrower OK region. The narrower OK plateau of the deposit bin is due to the ROI of the End bin differing from pick bins both temporally and spatially. Temporally, deposit actions typically have shorter dwell times: some operators place the finished part with a fast motion and minimal hovering, resulting in fewer consecutive frames inside the End region. Consequently, increasing the persistence threshold m beyond a moderate

range can cause missed End events, contributing to the rise in error on the right-hand side of the calibration curve (Fig. 4.b).

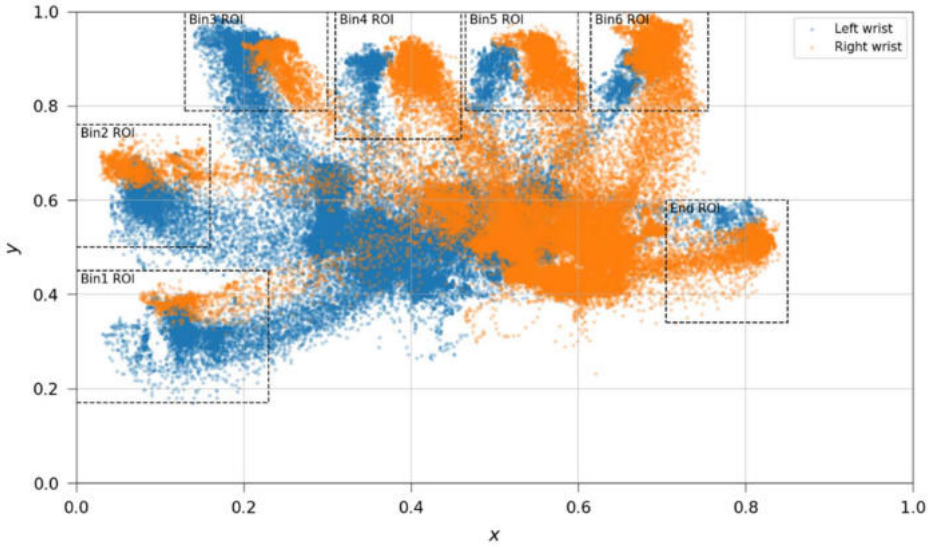
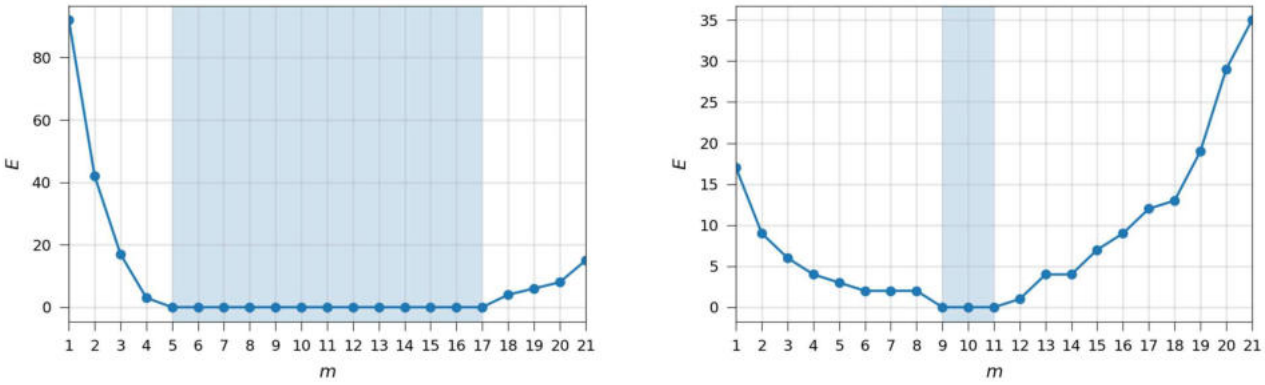


Fig. 3. Aggregated calibration trajectories and calibrated ROI configuration set. Normalized wrist keypoints from all runs are shown together with the calibrated ROIs for the six pick bins and the End bin, illustrating typical motion corridors and spatial separation between workstation zones.

Spatially, the deposit region (End bin) is located adjacent to a high-traffic corridor between the work area and Bin6. The aggregated trajectory plot with calibrated ROI shows that wrist paths frequently pass near or through the End region during transfers to or from Bin6. For smaller values of m, these brief incidental entries can satisfy the persistence rule and be counted as spurious End touches, which explains the elevated error on the left-hand side of $E(\mathcal{R}^*, m, g)$. Together, these effects compress the range of m values that both suppress flyover-induced false touch counts and still capture short deposit dwell times, resulting in a narrow End bin OK plateau. The OK plateau of the deposit bin thus allows selection of a parameter $m^*_{\text{deposit}} = 10$.



(a) (b)

Fig. 4. Calibration error vs. persistence threshold. Total absolute error E as a function of a persistence threshold m for (a) the pick region touches (Bin1-Bin6) and (b) the deposit region (End bin). The shaded zone indicates the OK plateau where perfect agreement with protocol targets is achieved

Table II summarises the aggregate counting performance at representative temporal thresholds m, including the number of reference actions, detected actions, and the corresponding Type I and Type II errors. In the event-counting interpretation used here, Type I errors denote excess detected actions relative to the protocol-defined reference count (overcounting), whereas Type II errors denote missing detected actions relative to that reference count (undercounting).

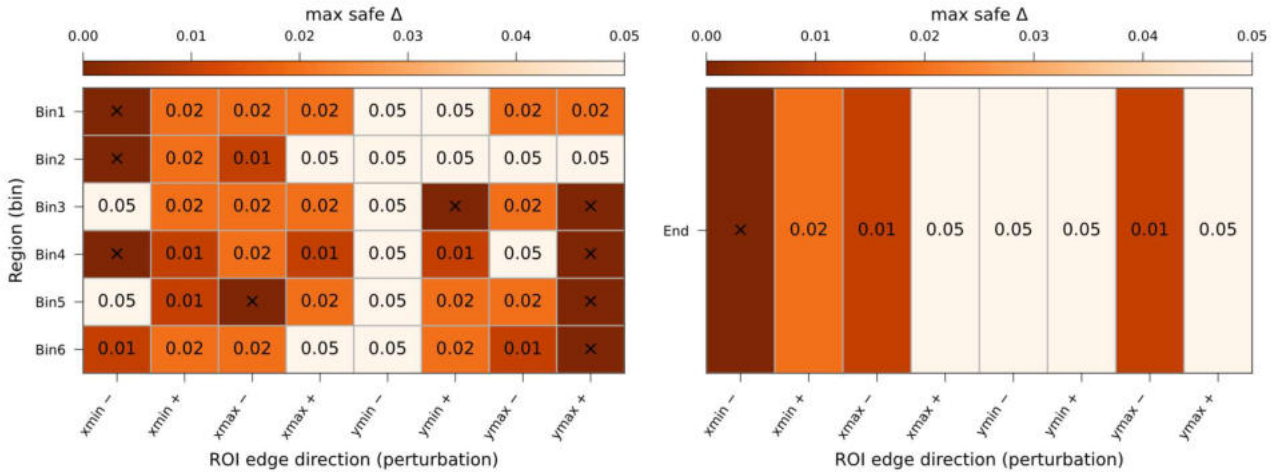
To quantify the sensitivity of the calibrated solution to ROI boundary placement, a structured perturbation analysis was conducted by shifting individual ROI edges and re-evaluating whether a robust operating region in m is maintained. Fig. 5 summarises the maximum safe ROI-edge change Δ that still preserves an acceptable OK plateau of required width $W \geq W_{\text{min}}$. In the present analysis, $W_{\text{min,pick}} = 3$ and $W_{\text{min,deposit}} = 2$. Smaller values indicate fragile edges where only minor boundary changes are tolerated. Cells marked "x" indicate that either the smallest tested perturbation fails to preserve the required plateau width or the corresponding edge cannot be perturbed further because it is already at a clipping limit of 0 or 1. The sensitivity map shows that robustness varies across bins and edges.

Several edges tolerate perturbations up to the maximum tested range, while certain edges exhibit near-zero tolerance, consistent with high wrist-traffic corridors and close spacing between neighbouring zones. These results provide actionable guidance for deployment: (i) fragile edges should be placed conservatively and verified first after any camera or bin repositioning, and (ii) sensitivity patterns can indicate where tighter ROI separation is needed to avoid systematic mislabelling.

TABLE II. AGGREGATE REFERENCE COUNTS, DETECTED COUNTS, AND TYPE I/TYPE II ERRORS AT REPRESENTATIVE TEMPORAL THRESHOLDS FOR PICK-BINS AND DEPOSIT-BIN EVENTS

Pick bins (Bin1 – Bin6)						*Deposit bin (End bin)*					
m	***Ref.***	***Detected***	***Type I***	***Type II***	***Error pattern***	***m***	***Ref.***	***Detected***	***Type I***	***Type II***	***Error pattern***
1	900	992	92	0	Overcount, +10.2%	1	150	167	17	0	Overcount, +11.3%
3	900	917	17	0	Overcount, +1.9%	5	150	153	3	0	Overcount, +2.0%
4	900	903	3	0	Overcount, +0.3%	8	150	152	2	0	Overcount, +1.3%
5-17	900	900	0	0	Zero error	9-11	150	150	0	0	Zero error
18	900	896	0	4	Undercount, -0.4%	12	150	149	0	1	Undercount, -0.7%
20	900	892	0	8	Undercount, -0.9%	15	150	143	0	7	Undercount, -4.7%
21	900	885	0	15	Undercount, -1.7%	21	150	115	0	35	Undercount, -23.3%

In summary, the calibration runs captured realistic operator variability while maintaining stable acquisition conditions. For the pick bins, the overall calibration curve shows a broad OK plateau, supporting robust mid-plateau selection of $m^*_{\text{pick}} = 11$. In contrast, the End (deposit) bin displays a narrower OK region due to short deposit dwell times and incidental corridor flyovers, motivating a separate persistence setting of $m^*_{\text{deposit}} = 10$. Finally, the ROI perturbation heatmaps quantify boundary tolerance and identify fragile edges that should be validated first after any repositioning of the camera or bins.



(a) $W_{\text{min,pick}} = 3$ (b) $W_{\text{min,deposit}} = 2$

Fig. 5. ROI edge sensitivity based on plateau preservation. Heatmaps show the maximum ROI-edge perturbation Δ that preserves an OK plateau of width at least W_{min}: (a) pick region ROIs (Bin1-Bin6); (b) deposit region (End bin). A cell value indicates the largest tested Δ that still satisfies the plateau criterion, while "x" denotes edges for which the smallest tested perturbation is not possible or fails to preserve the required plateau width W_{min}.

V. DISCUSSION AND CONCLUSIONS

This paper presents a calibration method for an edge vision measurement system that monitors bin interaction events in manual assembly using monocular video and pose-derived wrist keypoints. A standardised calibration procedure jointly selects ROI boundaries and temporal persistence parameters from protocol runs. Across five operators and varying postures, the calibration algorithm produced a robust operating region for counting pick-bin touches, while revealing that deposit events require separate temporal treatment due to shorter dwell times and corridor flyovers. ROI perturbation analysis quantified boundary tolerances and identified fragile edges that require particular attention during validation and layout adjustment.

While high accuracy is a prerequisite, the primary contribution is identification of a robust operating region. In production environments where lighting, posture, and lens cleanliness drift over time, a system calibrated at the edge of its performance envelope could fail frequently. By selecting parameters from the midpoint of the OK plateau, the system operates within a stable region rather than at a fragile optimum. The selection of a single parameter set valid across all five operators is critical for practical deployment. This universal calibration eliminates the need for operator-specific profiles or manual reconfiguration during shift changes. It ensures that a standard work event is defined

consistently across the workforce, enabling objective performance benchmarking and fair comparison across shifts. The method treats operator variability as an expected characteristic rather than noise to be eliminated.

The narrow OK plateau for the deposit bin is not only a technical limitation; it also serves as a diagnostic indicator of workstation layout efficiency. The calibration quantitatively demonstrates that the high-traffic corridor between the work area and Bin6 introduces systematic risk of false positives, while short dwell times increase the risk of false negatives. This provides engineers with data-driven justification for layout modifications. The calibration protocol therefore serves a dual purpose: it selects operational parameters for robust event counting and provides diagnostic evidence regarding workstation ergonomics and material flow. This dual role yields two tangible outputs:

- A robust event detector: spatiotemporal parameters that reliably count bin-interaction events events across a diverse workforce, remaining valid despite natural operator-to-operator variability.
- A process diagnostic tool: sensitivity maps that quantify layout robustness to boundary placement errors, transforming calibration from a one-time setup step into ongoing intelligence for continuous improvement.

Regarding limitations, the approach relies on a stable camera viewpoint and requires recalibration after any camera or bin movement. This can be addressed in practice by mechanically securing the mounting. Wide-angle lenses introduce geometric distortion, which may reduce the suitability of regions of interest near the image peripheries; future iterations could incorporate image undistortion to support metric-space definitions. Formal out-of-sample validation on an independently collected, protocol-conforming dataset remains an important direction for future work. Scalability can be challenging in densely packed layouts. In such cases, the sensitivity maps provide actionable feedback: edges with near-zero perturbation tolerance indicate insufficient physical separation and can guide workstation layout adjustments and future design rules. For layouts where depth ambiguity materially affects interaction assignment, future work could evaluate RGB-D sensing within the same calibration framework.

By framing vision-based monitoring as a measurement system requiring formal calibration, this work demonstrates that robust event detection is achievable with low-cost edge devices. The approach directly enables Quality 4.0 principles: it provides objective evidence for audit trails; its edge architecture allows flexible deployment where traditional automation is cost-prohibitive; and universal calibration ensures fair and standardized metrics, fostering data-driven improvement rather than individual surveillance.

In conclusion, this work provides the metrological foundation for treating operator motion as a reliable process signal. The calibration protocol ensures the transition from raw video to quality-relevant events is transparent, robust, and auditable, transforming the manual assembly station from a black box of human variability into a measurable, optimizable node within the connected factory. Future work will focus on integrating this calibrated sensor stream with digital twin platforms for real-time simulation and predictive quality feedback.

ACKNOWLEDGMENT

This work was funded in part by European union under NextGenerationEu initiative, project DREAM (Digital Readiness and Enhancement for Advanced Manufacturing) – uniri-iz-25-4.

REFERENCES

Andrianakos, G., Dimitropoulos, N., Michalos, G., Makris, S. (2019), "An approach for monitoring the execution of human based assembly operations using machine learning", Procedia CIRP, Vol. 86, pp. 198-203.

Ay, Ö. and Emel, E. (2025), "Real-Time Assembly Task Validation Using Deep Learning-Based Object Detection and Operator's Hand-Joints Trajectory Classification", IEEE Access, Vol. 13, pp. 57009-57029.

Berg, J., Reckordt, T., Richter, C., Reinhart, G. (2018), "Action Recognition in Assembly for Human-Robot-Cooperation using Hidden Markov Models", Procedia CIRP, Vol. 76, pp. 205-210.

Cao, Z., Hidalgo, G., Simon T., Wei S.-E., Sheikh Y. (2021), "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields", IEEE transactions on pattern analysis and machine intelligence, Vol. 43 No. 1, pp. 172-186.

Chen, H., Feng, R., Wu, S., Xu, H., Zhou, F., Liu, Z. (2022), "2D Human Pose Estimation: A Survey", Multimedia Systems, Vol. 29 No. 5, pp. 3115-3138.

Faccio, M., Ferrari, E., Galizia, F.G., Gamberi, M., Pilati, F. (2019), "Real-time assistance to manual assembly through depth camera and visual feedback", Procedia CIRP, Vol. 81, pp. 1254-1259.

Gao, Z., Chen, J., Liu, Y., Jin, Y, Tian, D. (2025), "A systematic survey on human pose estimation: upstream and downstream tasks, approaches, lightweight models, and prospects", Artificial Intelligence Review, Vol. 58 No. 3, p. 68.

Godina, R. and Matias J.C.O. (2019), "Quality control in the context of industry 4.0", Melão, N., Reis, J., Pinelas, S, Industrial Engineering and Operations Management II - XXIV IJCIEOM 2018, Springer, Cham, pp. 177-187.

Kaczmarek, S., Hogreve, S., Tracht, K. (2015), "Progress Monitoring and Gesture Control in Manual Assembly Systems Using 3D-image Sensors", Procedia CIRP, Vol. 37, pp. 1-6.

Liu, Y., Zhou, G., He, W., Zhu, H., Cui, Y. (2025), "Detail enhanced high-resolution network for human pose estimation", PLoS One, Vol. 20 No. 9, p. e0325540.

Mahin, M., Kadasah N., Alsabban, A, Albliwi, S. (2024), "Exploring the landscape of quality 4.0: a comprehensive review of its benefits, challenges, and critical success factors", Production and Manufacturing Research, Vol. 12 No. 1, p. 2373739.

Menanno, M., Riccio, C., Benedetto, V., Gissi, F., Savino, M. M., & Troiano, L. (2024), "An Ergonomic Risk Assessment System Based on 3D Human Pose Estimation and Collaborative Robot", Applied Sciences, Vol. 14 No. 11, p. 4823.

Mokuwe, M., Goorun, Y., Crafford, G. (2022), "Assembly Line Quality Assurance Through Hand Tracking and Object Detection", MATEC Web of Conferences, Vol. 370, p. 07003.

Mühlbauer, M., Kutzner, K., Sommer, A., Würschinger, H., Hanenkamp, N. (2022), "An Approach to Progress Monitoring of Industrial Manual Processes Based on Camera Recordings and Object Interactions", Procedia CIRP, Vol. 107, pp. 582-587.

Oyekan, J., Fischer, A., Hutabarat, W., Turner, C., Tiwari, A. (2020), "Utilising low cost RGB-D cameras to track the real time progress of a manual assembly sequence", Assembly Automation, Vol. 40 No. 6, pp. 925–939.

Princz, G., Shaloo, M., Erol, S. (2023), "A literature review on the prediction and monitoring of assembly and disassembly processes in discrete make-to-order production in SMEs with machine vision technologies", in ICIEAEU '23: Proceedings of the 2023 10th International Conference on Industrial Engineering and Applications, Association for Computing Machinery, New York, NY, USA, pp. 318-327.

Rückert, P., Papenberg, B., Tracht, K. (2021), "Classification of assembly operations using machine learning algorithms based on visual sensor data", Procedia CIRP, Vol. 97, pp. 110-116.

Zogopoulos, V., Birem, M., De Geest, R., Hofman, R., Jorissen, L., Vanherle, B., Gors, D. (2021), "Image-based state tracking in Augmented Reality supported assembly operations", Procedia CIRP, Vol. 104, pp. 1113-1118.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Investigating the impact of lean digital transformation on sustainability – The role of top management and green practices in the pharmaceutical industry

Evgenia Saltaferi [1], Maria Gianni [2,3] and Argyro Fragkaki [4]

[1] Hellenic Open University, Patras, Greece; esaltaferi@gmail.com
[2] Hellenic Open University, Patras, Greece; gianni.maria@ac.eap.gr
[3] International Hellenic University, Serres, Greece; mariagianni@ihu.gr
[4] Hellenic Open University, Patras, Greece; fragkaki.argyro@ac.eap.gr

Abstract* — The pharmaceutical industry currently faces strict regulations and traceability requirements, as well as a shift towards personalised therapies. This calls for smart and environmentally responsible factories. Digital transformation in the pharmaceutical industry has therefore grown rapidly into what is known as Pharma 4.0. In this context, this study aims to investigate whether and in what ways lean management practices can promote sustainability. It also examines the roles of top management commitment, employee training, green practices, and smart technologies. A review of the literature yielded a structural model where each variable is measured using a validated subscale. A five-point Likert scale questionnaire was first pilot-tested and then disseminated to collect data from employees of ten pharmaceutical companies (seven in Greece and three in the Netherlands). A total of 301 responses were analysed using the partial least squares structural equation modelling approach. The measurement model exhibited high internal consistency and satisfactory convergent validity. Most of the hypothesised relationships were proven to be statistically significant. Specifically, top management commitment positively influenced the implementation of green practices and lean management, while training had a significant positive effect on lean management and smart technologies. Additionally, smart technologies positively influenced digital transformation. The sequential effect of lean management and digitalisation on sustainability performance, confirmed that adopting Lean 4.0 accrues economic, environmental, and social benefits. This research demonstrates that applying lean and green perspectives on digital transformation can enhance sustainability in the pharmaceutical sector. Leadership commitment and continuous employee training are critical in steering the adoption of green practices, fostering the acceptance of smart technologies, and strengthening a culture of continuous improvement.

Keywords – lean, pharmaceutical, digital, sustainability

I. Introduction

In their efforts to survive and thrive, businesses are constantly looking for ways to increase their sustainability performance. This is particularly relevant in the pharmaceutical industry, where rapid technological developments and competition are advancing rapidly. Furthermore, strict regulatory requirements and the need for innovation and continuous process improvement make methodologies that increase efficiency and flexibility an imperative (Carneiro et al., 2026). According to Reinhardt et al. (2020), the pharmaceutical business model is transitioning to "a lean, focussed entity with a research footprint within key innovation bio-clusters and a growing revenue stream from specialty products and biologics through emerging markets".

In this context, lean management (LM) emerges as a strategic approach, which aims to reduce waste of resources, standardise and streamline processes, and increase efficiency at all levels of the drug manufacturing business. Meanwhile, digital transformation (DT) is bringing about radical changes in the way information and business processes are managed, strengthening organisations' ability to respond quickly to changing market demands, and improve communication and decision-making. Human resources play a crucial role in the successful implementation of these innovative practices, as they implement the changes (Gianni et al., 2026).

Accepting and integrating lean management principles alongside the possibilities offered by digital transformation can substantially improve operational efficiency and, consequently, help to achieve corporate sustainability. While the potential for increasing automation and autonomy in business operations is significant, notable gaps remain in the research on actual digitalisation progress (Miozza et al., 2024). This study explores the extent to which lean management principles affect sustainability in pharmaceutical companies, when integrated with digital applications. To this end, the following research questions are postulated:

Question 1: Do lean management practices affect the sustainability of pharmaceutical companies?

Question 2: What role does digital transformation play in facilitating the implementation of lean management and enhancing business results?

Question 3: What other organisational factors may influence the relationships between lean, digital and sustainability variables?

To answer these research questions, first the review of relevant literature led to the development of a conceptual framework. A survey was designed to collect data from employees of ten pharmaceutical companies, three of which are based in the Netherlands and seven in Greece. This approach allowed the differentiation of respondents in terms of their

demographic characteristics and their position in the sample organisations. Data processing tested the hypothesized relationships using structural equation modelling. Next, the results are discussed, generating practical and theoretical implications.

II. Literature Review

The philosophy of lean management aims to eliminate non-value-added activities and improve the efficiency of business processes. Originating from the Toyota Production System, the lean approach has been widely adopted in various industries such as the automotive, pharmaceutical and service industries. Over the years, lean has proven its effectiveness in minimising waste and increasing productivity (Santana et al., 2017). A key element of this approach is the idenfication of operational waste, that is, those activities that do not add value to the final product or service. For example, overproduction leads to the production of more units than required, but also creates unnecessary inventories and increased storage costs (Arunagiri and Gnanavelbabu, 2014). At the same time, waiting times – whether due to machine delays or insufficient information flow – reduce the efficiency of the system (Harish & Selvam, 2015). In sectors with high quality requirements, such as the pharmaceutical industry, the application of lean management principles has proven to be extremely effective. Reducing waste and optimizing processes contribute to achieving strict quality and accuracy standards, parameters that are critical for the production of medical and pharmaceutical products (Santana et al., 2017). Using tools such as Value Stream Mapping and Kaizen, companies can quickly identify deviations, thereby optimizing inventory management and reducing waiting times in production processes.

In pharmaceuticals, the practice of "Green Kaizen" attracts positive employee engagement and translates waste reduction into tangible environmental sectors (Kurdve and Bellgran, 2019). Furthermore, lean thinking is fuelled by green education that aligns individual behaviours with sustainability goals (Bellgran and Kurdve, 2021). Green practices translate internal lean logic into external environmental value. The integration of environmental design, z and 6R in a green supply chain simultaneously improves economic and ecological indicators (Delaney et al., 2022; Zhu et al., 2013). Integrating these practices into integrated management systems offers a holistic framework that combines quality, environment and health and safety standards, bridging triple bottom line measurement with operational management (Gianni et al., 2017). A recently proposed "lean-green" theoretical framework suggested that the coupling of these two axes of lean and green practices promotes increased sustainability performance compared to individual initiatives (Kosasih et al., 2023).

The smart technologies of Industry 4.0 act as a digital lever that multiplies the impact of lean management. The integration of IoT sensors, big-data analytics and AI transforms classic lean management tools such as Kanban, Andon and Poka-Yoke into digital learning platforms (Sony, 2018), improving quality and compliance with changes (Ding, 2018). The systematic review by Buer et al. (2021) documents that the combination of Lean and digital transformation leads to higher operational performance compared to individual applications, while empirical evidence shows that smart technologies directly enhance all dimensions of the triple bottom line (Kamble et al., 2020). Furthermore, studies indicate that the interaction of Industry 4.0 and sustainability is particularly strong in companies with a culture of learning and collaboration (Isensee et al., 2020) and is considered essential for the Pharma 4.0 transformation (Saha et al., 2022).

Top management is the key lever that supports lean, green and digital transformation. According to the theory of the upper echelons, the aspirations and vision of top executives determine the success of the transformation (Latan et al., 2018). A study in the pharmaceutical industry shows that strong leadership commitment is critical for the implementation of Lean 4.0 while its lack limits sustainability performance despite technological investments (Tetteh et al., 2024). As a result, training fosters green culture and equips staff with lean and digital skills, top management aligns strategy, resources and innovation, smart technologies offer continuous monitoring and full visibility of processes through data for continuous optimization, while green practices transform waste reduction into environmental and social metrics that, in turn, reinforce engagement and learning. Through this cycle, lean management becomes the backbone of operational excellence, digital transformation acts as an accelerator of innovation, and Sustainability gains strategic weight and social legitimacy. Together, they form an ecosystem of excellence capable of supporting the pharmaceutical industry in the era of Pharma 4.0 (Beier et al., 2022; Calabrese et al., 2022; Ologeanu-Taddei et al., 2025).

The aforementioned argumentation leads to a nexus of constructs with possible relationships illustrated in Fig. 1.

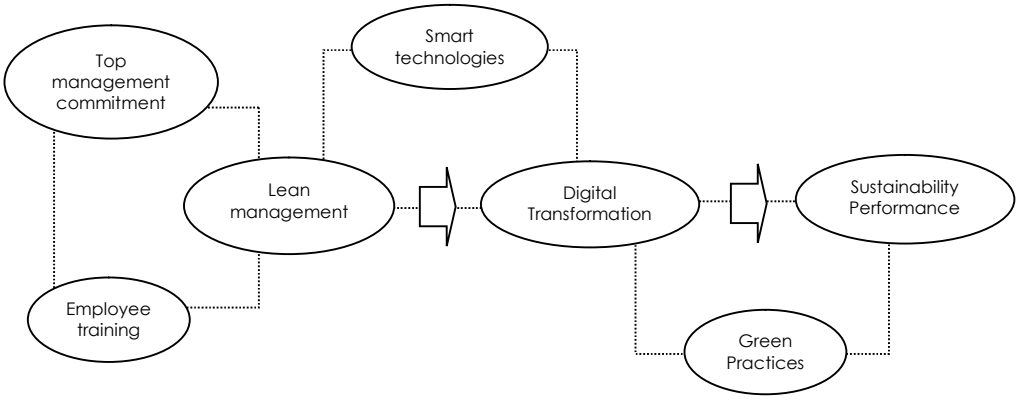


Fig. 1. Conceptual framework (Source: Authors' own work)

III. Research Methods and Tools

A. Measurement instrument

This study adopted reliable scales that have been developed, used, and validated by previous research. Lean management and employee training were measured using corresponding scale components of Bortolotti et al. (2015). For the green practices as well as for top management leadership and commitment in green and lean practices the questions were adopted from Kosasih et al. (2023). Digital transformation and smart technologies were measured with the subscales of Nasiri et al. (2020). Sustainability performance was measured using a set of questions originally composed by Siltori et al. (2021) and later tested using exploratory factor analysis by Baxevanidis Taros et al. (2025). This last scale includes questions on the impact of digital technologies on the economic, social and environmental aspects of business operations. Before dissemination, the questionnaire was shared with six experts – two academics and four practitioners to examine its wording, clarity, and structural adequacy. It was returned with minor recommendations for improvement (correction of grammatical errors due to translation of the original question, removal of duplicates).

B. Sample organisations

The survey questionnaire was administered to ten pharmaceutical companies, seven of which are based in Greece and three in the Netherlands. To facilitate the collection and processing of the results, and to preserve anonymity, each company was given a code name consisting of the Latin letter C and a serial number ranging from 1 to 10. The sample organisations are briefly described in Table I.

TABLE I. Sample Organisations Description

Case	Location / HQ	Founded	Sector / Focus	Main Products / Activities	Employees	Facilities / Reach	Standards / Certifications
C1	Attica, Greece (Athens)	1974	Pharmaceutical manufacturing	Non-sterile solids/liquids, creams, ointments, gels, sachets, dietary supplements	~200	2 production plants in Attica	ISO 9001, GMP
C2	Attica, Greece	1965	Pharmaceutical company	Development, production & distribution of injectables	>1,000	Exports to 85 countries; supplier to UN & WHO	ISO 9001, GMP
C3	Bilthoven, Netherlands	2023	Biotechnology / Dermatology	Endolysin-based skincare and dermatological solutions	~30	R&D and commercialization activities	EU clinical and biosafety regulations
C4	Netherlands	1961	Healthcare / OTC pharmaceuticals	Development, production & distribution of non-prescription medicines	~300	Presence in multiple European countries	ISO 9001, GMP
C5	Amsterdam, Netherlands	2018	Clinical-stage biotechnology	T-cell receptor therapies targeting cancer neoantigens	~100	Modern laboratories in Amsterdam	EU clinical and biosafety regulations
C6	Attica, Greece	1969	Pharmaceutical company	Slow-release injectables, controlled-release and ophthalmic formulations	>1,600	FDA & EU-approved facilities; 250 customers in 90+ countries	FDA, EU regulations
C7	Pikermi, Attica, Greece	1965	Pharmaceutical company	Original & generic medicines	>900	26,000 sq.m. facilities; exports to 60+ countries	ISO 9001, GMP
C8	Greece	1974	Pharmaceutical company	Ophthalmic & injectable formulations	>700	14,000 sq.m. facilities with 3 advanced buildings	GMP, ISO 14001
C9	Attica, Greece	1974	Pharmaceutical company	Ophthalmic & injectable formulations	~650	Modern production & research facilities	FDA, GMP
C10	Near Athens, Greece	1967	Generic pharmaceuticals	Development, production & distribution of generic medicines	~200	Manufacturing near Athens; warehouse 100 km away; Exports	International quality & safety standards

IV. Results and Discussion

A. Sample Characteristics

Data collection took place over the course of one month from January to February 2025. A total of 301 complete and usable responses were received from employees of the ten companies – 236 of whom reside in Greece and 65 of whom reside in the Netherlands. Of these, 27 were top management executives, 149 were medium-level blue collar employees, 46 were low-level blue-collar employees, 33 were medium-level white-collar employees, and 46 were low-level white-collar employees. Around half of the respondents were aged between 26 and 35 years, around 25% were aged between 36 and 45, around 10% were under 55. The proportion of people aged under 25 and over 55 was less than 5%. In terms of their roles, the majority (around 40%) worked in quality assurance, 33% in R&D, 16% in packaging and 7% in purchasing.

The business context in the two countries represented in the sample exhibits both similarities and differences. Both Greece and the Netherlands are members of the European Union, which imposes a unified legal and institutional framework. However, the Dutch pharmaceutical sector is strong in terms of R&D and innovation, while Greek pharmaceutical

companies have a clear export orientation. Relevant literature emphasizes the increasing importance of external innovation networks, alliance portfolios, and collaborative R&D models in sustaining pharmaceutical competitiveness (Schuhmacher et al., 2018). Dutch pharmaceutical companies also prioritise cost-effectiveness and regulated competition, with reforms encouraging economic evaluation and efficient drug pricing while preserving innovation incentives (Canoy and Tichem, 2018). In contrast, the Greek pharmaceutical industry operates in a challenging environment characterized by financial constraints and stringent regulatory frameworks. Under these conditions, drug manufacturers focus on remaining competitive by producing conventional medicines rather than allocating resources to innovative ones. Nevertheless, the sector has demonstrated notable agility, particularly in quality management processes, as well as resilience by expanding exports, increasing domestic production, and investing in research infrastructure. According to recent studies, pharmaceutical exports accounted for around 5.7% of total Greek exports in 2024, while R&D spending exceeded €160 million (FEIR, 2025).

B. Validity and Reliability Analysis

The Cronbach's alpha coefficient values for all constructs were above 0.8, well exceeding the acceptance threshold of 0.6-0.7. The Fornell-Larcker criterion confirmed adequate discriminant validity for all the variables in the model. The average variance extracted (AVE) was found to be above the threshold value of 0.5 for all variables except for sustainability performance, for which the AVE value was 0.473. This was due to the low outer loading (0.395) of one item in the scale corresponding to the statement "Digital technologies (will) make production more autonomous and efficient, leading to a reduction in jobs." Interestingly, this is the only question in the scale that has a negative semantic association with the impact of digital transformation on sustainability performance. As the model is reflective, this item was dropped to improve convergent validity. Nevertheless, this finding contributes to the discourse on the adverse impact of digital transformation on people (fear of job loss) and on the environment, e.g. the high-energy demand of AI data centres (Bogmans et al., 2026).

C. Hypothesis Testing

Bootstrapping with 5000 subsamples tested the hypothesised relationships. The results are illustrated in Fig. 2. For clarity, non-significant paths have been omitted. As expected, top management with a high level of commitment and engagement in promoting lean principles and green practices has a statistically significant positive effect on the respective variables in the model. Training of employees in lean practices has a significant positive effect on lean management. R-square values of LM and DT are quite high (60% and 43% respectively), indicating the model's high predictive power (Hair et al., 2022: p. 195). The hypothesis testing results are summarised in Table II.

TABLE II. HYPOTHESES TESTING

Hypothesis	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	P- value
Digital Transformation → Sustainability Performance	0.306	0.308	0.064	<0.001
Green Practices → Sustainability Performance	0.276	0.280	0.061	<0.001
Lean Management → Digital Transformation	0.238	0.239	0.046	<0.001
Lean Management → Smart Technologies	0.402	0.406	0.050	<0.001
Smart Technologies → Digital Transformation	0.521	0.522	0.041	<0.001
Top Management → Green Practices	0.774	0.774	0.026	<0.001
Top Management → Lean Management	0.359	0.359	0.044	<0.001
Training → Lean Management	0.539	0.541	0.041	<0.001

Mediation Analysis

LM has a significant direct positive effect on DT (0.447, p<0.001). Smart technologies (ST) have a significant indirect (specific) positive effect on lean management and digital transformation (LM → ST → DT: 0.209(=0.402*0.521), p<0.001). The product of the indirect and the direct effects is positive (0.402*0.521*0.447) confirming the mediation hypothesis. Specifically, this is a complementary mediation, which implies that other mediators may co-exist (Nitzl et al., 2016; Zhao et al., 2010). LM has a non-significant direct positive effect on SP, whereas the indirect effect of DT on LM and SP is positive and significant (0.238*0.301, p<0.001). This confirms the second mediation hypothesis (LM→DT→SP). Specifically, this means that DT fully mediates the LM to SP relationship Hair et al., 2022: p. 243).

Control factors

Given the sample taxonomies, this study tested two key control factors: company and country. Company's effect on SP was found significant (0.140, p=0.008). This finding shows that a company's internal context plays a significant role in influencing DT on SP. Significant differences were also found in the effects of country (1: Greece, 2: the Netherlands) on LM, ST and GP. The negative coefficient for GP (-0.277, p=0.003) implies that companies in the Netherlands report lower levels of implementation of green practices than companies in Greece with regard to top management influence. Similarly, the negative coefficient for ST (-0.610, p<0.001) implies that companies in the Netherlands report lower smart technology use than companies in Greece with regard to lean management influence. In fact, the country effect on ST is quite large (R-square increases from 0.162 to 0.220), thus improving model fit. The positive coefficient for LM (0.206, p=0.026) implies that companies in the Netherlands report higher LM practices than companies in Greece with regard to training influence.

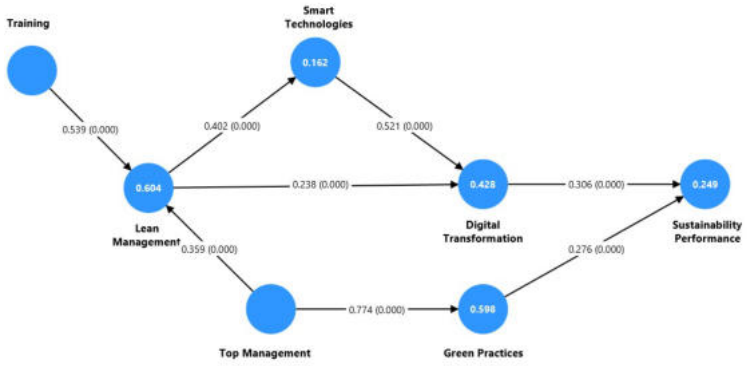


Fig. 2. Hypothesis testing – Bootstrapping results (Source: Authors' own work)

V. Research Implications

This study adds to prior research that reports high 4.0 awareness levels of top executives in the pharmaceutical sector (Reinhardt et al., 2020). Previous studies have emphasised the positive impact of top management commitment on lean or green initiatives (Januszek et al., 2024; Li et al., 2019). The findings underscore the role of top management in fostering lean and green practices together to enhance sustainability. Top managers should focus on such practices to digitally transform their organisations efficiently and effectively. Training has been included and examined as a distinct construct in the model. The findings corroborate that it is a critical factor, as people are considered important in every lean process. Therefore, they should receive continuous and synchronous training. The mediation analysis revealed that lean practices will maximise their effect on sustainability performance through the streamlined transformation of organisations that decide to digitalise their economic, social, and environmental processes. The analysis also identified smart technologies as a mediator for LM and DT. It is noteworthy that ST implementation is measured separately from DT, in line with the study of Nasiri et al. (2020). This emphasises a common misinterpretation of DT as a mere technical challenge. Implementation is one thing, while the process that people, machines and organisations undergo and the outcome of this process is another. In this context, the study verified smart technologies as antecedents of digital transformation within organisations. Apart from certain identified differences, the findings of the study indicate that in both countries the human factor, management commitment, and continuous employee training are critical prerequisites for the successful adoption of Lean 4.0 practices and the enhancement of corporate sustainability. The convergence of LM practices with digital transformation appears to represent a common strategic direction regardless of geographical or organizational context.

For researchers, this study proposed a model that associates managerial aspects with organisational processes and outcomes. Novel insights are offered into the long-lasting debate on the lean-digital-sustainable interplay (Gianni and Gotzamani, 2024a; Lei and Jialbao, 2024). Full mediation of DT on the relationship between LM and SP suggests that lean thinking alone cannot guarantee sustainable results (León and Calvo-Amodio, 2017). Advanced technologies support and promote lean practices amplifying and retaining such results. This draws on the socio-technical systems theory, which captures the interplay between technical and social factors in order to successfully implement processes and internalise knowledge within human-machine R&D and production systems (Gianni et al., 2026; Gianni and Gotzamani, 2024a;b). Further to the above, the findings address another challenge (Bhattacharya et al., 2019): how can lean and green integrate to influence SP? The model provides practical guidance on this issue, proposing a mechanism through which lean and green can enhance sustainability performance simultaneously.

Limitations

The limitations in this study provide opportunities to design and conduct future research on this topic. First, the cross-sectional data raise concerns regarding common method bias and causal inferences. Several precautionary measures were taken in the survey, in accordance with the guidelines of Rindfleisch et al. (2008). Multiple measurement scales and multiple respondents in different positions and hierarchical levels in each company were some of these measures. Second, regarding the generalisability of the study, collecting data from two countries is a step forward to meet such issues. Future studies can test the proposed model in different country contexts. Third, the mediation analysis highlighted the possible mediating role of other variables. Future research may expand the structural model of this study and include other potential mediators of LM and DT relationship. Lastly, as this model focused on the role of top executives. Future studies may investigate the role of middle managers which are also very much involved in lean success (Holmemo and Ingvaldsen, 2016; Netland et al., 2020; Reynders et al., 2022).

VI. Conclusion

This research demonstrates that combining lean management with digital transformation can boost sustainability in the pharmaceutical sector. Leadership commitment and continuous employee training are critical factors, steering the

adoption of green practices, fostering acceptance of smart technologies, and strengthening a culture of continuous improvement. This study bridges the domains of lean management, digital transformation and sustainability, offering concrete guidance on building resilient, competitive and responsible business models within the pharmaceutical industry.

Traceability and other challenges in medicine production require the assimilation of advances technologies in standards and regulations. Biotechnological companies face different challenges in drug discovery, where people have to deal with data sensitivity when using AI, deep-learning approaches or handling data that are gathered via online sensors and other online systems (Askr et al., 2023). Compatibility, security, interoperability, ethics are only some of the challenges that accompany smart technologies. Lean and green practices provide organisations with the resources and capabilities to address these challenges. The mechanisms proposed in this study can be used to this end.

Pharmaceuticals go hand-in-hand with healthcare. At the intersection of these two fields, information management systems and standards, augmented reality applications, big data analytics, machine learning, and robots are being gradually incorporated into daily operations. Top executives develop strategies, make decisions and allocate resources to promote digital transformation based on a systematic, regulatory-compliant, environmentally friendly groundwork. This study focuses on the role of top management in facilitating the digital transformation of pharmaceutical companies. Future studies could modify the model to investigate the role of middle managers.

REFERENCES

- Arunagiri P., and Gnanavelbabu A. (2014), "Identification of major lean production waste in automobile industries using weighted average method", Procedia Engineering, Vol. 97, pp. 2167–2175. https://doi.org/10.1016/j.proeng.2014.12.460
- Askr, H., Elgeldawi, E., Aboul Ella, H., Elshaier, Y.A.M.M., Gomaa, M.M., and Hassanien, A.E. (2023), "Deep learning in drug discovery: an integrative review and future challenges", Artificial Intelligence Review, Vol. 56, pp. 5975–6037. https://doi.org/10.1007/s10462-022-10306-1
- Baxevanidis Taros, P., Gianni, M., Gotzamani, K., and Fragkaki, A. (2026), "Embedding Lean Principles and Digital Transformation into Organizational Culture: Effects on Industry 4.0 Adoption and Corporate Sustainability Performance – A Case Study in Metal Processing", In Cantini, A., Caterino, M., Fantozzi, I.C., Leoni, L., and Mancusi, F. (Ed.) Advances in Performance Management and Measurement for Industrial Applications and Emerging Domains - Proceedings of the Third Conference on Performance Measurement and Management (COPERMAN), Lecture Notes in Production Engineering, Springer Cham.
- Beier G., Matthess M., Guan T., Grudzien D. I.O.P., Xue B., Lima E.P., and Chen L. (2022), "Impact of Industry 4.0 on corporate environmental sustainability: Comparing practitioners' perceptions from China, Brazil and Germany", Sustainable Production and Consumption, Vol. 31, pp. 287–300. https://doi.org/10.1016/j.spc.2022.02.017
- Bellgran M., and Kurdve M. (2021), "Green lean operationalisation of the circular economy concept on production shop floor level", Journal of Cleaner Production, Vol. 278, 123223. https://doi.org/10.1016/j.jclepro.2020.123223
- Bhattacharya, A., Nand, A., and Castka, P. (2019), "Lean-green integration and its impact on sustainability performance: A critical review", Journal of Cleaner Production, Vol. 236, 117697.
- Bogmans, C., Ganpurev, G., Gomez-Gonzalez, P., Melina, G., Pescatori, A., and Thube, S. (2026), "Power hungry: How AI will drive energy demand", Energy Economics, 109278. https://doi.org/10.1016/j.eneco.2026.109278
- Bortolotti, T., Boscari, S. and Danese, P. (2015), "Successful lean implementation: organizational culture and soft lean practices", International Journal of Production Economics, Vol. 160, pp. 182-201, doi: 10.1016/j.ijpe.2014.10.013.
- Buer S.-V., Semini M., Strandhagen J.O., and Sgarbossa F. (2021), "The complementary effect of lean manufacturing and digitalisation on operational performance", International Journal of Production Research, Vol. 59 No. 7, pp. 1976–1992. https://doi.org/10.1080/00207543.2020.1790684
- Calabrese A., Costa R., Tiburzi L., and Brem A. (2022), "Merging two revolutions: A human–artificial intelligence method to study how sustainability and Industry 4.0 are intertwined", Technological Forecasting & Social Change, Vol. 188, 122265. https://doi.org/10.1016/j.techfore.2022.122265
- Canoy, M., and Tichem, J. (2018), "Lower drug prices can improve innovation", European Competition Journal, Vol. 14 No. 2–3, pp. 278–304. https://doi.org/10.1080/17441056.2018.1512231
- Carneiro, F., Miguéis, V., Nóvoa, H., Carvalho, A.M., Ferreira, D., Antony, J., Tortorella, G., and Furterer, S. (2026), "A systematic approach to classify and reduce recurrent deviations in the pharmaceutical industry: A detailed case study", Quality Management Journal, Vol. 33 No. 1, pp. 41–60. https://doi.org/10.1080/10686967.2025.2578507
- Delaney, E., Liu, W., Zhu, Z., Xu, Y., & Dai, J. S. (2022), "The investigation of environmental sustainability within product design: a critical review", Design Science, Vol. 8, e15. doi:10.1017/dsj.2022.11
- FEIR (2025), "The pharmaceutical market in Greece: Facts and Figures 2024", Foundation for Economic and Industrial Research, Greece. https://iobe.gr/wp-content/uploads/2026/01/RES_05_A_25092025_REP_EN.pdf
- Gianni M., Gotzamani K., and Tsiotras G. (2017), "Multiple perspectives on integrated management systems and corporate sustainability performance", Journal of Cleaner Production, Vol. 168, pp. 1297–1311. https://doi.org/10.1016/j.jclepro.2017.09.061
- Gianni, M. and Gotzamani, K. (2024a), "Lean Digital Culture as an Enabler of Corporate Sustainability Performance: The Mediating Role of Intention to Use Industry 4.0 Technologies." In: Ciasullo, M.V., Martin, J., and Brunetti, F. (Eds.) Embracing Sustainability Management through Excellence in Services. EISIC 2023. Springer Proceedings in Business and Economics. Springer Nature, Cham. pp. 58-74. doi: 10.1007/978-3-031-65115-1_3

- Gianni, M. and Gotzamani, K. (2024b), "The internalisation of integrated management systems as a key for sustainable companies: Empirical evidence", Total Quality Management & Business Excellence, Vol. 35 No. 1-2, pp. 273-296. https://doi.org/10.1080/14783363.2023.2290659
- Gianni, M., Gotzamani, K., Konstantaras, I., Kessopoulou, E., and Xanthopoulou, S. (2026), "A lean management perspective on digital transformation: The role of culture in technology adoption", International Journal of Quality & Reliability Management, Vol. 43 No. 4, pp. 981-1004. https://doi.org/10.1108/IJQRM-12-2024-0444
- Isensee C., Teuteberg F., Griese K.-M., and Topi C. (2020), "The relationship between organizational culture, sustainability, and digitalization in SMEs: A systematic review", Journal of Cleaner Production, Vol. 275, 122944. https://doi.org/10.1016/j.jclepro.2020.122944
- Hair, J.F., Hult, T.M., Ringle, C.M., and Sarstedt, M. (2022), "A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)", 3nd ed., Sage Publications: Thousand Oaks, CA, USA.
- Harish, K.A., and Selvam, M. (2015), "Lean wastes: A study of classification from different categories and industry perspectives", The Asian Review of Civil Engineering, Vol. 4 No. 2, pp. 7–12. https://doi.org/10.51983/tarce-2015.4.2.2220
- Holmemo, M.D.Q., and Ingvaldsen, J.A. (2016), "Bypassing the dinosaurs? – How middle managers become the missing link in lean implementation", Total Quality Management & Business Excellence, Vol. 27 No. 11–12, pp. 1332–1345. https://doi.org/10.1080/14783363.2015.1075876
- Januszek, S., Netland, T.H., Furlan, A. (2024), "The role of managerial perceptions and behaviors across hierarchical levels during lean implementation", International Journal of Operations & Production Management, Vol. 44 No. 1 pp. 54–74, doi: https://doi.org/10.1108/IJOPM-07-2022-0417
- Kamble S., Gunasekaran A., and Dhone N. C. (2020), "Industry 4.0 and lean manufacturing practices for sustainable organisational performance in Indian manufacturing companies", International Journal of Production Research, Vol. 58 No. 5, pp. 1319–1337. https://doi.org/10.1080/00207543.2019.1630772
- Kosasih, W., Pujawan, I.N., Karningsih P.D., and Shee, H. (2023), "Integrated lean-green practices and supply chain sustainability framework", Cleaner and Responsible Consumption, Vol. 11, 100143. https://doi.org/10.1016/j.clrc.2023.100143
- Kurdve M., and Bellgran M. (2019, "Cost-driven Green Kaizen in pharmaceutical production – Creating positive engagement for environmental improvements", Procedia CIRP, Vol. 81, pp. 1219–1224. https://doi.org/10.1016/j.procir.2019.03.297
- Latan, H., Jabbour, C.J.C., de Sousa Jabbour, A.B.L., Wamba, S.F., and Shahbaz, M. (2018), "Effects of environmental strategy, environmental uncertainty and top management's commitment on corporate environmental performance: The role of environmental management accounting", Journal of Cleaner Production, Vol. 180, pp. 297-306.
- Lei, L., and Jialbao, L. (2024), "Digital transformation for the sustainable development of firms: The role of green capability and green culture", Sustainable Development, Vol. 32 No. 3, pp. 1861-1875.
- León, H.C.M., and Calvo-Amodio, J. (2017), "Towards lean for sustainability: Understanding the interrelationships between lean and sustainability from a systems thinking perspective", Journal of Cleaner Production, Vol. 142, pp. 4384-4402.
- Li, Y., Ye, F., Dai, J., Zhao, X., and Sheu, C. (2019), "The adoption of green practices by Chinese firms: Assessing the determinants and effects of top management championship". International Journal of Operations & Production Management, Vol. 39 No. 4, pp. 550–572, doi: https://doi.org/10.1108/IJOPM-12-2017-0753
- Miozza, M., Brunetta, F., and Appio, F.P. (2024), "Digital transformation of the Pharmaceutical Industry: A future research agenda for management studies", Technological Forecasting and Social Change, Vol. 207, 123580.
- Nasiri, M., Ukko, J., Saulina M., and Rantala, T. (2020), "Managing the digital supply chain: The role of smart technologies", Technovation, Vol. 96-97, 102121. https://doi.org/10.1016/j.technovation.2020.102121
- Netland, T.H., Powell, D.J., and Hines, P. (2020), "Demystifying lean leadership". International Journal of Lean Six Sigma, Vol. 11 No. 3, pp. 543–554. doi: https://doi.org/10.1108/IJLSS-07-2019-0076
- Nitzl, C., Roldan, J.L., and Cepeda, G. (2016), "Mediation analysis in partial least squares path modeling: Helping researchers discuss more sophisticated models", Industrial Management & Data Systems, Vol. 116 No. 9, pp. 1849–1864.
- Ologeanu-Taddei R., Mittermeier F., Hönigsberg S., Tana S., Weritz P., Dang D., Wache H., Hautala-Kankaanpää T., and Pekkola S. (2025), "The relationship of digital transformation and corporate sustainability: Synergies and tensions", Technological Forecasting & Social Change, Vol. 210, 123809. https://doi.org/10.1016/j.techfore.2024.123809
- Reinhardt, I.C., Oliveira, J.C., and Ring, D.T. (2020), "Current perspectives on the development of industry 4.0 in the pharmaceutical sector", Journal of Industrial Information Integration, Vol. 18, 100131.
- Reynders, P., Kumar, M., and Found, P. (2022), "Lean on me': an integrative literature review on the middle management role in lean", Total Quality Management & Business Excellence, Vol. 33 No. 3–4, pp. 318–354. https://doi.org/10.1080/14783363.2020.1842729
- Saha E., Rathore P., Parida R., and Rana N.P. (2022), "The interplay of emerging technologies in pharmaceutical supply chain performance: An empirical investigation for the rise of Pharma 4.0", Technological Forecasting & Social Change, Vol. 181, 121768. https://doi.org/10.1016/j.techfore.2022.121768
- Santana A., Afonso P., Zanin A., and Wernke R. (2017), "Implementation of lean production systems in small and medium-sized pharmaceutical enterprises", Procedia Manufacturing, Vol. 21, pp. 814–821. https://doi.org/10.1016/j.promfg.2018.02.188
- Schuhmacher, A., Gassmann, O., McCracken, N., and Hinder, M. (2018), "Open innovation and external sources of innovation. An opportunity to fuel the R&D pipeline and enhance decision making?", Journal of Transnational Medicine, Vol. 16, Article 119. https://doi.org/10.1186/s12967-018-1499-2

- Siltori, P.F.S., Anholon, R., Rampasso, I.S., Quelhas, O.L.G., Santa-Eulalia, L.A., and Filho W.L., (2021), "Industry 4.0 and corporate sustainability: An exploratory analysis of possible impacts in the Brazilian context", Technological Forecasting & Social Change, Vol. 167, 120741. https://doi.org/10.1016/j.techfore.2021.120741
- Sony, M. (2018), "Industry 4.0 and lean management: a proposed integration model and research propositions", Production & Manufacturing Research, Vol. 6 No. 1, pp. 416–432. https://doi.org/10.1080/21693277.2018.1540949
- Tetteh M.G., Gupta S., Kumar M., Trollman H., Salonitis K., and Jagtap S. (2024), "Pharma 4.0: A deep dive into top management commitment to successful Lean 4.0 implementation in Ghanaian pharma manufacturing sector", Heliyon, Vol. 10, e36677. https://doi.org/10.1016/j.heliyon.2024.e36677
- Zhao, X., Lynch, J.G. and Chen, Q. (2010), "Reconsidering Baron and Kenny: myths and truths about mediation analysis", Journal of Consumer Research, Vol. 37 No. 3, pp. 197-206.
- Zhu, Q., Sarkis, J., and Lai, K. H. (2013), "Institutional-based antecedents and performance outcomes of internal and external green supply chain management practices", Journal of Purchasing and Supply Management, Vol. 19 No. 2, pp. 106-117.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Understanding Lean Manufacturing and Industry 4.0 integration: the interplay of sociotechnical structures

João José Medeiros Alfineto[1], Kleber Francisco Esposto[2], Fabiane Letícia Lizarelli[1], Susana Duarte[3], Igor Roberto Amancio[1] and Lígia Lobo Mesquita[4]

[1] Federal University of São Carlos; joaojjma@gmail.com
[2] São Carlos School of Engineering, University of São Paulo
[3] NOVA School of Science and Technology, NOVA University Lisbon, Portugal
[4] Federal University of Paraíba

***Abstract*— The integration of Lean Manufacturing and Industry 4.0 is often discussed through linear or technology-driven perspectives, which may overlook the mutual influence between organisational practices and digital technologies. This study addresses RQ1: What are the relative and interdependent structural roles of Lean Manufacturing practices and Industry 4.0 technologies within sociotechnical production systems? To answer this question, the study adopts a sociotechnical perspective and applies Interpretive Structural Modelling (ISM), complemented by Matrice d'Impacts Croisés Multiplication Appliquée à un Classement (MICMAC). The results show that Industry 4.0 technologies have high driving power, mainly by shaping information flows and coordination mechanisms, while Lean Manufacturing practices play important mediating roles by translating digital capabilities into operational routines. These findings suggest that Lean–Industry 4.0 integration should not be understood as a linear sequence, but as an interdependent configuration in which practices and technologies influence one another. The study contributes to debates on Quality 4.0 and Operational Excellence by clarifying how Lean practices and Industry 4.0 technologies jointly support improvement systems in production environments.**

Keywords – Lean Manufacturing, Industry 4.0, Interpretive Structural Modelling, Operational Excellence.

I. Introduction

The intensification of competition and complexity in production systems has led organisations to seek approaches that address operational efficiency, flexibility, and customisation simultaneously. In this context, Lean Manufacturing and Industry 4.0 technologies have been widely discussed as means to improve organisational performance (Kamble et al., 2018a; Sasso et al., 2025). Lean Manufacturing has been associated with the development of efficient and stable production processes through waste reduction, standardisation, and continuous improvement (Jesus et al., 2025), while Industry 4.0 technologies have been linked to increased connectivity, transparency, and information-processing capabilities within production systems (Perales et al., 2018).

Despite the potential of these approaches, studies indicate that the isolated adoption of Industry 4.0 technologies does not guarantee sustainable performance improvements, particularly when production processes are not sufficiently structured (Dombrowski et al., 2017; Buer et al., 2018). Conversely, although Lean Manufacturing contributes to operational stability and process discipline, it may present limitations when facing increasing complexity, variability, and data intensity in contemporary production environments. These limitations have stimulated growing interest in the integration between Lean Manufacturing and Industry 4.0.

Empirical evidence from a large-scale global survey involving over 1,000 industry practitioners further supports this issue. Hines et al. (2026) found that Industry 4.0 adoption in industrial settings remains largely piecemeal, with Lean and digital technologies often managed by separate teams, while fully integrated Lean–Industry 4.0 systems remain rare in practice. These findings reveal a gap between academic conceptualisations and industrial realities, pointing to the need for approaches that examine the relationships between Lean practices and Industry 4.0 technologies at the system level, rather than through linear or sequential frameworks.

Recent literature suggests that Lean Manufacturing and Industry 4.0 are not conflicting approaches, but potentially complementary (Buer et al., 2018; Jesus et al., 2025). However, there is still no consensus on how Lean practices (LP) and Industry 4.0 (I4.0) technologies relate within production systems, particularly regarding their interdependencies and relative roles throughout the transformation process (Tortorella et al., 2020). While some studies indicate that Lean Manufacturing acts as a foundation for Industry 4.0 technologies (Tortorella and Fettermann, 2018), others suggest that Industry 4.0 technologies may precede or even drive the adoption of Lean practices (Buer et al., 2018). These contrasting views reveal a broader limitation in understanding how Lean Manufacturing practices and I4.0 technologies co-evolve within production systems.

Since early developments in sociotechnical thinking, research has emphasised that technological change and organisational arrangements evolve in close interaction rather than in isolation. Initial contributions in this field examined the implications of new technologies for work activities, tasks, and job design, showing how technical artefacts reshape day-to-day operational practices. However, broader organisational implications related to structural configurations, coordination mechanisms, and organisational forms have received less attention. As a result, the literature has yet to reach consensus on whether organisational structures adjust reactively to technological innovation or whether social and technical elements co-adapt through reciprocal and iterative processes of change (Cimini et al., 2021). In the context of Lean Manufacturing and

Industry 4.0 integration, this limitation appears in linear narratives that treat Lean either as a prerequisite for digitalisation or Industry 4.0 as a technological layer added to existing Lean systems, without sufficiently explaining how social and technical elements interact structurally.

The purpose of this study is therefore to examine the structural interdependencies between Lean Manufacturing practices and Industry 4.0 technologies within sociotechnical production systems.

Against this background, this study seeks to address RQ1: What are the relative and interdependent structural roles of Lean Manufacturing practices and Industry 4.0 technologies within sociotechnical production systems? By adopting a sociotechnical perspective, this study departs from dominant linear narratives that frame Lean Manufacturing as a prerequisite for digitalisation or Industry 4.0 as a purely technology-driven extension of existing practices (Kolberg and Zühlke, 2015; Tortorella and Fettermann, 2018). Instead, Lean practices and Industry 4.0 technologies are conceptualised as interdependent elements within a single production system, whose roles cannot be reduced to temporal sequences or isolated implementation stages. In this context, structural influence, as captured by driving and dependence power, does not imply technological primacy or organisational precedence, but reflects how certain elements shape information flows, coordination mechanisms, and operational routines. To address RQ1, the study employs Interpretive Structural Modelling (ISM) to identify and structure the hierarchical interdependencies between Lean Manufacturing practices and Industry 4.0 technologies, complemented by MICMAC analysis to assess their relative driving and dependence power within the proposed sociotechnical configuration.

The remainder of this paper is organised as follows. Section 2 reviews the literature on Lean Manufacturing, Industry 4.0, sociotechnical systems, and ISM–MICMAC applications. Section 3 describes the methodological procedures, including variable selection, expert panel composition, SSIM construction, reachability matrix development, level partitioning, and MICMAC analysis. Section 4 presents the results. Section 5 discusses the findings, and Section 6 concludes the paper by outlining theoretical contributions, practical implications, limitations, and future research directions.

II. Literature Review

A. Lean Manufacturing and Industry 4.0

Lean Manufacturing has been widely recognised as an approach for structuring production systems through process standardisation, systematic waste elimination, and continuous improvement. By establishing clear operational routines and well-defined workflows, Lean practices influence how activities are coordinated and how information is generated and used within production systems (Dombrowski et al., 2017). For this reason, Lean Manufacturing has been associated with operational stability and improved organisational performance across different industrial contexts.

Industry 4.0, in turn, refers to a broader technological paradigm involving the integration of cyber-physical systems, Internet of Things, cloud computing, and advanced data analytics across industrial and organisational settings. In this study, Industry 4.0 is examined specifically within production systems, where these technologies enable real-time data exchange, connectivity, and decentralised decision-making (Hermann et al., 2016). The terms "Industry 4.0 technologies" and "Industry 4.0" refer, respectively, to specific technological elements and to the broader technological paradigm. They are not treated as equivalent to the wider concept of digitalisation, which encompasses broader socio-organisational transformation processes.

However, several studies argue that adopting such technologies in isolation does not ensure sustainable performance improvements, particularly when production processes are not sufficiently structured (Dombrowski et al., 2017; Buer et al., 2018). In this context, the relationship between Lean Manufacturing and Industry 4.0 has received growing attention. Studies indicate that these approaches are not conflicting, but potentially complementary, as they address different dimensions of production systems (Kolberg and Zühlke, 2015; Buer et al., 2018).

B. Lean Manufacturing and Industry 4.0 as interdependent sociotechnical elements

Although Lean Manufacturing and Industry 4.0 are often described as complementary approaches, their relationship has been interpreted in different ways. One stream of literature suggests that Lean Manufacturing provides a foundation for Industry 4.0 implementation by stabilising processes, reducing waste, standardising routines, and creating the organisational discipline required for digital technologies to generate value (Kolberg and Zühlke, 2015; Buer et al., 2018; Tortorella and Fettermann, 2018). From this perspective, Lean practices help prepare production systems for digitalisation by reducing process variability and improving operational visibility.

Another stream of research indicates that Industry 4.0 technologies can also support, extend, and reshape Lean practices by increasing data availability, enabling real-time monitoring, improving feedback loops, and supporting decentralised decision-making (Dombrowski et al., 2017; Frank et al., 2019; Tortorella et al., 2020). In this sense, digital technologies are not simply tools added after Lean implementation; they may also change how Lean routines are performed, monitored, and improved within production systems.

These contrasting perspectives suggest that the integration between Lean Manufacturing and Industry 4.0 should not be understood as a simple linear sequence. Instead, Lean practices and Industry 4.0 technologies may influence each other through reciprocal relationships: organisational routines shape the use of technologies, while technologies reshape information flows, coordination mechanisms, and improvement practices. This supports the view that Lean–Industry 4.0 integration involves complex interdependencies between social and technical elements within production systems.

C. ISM and MICMAC in operational excellence research

ISM and MICMAC have been widely used in operations management and operational excellence research to structure complex relationships among barriers, enablers, practices, and technologies. In Lean Six Sigma contexts, Singh et al. (2019)

applied ISM and MICMAC to model barriers to LSS implementation in Indian micro, small, and medium enterprises, classifying them according to their driving and dependence power. More recently, Singh et al. (2024) used ISM–MICMAC to analyse strategies for overcoming barriers to LSS implementation in Indian small manufacturing enterprises. These studies show that ISM–MICMAC is useful for examining continuous improvement initiatives involving multiple interdependent organisational, managerial, and operational factors.

In Industry 4.0 research, ISM and MICMAC have also been applied to model adoption barriers and implementation challenges. For example, Kamble et al. (2018b) analysed barriers to Industry 4.0 adoption in Indian manufacturing using ISM and fuzzy MICMAC, while Attiany et al. (2023) examined barriers to Industry 4.0 adoption in supply chains using ISM and MICMAC. These studies demonstrate the relevance of ISM–MICMAC for analysing technology-enabled transformation in operations contexts. They also support its use in this study to examine the structural roles and interdependencies between Lean Manufacturing practices and Industry 4.0 technologies.

III. Methodology

This study adopts a qualitative and structural research approach to examine the integration between Lean Manufacturing practices and Industry 4.0 technologies in production systems (Buer et al., 2018; Kolberg and Zühlke, 2015). This integration is not understood as a linear or purely technical process, but as the result of interactions among organisational structures, work routines, and technological artefacts (Dombrowski et al., 2017; Kolberg and Zühlke, 2015). In line with the sociotechnical tradition, which emphasises the interdependence and co-evolution of social and technical subsystems, this study analyses the structural roles and interdependencies that characterise the relationship between Lean Manufacturing practices and Industry 4.0 technologies (Buer et al., 2018; Tortorella and Fettermann, 2018).

To answer RQ1, the study employs Interpretive Structural Modelling (ISM), complemented by MICMAC analysis, to examine the hierarchical relationships, driving power, dependence power, and structural roles of Lean Manufacturing practices and Industry 4.0 technologies (Singh et al., 2010; Attri et al., 2013). ISM is appropriate for research contexts marked by high relational complexity and limited theoretical consolidation, as it enables the structuring of systems composed of multiple interdependent elements based on informed expert judgements (Singh et al., 2010; Attri et al., 2013).

The ISM analysis was grounded on the judgement of a heterogeneous panel composed of nine specialists with at least five years of experience in production systems, Lean Manufacturing and/or Industry 4.0, as presented in Table I. The panel was designed to combine industrial and academic perspectives. The industrial experts had experience in operational excellence, process engineering, quality management, Lean implementation, production management, and digitalisation-related initiatives in manufacturing contexts. The academic experts contributed knowledge in quality management, Lean and Six Sigma, operations and supply chain management, modelling and simulation, sustainability, and digital servitization. This composition supported a balanced interpretation of the structural relationships between Lean practices and Industry 4.0 technologies.

TABLE I. Composition of the Experts Panel

Expert	Position/Role	Experience	Main Background	Lean experience	Industry 4.0 experience
E1	Cylinder Block Machining Group Manager	8 years	Industry	Lean practices applied to machining operations, shop-floor routines, process standardisation, and production flow improvement	Digital technologies applied to machining and production monitoring
E2	Process Engineering Group Manager	10 years	Industry	Lean-related experience in process engineering, standardisation, continuous improvement, and production process optimisation	Experience with digitalisation-related initiatives in process engineering, production support, process monitoring, and data-based process improvement
E3	Quality and New Product Launch Director	20 years	Industry	Experience in quality management and new product launch processes, including leading Lean projects, standardised work, root cause analysis, problem-solving methods, visual management, continuous improvement, and process stability	Experience with digital support systems related to quality management
E4	Operational Excellence Leader in South America	13 years	Industry	Lean-related experience in operational excellence, waste elimination, problem-solving, efficiency improvement, value stream management, and integration of teams and processes	Experience with Industry 4.0-related initiatives, digitalisation of processes, innovation, decision-making support, and transformation of industrial operations
E5	Lean Manufacturing Specialist	18 years	Industry	Direct experience in Lean Manufacturing implementation, Lean tools, waste reduction, standardised work, continuous improvement, and operational performance improvement	Experience with digital technologies applied to manufacturing
E6	Operational Excellence Analyst	5 years	Industry	Experience in operational improvement, performance indicators, continuous improvement support, and routine standardisation	Experience with digital support for operational excellence activities
E7	Researcher in Quality Management, Lean, and Six Sigma	15 years	Academia	Academic experience in Lean Manufacturing, Six Sigma, quality management, continuous improvement, process improvement, and operational performance	Academic experience in digital transformation, Quality 4.0, operations management, and the relationship between digital technologies, Lean,

Expert	Position/Role	Experience	Main Background	Lean experience	Industry 4.0 experience
					Six Sigma, and quality management
E8	Researcher in Operations and Supply Chain Management; Modelling and Simulation; Sustainability	10 years	Academia	Academic experience in operations and supply chain management, process analysis, modelling and simulation, and sustainability-oriented operational improvement	Academic experience in simulation, operations and supply chain management, and digital approaches applied to production and sustainability-oriented systems
E9	Lean Process Analyst / Researcher in Digital Servitization	8 years	Industry/Academia	Lean-related experience in process analysis, Lean routines, continuous improvement, operational improvement, and integration between process improvement and service-oriented operations	Experience in digital servitization, digital transformation research, Industry 4.0-related topics, and the integration between process improvement and digital service-oriented operations

The industrial profile of the panel reflects predominantly discrete manufacturing contexts, with expertise concentrated in automotive and mechanical manufacturing sectors. These sectors are well-established domains for both Lean Manufacturing and Industry 4.0 adoption, providing a relevant basis for the structural judgements underlying the ISM model. This sectoral focus represents a contextual boundary condition that should be considered when interpreting the results.

Since Lean Manufacturing is generally more mature and widely established in organisations than Industry 4.0, the study considered the possibility that expert assessments could be influenced by differences in familiarity, maturity, or current managerial emphasis. Industry 4.0 technologies could be perceived as more strategically relevant due to their recent prominence, while Lean practices could be perceived as more familiar due to their longer history of implementation. To reduce this potential bias, experts were instructed to evaluate structural influence among the variables rather than perceived importance, maturity level, novelty, or chronological order of adoption. Therefore, driving power and dependence power were interpreted as indicators of systemic influence within the sociotechnical production system, not as measures of priority, technological superiority, or implementation sequence.

The number of experts is consistent with prior ISM applications in operations management, which emphasise depth of expertise and consensus building rather than large sample sizes (Ahmad and Qahmash, 2021). In this study, ISM is used to organise Lean Manufacturing practices and Industry 4.0 technologies identified from the literature into a hierarchical structure that reveals relationships of precedence, dependence, and mutual influence (Singh et al., 2010; Attri et al., 2013). MICMAC analysis is then applied as a complementary step to examine the driving power and dependence power of the identified elements, allowing their classification into autonomous, dependent, linkage, and independent variables (Attri et al., 2013; Singh et al., 2010). This combination provides a structured interpretation of the systemic roles played by Lean practices and Industry 4.0 technologies within the proposed framework. The overall methodological procedure is illustrated in Figure 1.

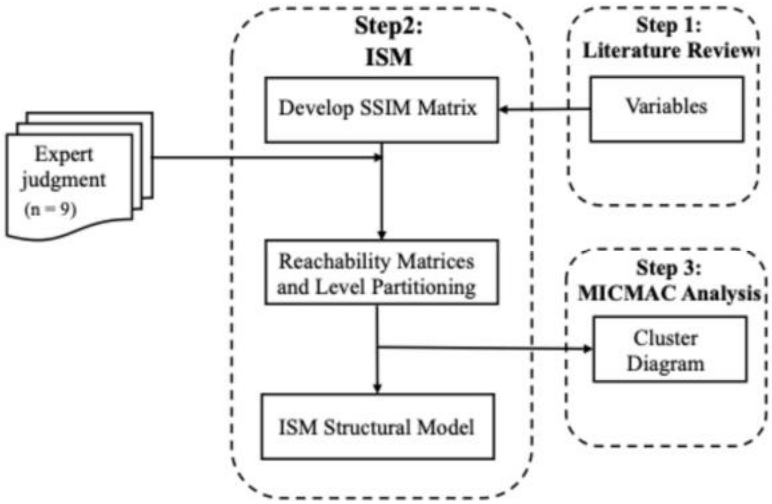


Fig. 1. Methodological framework of the ISM–MICMAC approach

In this study, driving power and dependence power are interpreted as indicators of structural influence within the sociotechnical system, rather than as indicators of temporal precedence or implementation sequence. Accordingly, elements with high driving power are not assumed to be adopted earlier. Instead, they are understood as playing a central role in configuring system-level interdependencies, information flows, coordination mechanisms, and operational routines within production systems.

The selection of variables was based on established literature on Lean Manufacturing and Industry 4.0 technologies. The Lean Manufacturing practices were mainly derived from the literature review conducted by Bhamu and Sangwan (2014), who identified a broad set of Lean practices discussed in previous studies. From this broader set, the practices included in this study were selected according to three criteria: recurrence in the Lean Manufacturing literature, relevance to production systems, and suitability for relational analysis through ISM. The Industry 4.0 technologies were selected based on recognised classifications of Industry 4.0 technologies, particularly those proposed by Rüßmann et al. (2015), Hermann et al. (2016), and Frank et al. (2019), which emphasise technologies related to connectivity, data processing, cyber-physical integration, and digital support for production systems. This procedure ensured that the selected variables represented both the organisational-practice dimension of Lean Manufacturing and the technological dimension of Industry 4.0.

The Lean Manufacturing practices analysed comprise standardised work (LP1), total productive maintenance (LP2), multifunctional teams and employee involvement (LP3), just-in-time (LP4), poka-yoke (LP5), continuous improvement (LP6), production smoothing (LP7), value stream mapping (LP8), visual control (LP9), and kanban (LP10). In parallel, the Industry 4.0 technologies included in the analysis encompass autonomous robots (IT1), simulation (IT2), system integration (IT3), Internet of Things (IT4), cloud computing (IT5), augmented reality (IT6), big data and analytics (IT7), and cybersecurity (IT8). These practices and technologies are treated as interrelated elements of a single sociotechnical production system, whose interactions shape information visibility, coordination mechanisms, and operational routines.

In total, the study considered 18 variables, comprising 10 Lean Manufacturing practices and 8 Industry 4.0 technologies. Since ISM is based on pairwise comparisons among the selected elements, 153 pairs were formed for expert evaluation, calculated as $n(n-1)/2$, where n represents the total number of variables. All variables were presented within the same evaluation instrument, allowing experts to assess the relationships from a systemic perspective and to consider the full set of Lean practices and Industry 4.0 technologies simultaneously.

The application of ISM followed the classical procedural steps of the method (Singh et al., 2010; Attri et al., 2013). Data collection from experts was conducted using a structured Excel-based instrument composed of three worksheets. The first worksheet presented the study objective, completion instructions, and a brief explanation of the ISM method. The second worksheet contained the pairwise comparison matrix used to construct the Structural Self-Interaction Matrix (SSIM). To reduce filling errors, the cells were locked, and experts were allowed to select only one of the four standard ISM symbols: V, A, X, or O. Brief comments were also provided to define each Lean Manufacturing practice and Industry 4.0 technology. The third worksheet allowed experts to comment on the clarity of the instrument and the completion process. This step was used as a quality control procedure, since responses indicating uncertainty or lack of clarity were excluded from the analysis. The instrument was sent to 20 specialists, of whom 12 returned completed responses. Three responses were excluded because they were incomplete or because the experts reported uncertainty regarding the completion of the instrument. Therefore, the final SSIM was based on nine valid expert responses.

Initially, a Structural Self-Interaction Matrix (SSIM) was constructed to capture the influence relationships among pairs of elements. The SSIM was populated using standard ISM symbols: V indicates that element i influences element j; A indicates that element j influences element i; X denotes mutual influence between elements i and j; and O indicates the absence of a direct relationship. All possible directions of influence were considered, making it possible to identify not only driving and dependent elements, but also mutually reinforcing relationships between Lean practices and Industry 4.0 technologies (Singh et al., 2010; Attri et al., 2013).

The SSIM was then transformed into an initial binary reachability matrix by systematically replacing the qualitative symbols with binary values. The principle of transitivity was subsequently applied, resulting in a final reachability matrix that incorporates indirect relationships logically inferred from the direct ones. The ISM data were processed using Microsoft Excel, without dedicated ISM software. The conversion of the SSIM into the initial reachability matrix followed standard ISM rules, and the final reachability matrix was obtained by applying the transitivity principle. Transitive relationships were evaluated by the authors and marked with an asterisk in the final reachability matrix, distinguishing indirect relationships from direct expert judgements. No additional consultation round was conducted after the identification of transitive relationships, since these links were treated as logical implications derived from the initial SSIM rather than as new direct judgements. This procedure preserved the interpretive basis of the expert evaluations while ensuring consistency with the formal logic of the ISM method.

Based on this matrix, an iterative level-partitioning procedure was conducted, leading to the development of the interpretive structural model (Singh et al., 2010; Attri et al., 2013). This model provides a graphical representation of the relative positioning of Lean Manufacturing practices and Industry 4.0 technologies within the analysed system.

Following the ISM analysis, MICMAC was applied using the final reachability matrix to classify the elements according to their driving power and dependence power (Attri et al., 2013). The outcomes of the ISM and MICMAC analyses provide the foundation for the subsequent section, in which the structural model is presented and discussed, with emphasis on the sociotechnical roles of Lean Manufacturing practices and Industry 4.0 technologies in their integration.

IV. Results

The results are organised into two subsections. The first presents the ISM results, including the Structural Self-Interaction Matrix (SSIM), the reachability matrices, the level partitioning, and the structural model. The second presents the MICMAC results, including the classification of the analysed elements according to their driving and dependence power.

A. ISM results

The Structural Self-Interaction Matrix (SSIM) was built using the most frequent responses provided by the nine experts. The mode of the expert judgements was adopted to represent the final relationship between each pair of variables. To illustrate

the interpretation of the SSIM, the relationship between LP1 and IT8 is represented by "V", indicating that most experts considered LP1 to influence IT8. The final SSIM is presented in Table II.

TABLE II. FINAL STRUCTURAL SELF-INTERACTION MATRIX (SSIM)

	IT8	IT7	IT6	IT5	IT4	IT3	IT2	IT1	LP10	LP9	LP8	LP7	LP6	LP5	LP4	LP3	LP2	LP1
LP1	V	A	A	A	A	A	A	A	X	A	X	V	X	X	V	X	X	
LP2	O	A	A	A	A	A	A	A	X	A	X	X	X	X	V	A		
LP3	V	V	V	A	A	X	V	O	X	A	X	V	X	V	V			
LP4	O	O	O	A	A	A	A	A	V	A	A	X	A	A				
LP5	O	O	A	O	X	A	A	A	O	A	O	V	V					
LP6	O	X	A	A	A	X	A	X	X	A	X	V						
LP7	O	O	O	A	A	A	A	A	A	A	A							
LP8	O	A	O	A	A	A	A	O	X	V								
LP9	O	A	A	O	A	A	O	O	X									
LP10	O	A	O	A	A	A	A	A										
IT1	A	X	O	A	A	X	X											
IT1	O	X	X	A	X	A												
IT3	A	X	X	A	X													
IT4	A	X	V	X														
IT5	X	X	V															
IT6	A	A																
IT7	X																	
IT8																		

The SSIM was converted into the Initial Reachability Matrix, presented in Table III, by applying the standard ISM conversion rules. In this procedure, a V in position (i, j) indicates that the corresponding entry in the reachability matrix is assigned as 1 for (i, j) and 0 for (j, i). An A indicates 0 for (i, j) and 1 for (j, i). An X indicates 1 for both (i, j) and (j, i), whereas an O indicates 0 for both positions (Attri, Dev and Sharma, 2013). Subsequently, the Final Reachability Matrix was obtained by incorporating the transitivity property and is presented in Table IV. In a transitivity relationship, if element "A" influences element "B" and element "B" influences element "C," it can be inferred that "A" indirectly influences "C". The indirect influence of variables is represented in the final reachability matrix by 1* (Attri et al., 2013).

TABLE III. INITIAL REACHABILITY MATRIX

	LP1	LP2	LP3	LP4	LP5	LP6	LP7	LP8	LP9	LP10	IT1	IT2	IT3	IT4	IT5	IT6	IT7	IT8
LP1	1	1	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0
LP2	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
LP3	0	0	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0
LP4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LP5	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
LP6	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
LP7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
LP8	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
LP9	1	1	0	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0
LP10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
IT1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
IT2	1	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0
IT3	0	0	0	1	0	0	1	1	1	1	1	0	1	0	0	1	0	0
IT4	0	0	0	0	1	1	1	0	1	0	0	0	0	1	0	1	1	0
IT5	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0
IT6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
IT7	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0
IT8	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1

TABLE IV. FINAL REACHABILITY MATRIX WITH DRIVING AND DEPENDENCE POWER

	LP1	LP2	LP3	LP4	LP5	LP6	LP7	LP8	LP9	LP10	IT1	IT2	IT3	IT4	IT5	IT6	IT7	IT8	DRI POW
LP1	1	1	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	**6**
LP2	0	1	0	1	0	1	1*	0	0	0	0	0	0	0	0	0	0	0	**4**
LP3	0	0	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	**6**
LP4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**
LP5	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	**2**
LP6	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	**3**
LP7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	**1**
LP8	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	**3**
LP9	1	1	0	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0	**7**
LP10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	**1**
IT1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	**2**
IT2	1	1*	0	1	0	1	1	0	0	1*	0	1	0	0	0	0	0	0	**7**
IT3	1*	1*	0	1	0	1*	1	1	1	1	1	0	1	0	0	1	0	0	**11**
IT4	1*	1*	0	1*	1	1	1	0	1	1*	1*	1*	1*	1	0	1	1	0	**14**
IT5	1*	0	0	1*	0	1*	1*	0	0	0	1	1	1*	0	1	1	1	0	**10**
IT6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	**1**
IT7	1*	0	0	1*	0	1*	1*	1*	1*	1*	1	1	1	0	0	1*	1	0	**12**
IT8	0	0	0	0	0	0	1*	0	0	0	1	1*	1*	0	0	1	1	1	**7**
DEP POW	**7**	**6**	**1**	**12**	**2**	**10**	**15**	**4**	**4**	**8**	**6**	**5**	**5**	**1**	**1**	**6**	**4**	**1**	

Based on the Final Reachability Matrix, a level partitioning procedure was first performed to identify the hierarchical levels of the analysed elements, as presented in Figure 2. These levels then provided the basis for constructing the ISM structural model, presented in Figure 3.

Variables	Reachability set	Antecedent set	Intersection set	Level
LP1	LP1, LP9, IT2, IT3, IT4, IT5, IT7	LP1, LP2, LP4, LP6, LP7, LP10	LP1	IV
LP2	LP1, LP2, LP9, IT2, IT3, IT4	LP2, LP4, LP6, LP7	LP2	III
LP3	LP3	LP3, LP4, LP6, LP7, LP8, LP10	LP3	III
LP4	LP1, LP2, LP3, LP4, LP6, LP8, LP9, IT2, IT3, IT4, IT5, IT7	LP4	LP4	I
LP5	LP5, IT4	LP5, LP7	LP5	II
LP6	LP1, LP2, LP3, LP6, LP9, IT2, IT3, IT4, IT5, IT7	LP4, LP6, LP7	LP6	II
LP7	LP1, LP2, LP3, LP5, LP6, LP7, LP8, LP9, IT1, IT2, IT3, IT4, IT5, IT7, IT8	LP7	LP7	I
LP8	LP3, LP8, IT3, IT7	LP4, LP7, LP8	LP8	II
LP9	LP9, IT3, IT4, IT7	LP1, LP2, LP4, LP6, LP7, LP9, LP10	LP9	V
LP10	LP1, LP3, LP9, LP10, IT2, IT3, IT4, IT7	LP10	LP10	I
IT1	IT1, IT3, IT4, IT5, IT7, IT8	LP7, IT1	IT1	II
IT2	IT2, IT4, IT5, IT7, IT8	LP1, LP2, LP4, LP6, LP7, LP10, IT2	IT2	V
IT3	IT3, IT4, IT5, IT7, IT8	LP1, LP2, LP4, LP6, LP7, LP8, LP9, LP10, IT1, IT3, IT6	IT3	VI
IT4	IT4	LP1, LP2, LP4, LP5, LP6, LP7, LP9, LP10, IT1, IT2, IT3, IT4, IT6, IT7	IT4	VIII
IT5	IT5	LP1, LP4, LP6, LP7, IT1, IT2, IT3, IT5, IT6, IT7	IT5	VIII
IT6	IT3, IT4, IT5, IT6, IT7, IT8	IT6	IT6	I
IT7	IT4, IT5, IT7, IT8	LP1, LP4, LP6, LP7, LP8, LP9, LP10, IT1, IT2, IT3, IT6, IT7	IT7	VII
IT8	IT8	LP7, IT1, IT2, IT3, IT6, IT7, IT8	IT8	VIII

Fig. 2. Level partitioning of the ISM model

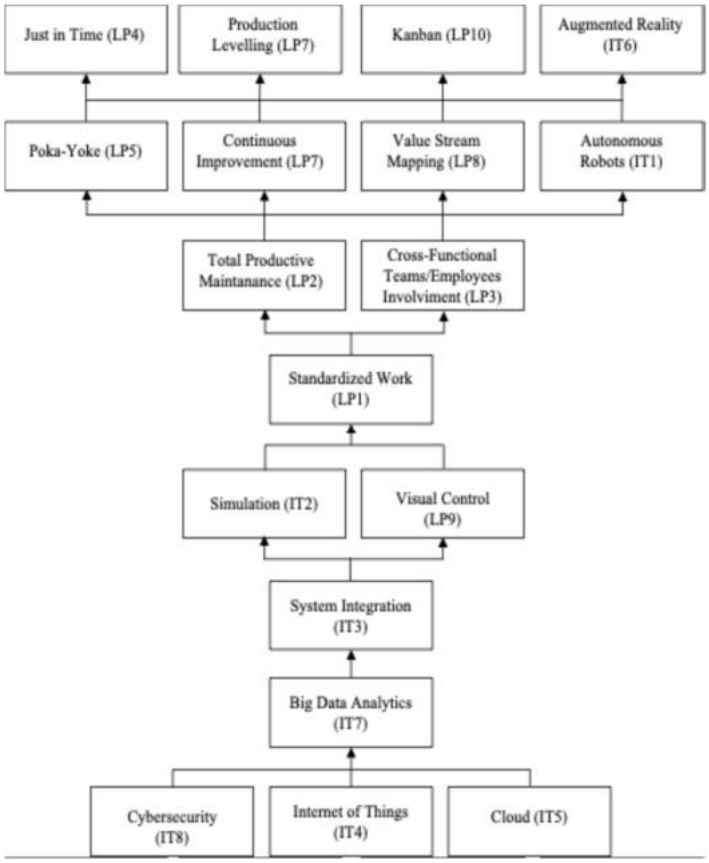


Fig. 3. ISM structural model

B. MICMAC results

MICMAC analysis was conducted using the Final Reachability Matrix presented in Table IV, which incorporates both direct and transitive relationships among the analysed elements. Driving power was calculated as the row sum of each element, while dependence power was calculated as the column sum. These values were used to classify the elements into four groups: autonomous variables, dependent variables, linkage variables, and independent variables. The resulting classification is presented in Figure 4.

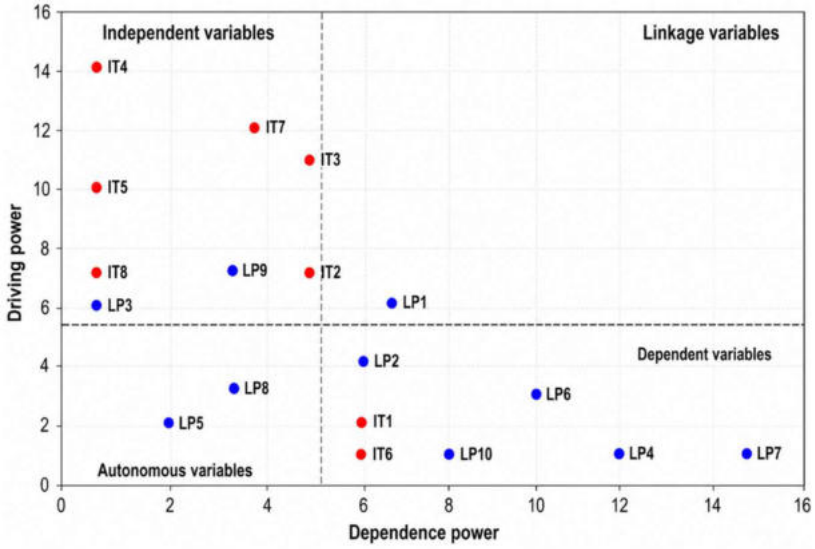


Fig. 4. MICMAC classification of Lean Manufacturing practices and Industry 4.0 technologies according to driving and dependence power

V. DISCUSSION

The results provide important insights into the integration between Lean Manufacturing practices and Industry 4.0 technologies from a sociotechnical perspective. The ISM model and MICMAC analysis reveal a non-linear and interdependent configuration, in which Industry 4.0 technologies are mostly positioned at lower hierarchical levels and show greater driving power. Lean Manufacturing practices, in turn, are mainly positioned at intermediate or upper levels, indicating greater dependence. This differentiation suggests that technologies and practices do not play equivalent roles in the system but contribute through distinct functions that shape integration dynamics. At first glance, this finding may seem to contrast with studies that frame Lean Manufacturing as a prerequisite for Industry 4.0 implementation (Buer et al., 2018). However, rather than invalidating these studies, the results suggest that sequential and structural perspectives capture different aspects of the same sociotechnical system.

From a sociotechnical standpoint, the position of elements in the ISM hierarchy does not necessarily indicate temporal precedence or conceptual primacy. Instead, it reflects their role in shaping system-level relationships. In this sense, the position of Industry 4.0 technologies in the model indicates that they act as important configurators of contemporary production systems, reshaping information flows, decision-making processes, and coordination mechanisms. These technologies mainly operate within the technical subsystem by expanding information visibility, enabling real-time feedback, and increasing the system's capacity for data-driven control (Frank et al., 2019; Beier et al., 2020). As a result, they influence how Lean practices are enacted, monitored, and reinforced within operational routines.

The literature that conceptualises Lean Manufacturing as a foundation for Industry 4.0 usually emphasises the importance of stable, standardised, and waste-free processes as prerequisites for successful digitalisation (Buer et al., 2018). This argument remains valid and is not contradicted by the results of this study. Lean practices continue to play a critical enabling role by providing organisational discipline, process clarity, and problem-solving routines required for digital technologies to generate value (Dombrowski et al., 2017). Within the sociotechnical architecture, these practices are mainly associated with the social subsystem, as they structure routines, define roles, and stabilise interaction patterns among workers and teams. However, the findings suggest that once digital technologies are introduced, they also influence the evolution and reconfiguration of Lean practices themselves.

This apparent inversion in the ISM–MICMAC results can be explained by recognising that Industry 4.0 technologies change the sociotechnical architecture of production systems, especially in relation to information visibility, feedback loops, and real-time control (Frank et al., 2019). Advanced data analytics, cyber-physical systems, and integrated information platforms do not merely support existing Lean routines; they reshape how performance is measured, how abnormalities are detected, and how continuous improvement activities are prioritised. By changing the speed, detail, and reliability of information flows, these technologies alter the conditions under which social routines operate. Consequently, they influence behavioural responses, coordination patterns, and managerial attention within the operational environment in which Lean practices are embedded (Kolberg and Zühlke, 2015; Frank et al., 2019).

The intermediate positioning of several Lean Manufacturing practices in the ISM hierarchy also highlights their role as mediators between technological capabilities and organisational routines. Practices such as standardisation, visual

management, and continuous improvement function as translation mechanisms that convert digital information into actionable operational responses (Shah and Ward, 2007). They enable technical signals generated by digital systems to be incorporated into shared understandings, collective sense-making, and coordinated action on the shop floor. Rather than serving only as antecedents to digitalisation, these practices coevolve with technologies, adapting their form and function in response to greater data availability and system connectivity (Buer et al., 2018).

The MICMAC analysis indicates that most elements are embedded in the system, with only a limited number of autonomous variables. Specifically, LP5 and LP8 appear in the autonomous quadrant, suggesting relatively low driving and dependence power compared with the other analysed elements. In contrast, most Industry 4.0 technologies and some Lean Manufacturing practices occupy the independent, dependent, or linkage quadrants, reinforcing the interpretation of Lean–Industry 4.0 integration as a strongly interconnected sociotechnical system.

The results also help clarify inconsistencies observed in prior empirical studies. While some authors argue that Lean Manufacturing must precede Industry 4.0 (Tortorella and Fettermann, 2018), others suggest that digital technologies can act as catalysts for Lean adoption (Frank et al., 2019). The structural configuration identified in this study indicates that both interpretations can coexist, depending on the stage of system evolution and the level of analysis adopted. From a structural perspective, preparatory Lean practices may stabilise the social subsystem before digitalisation. At the same time, subsequent technology deployment can reconfigure these practices by introducing new forms of control, feedback, and coordination.

Therefore, Lean practices may be necessary to prepare the organisational context for digitalisation, but once technologies are deployed, they can become powerful drivers that reshape the paths of Lean implementation. Rather than reflecting a linear implementation logic, the ISM–MICMAC results suggest that integration unfolds through reciprocal adjustments between technical infrastructures and social routines, which jointly shape operational outcomes over time. This perspective reconciles seemingly contradictory findings in the literature and provides a more nuanced foundation for future empirical and theoretical research on digital transformation in production systems.

From a theoretical perspective, this study contributes by reframing Lean Manufacturing and Industry 4.0 integration as a structural configuration problem, highlighting the differentiated roles of practices and technologies within sociotechnical production systems.

VI. Conclusion

This study advances the understanding of Lean Manufacturing and Industry 4.0 integration by conceptualising their relationship as an interdependent sociotechnical configuration rather than a linear or sequential transformation path. By combining Interpretive Structural Modelling (ISM) with MICMAC analysis, the research provides a systemic interpretation of how organisational practices and digital technologies jointly shape contemporary production systems.

The findings reconcile competing views in the literature by showing that Lean Manufacturing and Industry 4.0 assume distinct but mutually conditioning roles. Lean practices contribute organisational discipline, process stability, and operational routines, while Industry 4.0 technologies reshape information flows, coordination mechanisms, and feedback structures. This does not imply technological primacy, but indicates that digital technologies act as configurators within evolving sociotechnical systems.

From a methodological perspective, the study demonstrates the usefulness of integrating ISM and MICMAC to examine complex and under-theorised phenomena in operations management. From a managerial perspective, the results suggest that organisations should avoid rigid implementation sequences and adopt integrated approaches that recognise the reciprocal adjustment between Lean practices and Industry 4.0 technologies over time.

The findings are particularly relevant to Quality 4.0 and Operational Excellence initiatives, which increasingly seek to combine advanced digital technologies with Lean and continuous improvement practices. By clarifying the structural roles of Lean Manufacturing practices and Industry 4.0 technologies, this study provides insights for organisations pursuing integrated quality and operational improvement strategies.

This study has limitations. First, the expert panel was drawn from specific industrial and academic contexts, which may affect the generalisability of the structural model to other sectors or geographic settings. Second, ISM captures a static representation of interdependencies and does not fully account for the dynamic co-evolution of Lean practices and Industry 4.0 technologies over time. Third, the predefined set of practices and technologies reflects current literature coverage and may not represent all production contexts. Future research should address these limitations through longitudinal studies, sector-specific case studies, and quantitative validation of the proposed structural model.

Acknowledgment

This work was carried out with the support of the National Council for Scientific and Technological Development (CNPq) - process no. 160823/2022-6. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001

References

Ahmad, R. and Soberi, M.S.F. (2018) Changeover process improvement based on modified SMED method and other process improvement tools application: An improvement project of 5-axis CNC machine operation in advanced composite manufacturing industry. The International Journal of Advanced Manufacturing Technology, 94(1), pp. 433–450.

Attri, R., Dev, N. and Sharma, V. (2013) Interpretive structural modelling (ISM) approach: an overview. Research Journal of Management Sciences, 2(2), pp. 1–8.

Attiany, M.S., Al-Abdallah, G.M., Salameh, A.A. and Bani-Melhem, S. (2023) Barriers to adopt Industry 4.0 in supply chains using interpretive structural modeling. Uncertain Supply Chain Management, 11, pp. 299–306.

Beier, G., Ullrich, A., Niehoff, S., Reißig, M. and Habich, M. (2020) Industry 4.0: How it is defined from a sociotechnical perspective and how much sustainability it includes – A literature review. Journal of Cleaner Production, 259, 120856.

Bhamu, J.; Sangwan, K. S. Lean Manufacturing: Literature review and research issues. International Journal of Operations and Production Management, v. 34, n. 7, p. 876–940, 2014.

Buer, S.-V., Strandhagen, J.O. and Chan, F.T.S. (2018) The link between Industry 4.0 and Lean Manufacturing: Mapping current research and establishing a research agenda. International Journal of Production Research, 56(8), pp. 2924–2940.

Cimini, C., Pinto, R., Pezzotta, G. and Gaiardelli, P. (2021) How do Industry 4.0 technologies influence organisational change? An empirical analysis of Italian SMEs. Journal of Manufacturing Technology Management, 32(3), pp. 695–721.

Dombrowski, U., Richter, T. and Krenkel, P. (2017) Interdependencies of Industrie 4.0 and Lean Production Systems: A use case analysis. Procedia Manufacturing, 11, pp. 1061–1068.

Frank, A.G., Dalenogare, L.S. and Ayala, N.F. (2019) Industry 4.0 technologies: Implementation patterns in manufacturing companies. International Journal of Production Economics, 210, pp. 15–26.

Hines, P. et al., (2026) A global industrial perspective on lean industry 4.0: a qualitative wide-angle lens approach. Production Planning & Control, 37(3), pp. 225-246.

Hermann, M., Pentek, T. and Otto, B. (2016) Design principles for Industrie 4.0 scenarios. In: Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, pp. 3928–3937.

Jesus, C., Silva, R., Pereira, M. and Costa, E. (2025) Impact of adopting Industry 4.0 practices and technologies on time and motion study: A pilot study revealing efficiency gains in a multi-operation workstation. Production & Manufacturing Research, 13(1), 2476436.

Kamble, S.S., Gunasekaran, A. and Gawankar, S.A. (2018a) Sustainable Industry 4.0 framework: A systematic literature review identifying current trends and future perspectives. Process Safety and Environmental Protection, 117, pp. 408–425.

Kamble, S.S., Gunasekaran, A. and Sharma, R. (2018b) Analysis of the driving and dependence power of barriers to adopt Industry 4.0 in Indian manufacturing industry. Computers in Industry, 101, pp. 107–119.

Kolberg, D. and Zühlke, D. (2015) Lean automation enabled by Industry 4.0 technologies. IFAC-PapersOnLine, 48(3), pp. 1870–1875.

Perales, D.P., Valero, F.A. and Boza García, A. (2018) Industry 4.0: A classification scheme. In: Viles, E. et al. (eds.) Closing the Gap Between Practice and Research in Industrial Engineering. Cham: Springer, pp. 343–350.

Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P. and Harnisch, M. (2015) 'Industry 4.0: The future of productivity and growth in manufacturing industries', Boston Consulting Group, 9(1), pp. 54–89.

Sasso, R.A., Filho, M.G. and Ganga, G.M.D. (2025) Synergizing lean management and circular economy: Pathways to sustainable manufacturing. Corporate Social Responsibility and Environmental Management, 32(1), pp. 543–562.

Shah, R. and Ward, P.T. (2007) Defining and developing measures of lean production. Journal of Operations Management, 25(4), pp. 785–805.

Singh, M., Rathi, R. and Garza-Reyes, J.A. (2019) Modelling the barriers of Lean Six Sigma for Indian micro-small medium enterprises: An ISM and MICMAC approach. The TQM Journal, 32(2), pp. 313–335.

Singh, M., Kaswan, M.S.S. and Rathi, R. (2024) Empirical analysis of strategies to overcome the barriers of LSS implementation in Indian small manufacturing enterprises: an application of ISM–MICMAC. International Journal of Lean Six Sigma, 15(4), pp. 886–905.

Tortorella, G.L. and Fettermann, D.C. (2018) Implementation of Industry 4.0 and lean production in Brazilian manufacturing companies. International Journal of Production Research, 56(8), pp. 2975–2987.

Tortorella, G.L., Frank, A.G., Fettermann, D.C. and Sawhney, R. (2020) Towards the proposition of a lean automation framework: Integrating Industry 4.0 into lean production. Journal of Manufacturing Technology Management, 31(3), pp. 615–640.

Wang, S., Wan, J., Li, D. and Zhang, C. (2016) Implementing smart factory of Industrie 4.0: An outlook. International Journal of Distributed Sensor Networks, 12(1), pp. 1–10.

Zhu, Y., Shang, Z. and Wu, J. (2025) Enhanced tool detection in Industry 4.0 via deep learning-augmented human intent recognition: Introducing the Industry-RetinaNet model. Traitement du Signal, 40(4), pp. 1723–1734.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Use of Lean Tools to Propose Improvements in the Transport Management Process of a Public Higher Education Institution

Marcelo Rodrigues [1], Fabiane Lizarelli [1], João Alfineto [1] and Gilberto Ganga [1]

[1] Federal University of São Carlos, Brazil; fabiane@dep.ufscar.br

***Abstract*— This study investigates the application of Lean principles and tools to improve the transportation management process of a public higher education institution (HEI). The research aims to analyze how Lean practices, particularly Value Stream Mapping (VSM) and the 5W1H action plan, can identify waste and improve operational efficiency in administrative service processes. The study adopts a qualitative and applied case study procedure. The unit of analysis is the transportation department of the Federal University of Mato Grosso (Universidade Federal de Mato Grosso UFMT), Sinop campus, Brazil. Data was collected from institutional documents, digital records, internal regulations, and process observations. A hybrid Business Process Model and Notation – Value Stream Mapping (BPMN–VSM) model was used to map and analyze the transportation service process. Process performance was evaluated through indicators such as processing time, waiting time, necessary non-value-added activities, and total lead time. The analysis of the current-state process revealed several inefficiencies, including overprocessing, unnecessary movement, and long waiting times. Improvement proposals were developed by integrating activities, improving communication, and adopting digital tools to streamline information flow. The future-state BPMN–VSM suggests potential improvements, including a reduction in processing time, necessary non-value-added time, and overall lead time. The proposed changes demonstrate the potential of Lean practices to improve administrative service processes in public HEIs. This study contributes to the literature on Lean in services and higher education by examining the application of Lean tools to transportation management processes in a public HEI—an area that has received limited attention in previous research. Additionally, the study provides practical insights for university managers seeking to improve operational efficiency in administrative processes through Lean methodologies.**

Keywords – Lean, Value Stream Mapping, Higher Education Institution, Public.

I. Introduction

Lean has brought benefits and improvements to several sectors, as it is not limited to a toolbox for quality improvement but rather represents a philosophy and a set of principles aimed at cultural change within organizations (Antony et al., 2012; Klein et al., 2024). Lean philosophy requires a transformation in people's mindset, while practices and tools serve as mechanisms to translate these principles into concrete actions (Bhasin and Burcher, 2006; Gupta et al., 2016; Lima et al., 2023). The integration of principles, practices, and tools facilitates efficiency improvements across different productive sectors, particularly in services (Danese et al., 2018).

The service sector differs from manufacturing due to characteristics such as intangibility, heterogeneity, inseparability, simultaneity, and perishability (Lovelock and Gummesson, 2004; Gupta et al., 2016). Services can be defined as acts, processes, and performances offered or co-produced by one entity for another (Zeithaml et al., 2014). Despite the importance of the service sector to the global economy—representing more than 50% of GDP in most countries—the literature on Lean in services remains limited compared to other sectors (Danese et al., 2018; Fenner and Netland, 2023).

Lean in service contexts, often referred to as Lean Service, can be defined as a multidimensional approach focused on eliminating waste through continuous improvement in service operations and processes (Gupta et al., 2016; Fenner and Netland, 2023). Unlike traditional Lean principles developed for manufacturing, Lean in services emphasizes adapting Lean methodologies to the specific characteristics and requirements of service organizations. This adaptation involves process standardization, the use of Lean tools and techniques, and cultural changes aimed at continuous improvement and waste reduction in service delivery (Gupta et al., 2016; Vadivel et al., 2022).

In recent decades, Lean principles and practices have increasingly been adopted in both private and public service organizations (Balzer et al., 2016; Klein et al., 2024). Public organizations have sought to improve and refine procedures and tasks to deliver more efficient and effective services to citizens (Biazzi et al., 2011; Cano et al., 2022). Lean has proven suitable for improving performance and fostering a culture of continuous improvement in this context, generating innovative solutions and simplifying processes in public organizations (Radnor and Boaden, 2008; Cano et al., 2022). Applications of Lean in the service sector include areas such as healthcare, security, justice, social assistance, and education (Antony et al., 2017).

Within higher education institutions (HEIs), Lean can support improvements aimed at meeting the expectations of students, staff, and other stakeholders (Balzer et al., 2016; Lima et al., 2023). Its implementation can help reduce costs in contexts of increasing expenses and limited financial resources, meet public accountability requirements regarding efficiency and effectiveness, and strategically leverage institutional resources to fulfill teaching, research, and extension missions (Balzer et al., 2016; Antony et al., 2012; Zighan and El-Qasem, 2021; Mundth et al., 2020; Lima et al., 2023).

However, Lean implementation in HEIs also faces significant challenges. These include limited leadership commitment, difficulties in establishing and evaluating quality, and low staff engagement (Antony, 2020; Klein et al., 2024). Additionally, adapting terminology and concepts originally developed for manufacturing to the higher education context requires creative thinking and specific training (Antony, 2020; Lima et al., 2023). Lean implementation in HEIs should therefore be understood as a gradual and multifaceted process that requires strong support from top management, effective communication across organizational levels, and integrated planning of improvement initiatives (Antony, 2020; Antony et al., 2012; Cross et al., 2022; Gastelum-Acosta et al., 2024).

Despite these challenges, Lean presents significant potential for application in public HEIs (Klein et al., 2021). Among the benefits reported are cost reduction, better information control, improved use of physical space, increased productivity, and shorter service processing times (Lima et al., 2023). Nevertheless, further research is still needed on the implementation of Lean practices in HEIs and their impact on organizational performance (Klein et al., 2023). However, although interest in Lean applications in higher education has increased, the literature remains dominated by conceptual discussions and isolated case studies, with relatively few empirical investigations of specific administrative support processes (Malmbrandt and Åhlström, 2011; Gómez-Molina and Moyano-Fuentes, 2022).

In particular, transportation management services have received little attention despite their importance in supporting teaching, research, and extension activities in public HEIs. Furthermore, limited evidence exists regarding the integration of Business Process Model and Notation (BPMN) and Value Stream Mapping (VSM) to analyze and improve administrative processes in this context. This study addresses these gaps by examining the application of Lean principles to a transp(ortation management process in a public HEI through an integrated BPMN–VSM approach. Furthermore, although Value Stream Mapping has been applied in some HEI contexts (Waterbury, 2010; Mundth et al., 2020; Lima et al., 2023), limited evidence exists regarding its integration with BPMN to analyze and improve administrative support processes in higher education settings. Therefore, further empirical research is needed to expand the understanding of how Lean principles and tools can be adapted to underexplored support services in public higher education institutions.

In this context, Lean practices and tools can be applied to several processes within public HEIs, including transportation management. Transportation services in HEIs generally involve the coordination of people, resources, and time, which could benefit from Lean practices. To date, no studies were identified addressing Lean implementation in transportation management process within public HEIs. Therefore, this study aims to answer the following research question: What benefits can Lean tools, principles, and practices generate when applied to the transportation management process of a public higher education institution? Accordingly, the objective of this study is to present a case of Lean application and its potential benefits in the transportation management process of a public HEI.

The remainder of this paper is structured as follows. Section II presents the literature review, discussing Lean applications in public higher education institutions and the main Lean tools used in administrative process improvement. Section III describes the research methodology, including the case study context, data collection procedures, and analytical approach. Section IV presents the results of the BPMN–VSM analysis, including the current-state process, the identification of inefficiencies, the proposed improvements and future-state mapping, and the implementation action plan. Finally, Section V discusses the conclusions, research limitations, and directions for future studies.

II. Literature Review

A. Lean applied to the Public Higher Education sector

Lean philosophy (also known as Lean thinking, Lean manufacturing, or the Toyota Production System) emerged in Japan in the 1950s and has since been widely adopted worldwide (Holweg, 2007; Gil-Vilda et al., 2021). Initially focused on manufacturing, Lean later expanded to other contexts such as services, commerce, and public administration (Womack and Jones, 2003; Leite and Vieira, 2015). The publication of The Machine That Changed the World highlighted the efficiency of Japanese automotive production and popularized the concept of Lean Production (Comm and Mathaisel, 2005; Holweg, 2007). Over time, Lean became a management philosophy focused on distinguishing value-adding from non-value-adding activities to minimize waste and improve flow and productivity (Seth et al., 2017).

The adaptation of Lean principles to service contexts led to the concept of Lean Service, which aims to eliminate waste, improve efficiency, reduce lead time, and enhance service quality while focusing on value delivery to customers (Bowen and Youngdahl, 1998; Vadivel et al., 2022; Fenner and Netland, 2023). However, Lean principles cannot be applied directly, since services differ from manufacturing due to characteristics such as intangibility, heterogeneity, and customer participation in service delivery (Gupta et al., 2016). In addition, employee engagement and respect for people are critical elements in Lean Service, as service quality depends strongly on interactions between employees and customers (Gupta et al., 2016).

Lean thinking can also be applied to non-manufacturing and information-based processes (McManus, 2003; Amin et al., 2024). In these contexts, value streams consist mainly of information and knowledge flows rather than physical materials, making waste identification more complex but still feasible and beneficial (Zighan and El-Qasem, 2021).

Within public Higher Education Institutions (HEIs), Lean can be applied to a wide range of administrative and academic processes, including student services, hiring, procurement, curriculum development, research activities, and resource management (Balzer et al., 2016; Antony et al., 2012; Zighan and El-Qasem, 2021; Mundth et al., 2020; Lima et al., 2023). Studies report benefits such as cost reduction, improved information management, better use of physical space, and increased productivity (Lima et al., 2023). Successful implementation requires leadership support,

employee engagement, a focus on internal customers, and a culture of continuous improvement (Klein et al., 2023, 2024; Antony et al., 2012; Gómez-Molina and Moyano-Fuentes, 2022).

However, Lean implementation in HEIs also faces several challenges, including lack of visionary leadership, difficulty defining and measuring quality, resistance to change, and limited understanding of Lean benefits (Antony, 2020; Douglas et al., 2015; Klein et al., 2024). Traditional academic management practices often focus on individual work, whereas Lean emphasizes teamwork and project-based continuous improvement (Antony, 2020). For this reason, understanding the application methods, tools, and potential of Lean in HEIs can help other HEI managers succeed in its implementation.

B. Lean tools applied in HEIs

Lean philosophy relies on several tools designed to eliminate waste and optimize processes (Lima et al., 2023). Among the tools commonly applied in HEIs are Value Stream Mapping (VSM) and the 5W1H action plan. These tools aim to improve efficiency, reduce variability and waste, and increase stakeholder satisfaction.

Previous studies have demonstrated the potential of Value Stream Mapping (VSM) in higher education institutions (HEIs). For example, Zighan and El-Qasem (2021) applied VSM to analyze the curriculum of an academic program in a Jordanian HEI, identifying value-adding and non-value-adding activities and improving graduates' employability. Similarly, Mundth et al. (2020) applied VSM in the human resources department of a public HEI to support continuous improvement initiatives.

VSM is one of the most widely used Lean tools for process improvement due to its adaptability across different sectors. In recent years, it has been applied in areas such as transportation, construction, healthcare, services, software development, innovation management, mining, manufacturing, food production, maintenance, and educational services (Lima et al., 2023). The tool provides a visual representation of the flow of materials and information in the delivery of products or services within an organization. By mapping these flows, VSM helps identify inefficiencies and sources of waste, enabling organizations to redesign processes and improve operational efficiency (Tapping and Shuker, 2018).

In administrative environments, Lean principles aim to reduce costs, minimize rework, improve communication, and eliminate non-value-adding activities that hinder productivity (Tapping and Shuker, 2018). According to Nallusamy (2016), the VSM approach typically involves two stages: the current-state map, which documents existing activities and measures cycle time and lead time, and the future-state map, which is developed after critical analysis to propose improvement actions and eliminate identified waste. In this sense, VSM serves as a strategic starting point for Lean transformation initiatives by providing detailed process insights that support managerial decision-making and continuous improvement (Lima et al., 2023).Another commonly used tool is the 5W1H action plan, which helps structure improvement initiatives by defining what should be done, why, when, where, how, and by whom. This approach facilitates project planning, clarifies responsibilities, and supports effective implementation of improvement actions (Neves et al., 2018).

III. RESEARCH METHOD

This study adopts a qualitative and applied research approach. In terms of objectives, it is classified as exploratory, and regarding procedures, it follows a case study approach (Sunder et al., 2020; Belhadi et al., 2023). Case study research provides flexibility in research design and application, allowing the exploration of complex organizational phenomena through qualitative and quantitative evidence (Sunder et al., 2020; Belhadi et al., 2023). It is an empirical method that investigates a phenomenon within its real-life context through an in-depth analysis of one or more units of analysis, enabling a closer interaction between theory and empirical evidence (Yin, 2009).

Following Voss et al. (2002), Eisenhardt (1989), and Yin (2009), this study adopts an exploratory case study approach with a theory extension and refinement purpose. The case study method is appropriate because it enables an in-depth investigation of a contemporary organizational phenomenon within its real-life context. Rather than testing a predefined theory, the study seeks to expand the understanding of Lean applications in public higher education institutions by examining a transportation management process that has received limited attention in the literature. In addition, the study contributes methodologically through the integrated application of BPMN and Value Stream Mapping (VSM) for analyzing administrative service processes in a public HEI context.

In the context of this study, the case study method has been considered appropriate for examining the application and effectiveness of Lean methodology and for reporting the experiences and lessons learned from its implementation in public HEIs (Antony et al., 2018). This approach combines multiple sources of qualitative and quantitative data, such as institutional records, process times, observations, and discussions, providing measurable operational evidence grounded in real-life contexts (Seth et al., 2017). For this research, the unit of analysis is the transportation sector of a public higher education institution, described in the following section.

A. Unit of Analysis

The Transportation Department is located at the Sinop Campus of the Federal University of Mato Grosso (Universidade Federal do Mato Grosso - UFMT). UFMT is a Brazilian federal public higher education institution established as a foundation by Law No. 5,647 of December 10, 1970, with headquarters in Cuiabá and additional campuses in Várzea Grande, Araguaia, and Sinop.

The Transportation Department operates under the Campus Administration and is responsible for managing the official fleet of the Sinop Campus. The fleet currently consists of 19 vehicles, including pickup trucks, buses, passenger

cars, a minibus, a truck, and a van. According to Resolution No. 396 of the National Traffic Council (CONTRAN), vehicles are classified as light vehicles (gross weight up to 3,500 kg) and heavy vehicles (above 3,500 kg). Based on this classification, the department manages 12 light vehicles and 7 heavy vehicles.

The department's main responsibilities include providing institutional transportation services, ensuring fleet maintenance programs, guiding users regarding regulations and procedures, and supervising service contracts related to drivers, fleet maintenance, and fueling. This study focuses on the most demanded process within the department: institutional transportation services, which support field classes, technical visits, administrative travel, and research and extension activities. This process involves several actors who perform essential roles in fleet operation and service execution, as summarized in Table I.

TABLE I. ACTORS AND THEIR ROLES IN THE TRANSPORTATION DEPARTMENT PROCESSES (UFMT)

Process actors	Description
Transportation Department	Linked to the Campus Administration and composed of two technical staff members and four outsourced drivers.
Drivers	Linked to the department through a contracted company that provides driver services.
Driver – Administrative Supervisor	Works as a driver and administrative supervisor, monitoring driver activities and organizing travel records.
Driver – Vehicle Maintenance Supervisor	Works as a driver and maintenance supervisor, responsible for identifying maintenance needs and monitoring fleet maintenance activities.
Requesters	Faculty members, administrative staff, and students enrolled in undergraduate or graduate programs who request transportation services.
Contracted Company (vehicle maintenance)	Company responsible for managing maintenance services for official vehicles through a contractual agreement with the university.
Accredited Workshop	Workshop authorized by the contracted maintenance company to perform vehicle repair and maintenance services.

B. Data collection procedures

Data were collected from multiple sources of evidence, including administrative documents from the Transportation Department, digital records stored in the institutional electronic system (Sistema Eletrônico de Informações - SEI), internal regulations, and contracts with outsourced companies. The use of multiple sources enabled data triangulation, increasing the reliability of the results and reducing potential bias (Yin, 2009). To analyze the selected processes, a hybrid Business Process Model and Notation – Value Stream Mapping (BPMN–VSM) model adapted from Santhiapillai and Ratnayake (2024) was used to represent the process flows of the Transportation Department. This integration allows the identification of bottlenecks and waste, supporting proposals for process improvement. The graphical models were developed using Bizagi Modeler.

The analysis followed four steps. First, the current-state VSM was developed to map the institutional transportation processe and collect activity time data. The BPMN–VSM integration was operationalized by first using BPMN to model the sequence of activities, decision points, actors, and information flows involved in the transportation management process. Subsequently, VSM metrics were incorporated into the BPMN model, allowing the measurement of processing times, waiting times, necessary non-value-added activities, and total lead time for each process step. This integrated representation enabled a simultaneous analysis of process logic and operational performance (Santhiapillai and Ratnayake, 2024).

The indicators shown in Table II were used to evaluate process performance (Tapping and Shuker, 2018; Santhiapillai and Ratnayake, 2024). Activities were classified based on Lean principles. Value-added activities were defined as those that directly contributed to fulfilling the transportation service requested by the user. Activities that did not directly create value but were required due to administrative, regulatory, or operational requirements were classified as necessary non-value-added activities (NNVA). Waiting periods and delays between activities were considered waste and were incorporated into the waiting time analysis.

TABLE II. QUANTITATIVE INDICATORS USED IN THE VSM ANALYSIS

Indicator	*Definition*	*Unit*
Processing Time (PT)	Time required to complete a process or activity.	Hours (h)
Waiting Time (WT)	Time a task waits before the next activity begins.	Hours (h)
Necessary Non-Value-Added Activities (NNVA)	Time spent on required activities that do not add value.	Hours (h)
Total Lead Time	Sum of processing and waiting times in the value stream.	Hours (h)

Second step, the analysis of the current-state map allowed the identification of problems and non-value-adding activities, such as waiting times and difficulties in accessing vehicle information. The current-state process mapping was supported by documentary analysis and individual semi-structured interviews with six Transportation Department employees, including two administrative staff members (the Administrative Supervisor and the Vehicle Maintenance

Supervisor) and four drivers. Each interview lasted approximately 1.5 to 2 hours and focused on understanding the activities performed, process flows, responsibilities, operational challenges, and improvement opportunities.

The PT, WT, and NVA indicators were calculated using the average processing times obtained from 10 transportation service requests conducted between June and November 2024. The process times were identified through documentary analysis, direct examination of service records, and validation with Transportation Department employees during the interviews and group meetings. All times were expressed in working hours, considering an 8-hour workday consistent with the workload of the employees involved in the process. This approach focuses on the effective working time required to perform each activity, excluding periods of inactivity, and provides a consistent basis for identifying bottlenecks and improvement opportunities.

Following the development of the BPMN-VSM current-state map, four group meetings were conducted with the Transportation Department team. The first meeting was dedicated to validating the current-state map and confirming the accuracy of the process representation. Third, future-state VSMs were developed to propose improvements aimed at reducing waste and increasing efficiency. The remaining three meetings focused on presenting, discussing, refining, and prioritizing the proposed improvement opportunities. These meetings enabled the identification and validation of bottlenecks, non-value-added activities, and feasible improvement actions, supporting the development of the future-state map and the corresponding action plan. Finally, a 5W1H action plan was designed to support the implementation of the proposed improvements.

IV. RESULTS

The results are structured into four subsections, which present the analysis of the current-state process, the identification of inefficiencies, the development of improvement proposals and the future-state BPMN–VSM, and the implementation action plan.

A. Analysis of the institutional transportation services process in its current state

Table III summarizes the activities of the institutional transportation service process. The process begins when the requester submits a travel request through the Electronic Information System (SEI), providing information such as origin, destination, date, time, number of passengers, and purpose of the trip. The Transportation Sector of UFMT then reviews the request, verifying the required documentation and assessing its feasibility according to institutional regulations.

Next, the availability of drivers and vehicles is checked, and the administrative supervisor assigns the appropriate resources. The requester is notified of the confirmation through the SEI system. For trips outside the municipality of Sinop-MT, driver travel allowances must be requested in advance. After the trip, the driver collects passengers' signatures, and the documents are digitized and attached to the SEI process, concluding the service request. Fig. 1 presents the current-state map based on BPMN-VSM.

TABLE III. ACTIVITIES OF THE INSTITUTIONAL TRANSPORT SERVICE PROCESS – CURRENT STATE

Activity	Description	Actor	Processing Time (h)	Waiting Time (h)	NNVA Time (h)[a]
1. Submit transport request	Request submitted through the system with required documents	Requester	–	2.00	–
2. Analyze request	Verify documents and signatures	Technical staff	0.13	–	0.13
3. Check transport availability	Check driver and vehicle availability in the schedule	Technical staff	0.08	–	0.08
4. Print transport request form	Print travel documents and deliver to supervisor	Technical staff	0.50	–	0.50
5. Assign driver and vehicle	Identify available driver and vehicle	Driver / Administrative supervisor	4.00	0.25	4.25
6. Record trip information in the schedule	Insert trip information into scheduling spreadsheet	Technical staff	0.08	–	0.08
7. Request driver allowance (per diem)	Check need for driver allowance and request it	Technical staff	0.16	–	0.16
8. Confirm transport service	Send confirmation to requester through SEI	Technical staff	0.08	–	0.08
9. Post transport requests on the drivers' board	Display printed requests in drivers' room	Driver / Administrative supervisor	0.08	16.00	16.08
10. Provide transport service	Perform the trip and collect passengers' signatures	Driver	6.00	1.00	1.00
11. Check trip report	Return signed documents and report service completion	Driver / Administrative supervisor	0.08	–	0.08
12. Attach documents in the system	Scan and attach documents to SEI process	Technical staff	0.08	–	0.08
13. Close transport request	Send final confirmation and close the process	Technical staff	0.08	–	0.08
Total			11,35 h	19,25 h	22,6h
Lead time (Processing Time+Waiting time)			***30,6 h***		

[a.] **NNVA:** (Necessary Non-Value-Added Activities - processing time of activities that do not add value but are necessary.

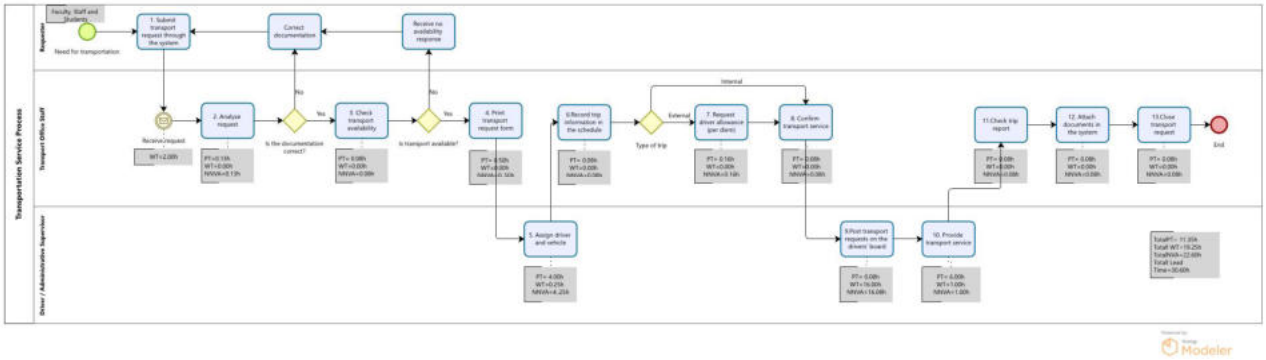


Fig. 1. Business Process Model and Notation – Value Stream Mapping (BPMN-VSM) Current State. Processing Time(PT): time required to perform the activity. Waiting Time(WT): time between activities. Necessary Non-Value-Added Time(NAV): processing time of activities that do not add value but are necessary.

B. Identification of problems in the activity flow

Four monthly meetings with Transport Department staff, drivers, and users were conducted to review and validated the current-state BPMN–VSM and identify bottlenecks and improvement opportunities. The analysis showed that nine of the thirteen activities involve waste, such as overprocessing, unnecessary motion, and waiting (Table IV).

Errors in Submit transport request may affect subsequent stages. Print transport request form may generate delays due to physical document handling, while Attach trip report to the process can be affected by pending actions in previous steps, such as collecting signatures. Additionally, Record trip information in the schedule may be delayed due to communication failures or manual updates. A significant issue occurs in Post transport requests on the drivers' board, where the processing time is 0.08 hours but the waiting time reaches 16 hours due to the contractual requirement to request driver per diem at least 16 working hours before trips outside the municipality. In Assign driver and vehicle, the processing time reaches 4 hours because driver and vehicle availability must be verified and communicated to technical staff. The lack of standardization in vehicle verification contributes to this delay.

The time analysis shows 11.35 hours of processing time (PT), 19.25 hours of waiting time (WT), and 22.6 hours of necessary non-value-added time (NNAV). As a result, the lead time reaches 30.6 hours, with a significant portion associated with waiting and administrative requirements typical of public universities (Comm and Mathaisel, 2005).

C. Improvement proposals and future-state BPMN–VSM

The improvement proposals (Table IV) aim to reduce waste and streamline the institutional transport service process. For Analyze transport request, the dissemination of guidelines and internal communications is proposed to reduce errors in Submit transport request, while also integrating the verification of the need to Request driver per diem, which was previously performed in a separate activity. This integration is expected to reduce overprocessing and improve process flow. In Check driver and vehicle availability, it is proposed that the Transport Office Staff also Record trip information in the schedule, avoiding duplicate data entry. For Print transport request form, placing a printer in the drivers' room and sharing documents digitally with the supervisor would reduce unnecessary movement.

The activity Assign driver and vehicle could be improved by sharing access to the scheduling spreadsheet with the Driver / Administrative Supervisor and implementing periodic vehicle inspections through a standardized inspection manual, ensuring updated information about fleet conditions. Additionally, Post transport requests on the drivers' board could be performed together with Assign driver and vehicle, reducing redundancy. Finally, Deliver signed documents could be scheduled on a fixed day of the week to improve process organization. Overall, these improvements aim to increase efficiency, reduce processing time, and eliminate waste, in line with Lean principles.

TABLE IV. IMPROVEMENT PROPOSALS FOR THE INSTITUTIONAL TRANSPORT SERVICE PROCESS

Activity	*Lean waste*	*Improvement proposal*
1. Submit transport request	–	–
2. Analyze transport request	Overprocessing	Disseminate internal guidelines to reduce errors in transport requests; integrate Request driver per diem verification into this activity (activity 7).
3. Check transport availability	Overprocessing	Integrate Record trip information (activity 6). in the schedule into this activity to avoid duplication.
4. Print transport request form	Unnecessary motion	Install a printer in the drivers' room and share documents digitally with the supervisor.
5. Assign driver and vehicle	Overprocessing / unnecessary motion	Share access to the scheduling spreadsheet with the Driver / Administrative Supervisor; implement periodic vehicle inspections using a checklist; integrate activity 9 into this step.
6. Record trip information in the schedule	Overprocessing	Integrate this activity with Check driver and vehicle availability (3).

Activity	*Lean waste*	*Improvement proposal*
7. Request driver allowance (per diem)	Overprocessing	Integrate the verification of per diem needs with Analyze transport request (2); keep this activity only for submitting the request.
8. Confirm transport service	–	–
9. Post transport requests on the drivers' board	Overprocessing	Integrate this activity with Assign driver and vehicle (5).
10. Provide transport service	–	–
11. Check trip report	Unnecessary motion	Establish a fixed day for document delivery and verify trip issues with the supervisor.
12. Attach documents in the system	–	–
13. Close transport request	Overprocessing	Integrate this activity with Attach trip report to the process (12).

Fig. 2 presents the proposed future-state BPMN–VSM. A comparison between the current-state BPMN–VSM (Fig. 1) and the future-state BPMN–VSM (Fig. 2) shows several proposed improvements aimed at reducing waste and processing time in the institutional transport service process. The main changes include the integration of activities, such as merging Analyze transport request (2) with Request driver per diem (7), and adjustments in the sequence of tasks, including Record trip information in the schedule before Assign driver and vehicle. It is important to note that the future-state map represents a proposed improvement scenario and the performance gains associated with the future-state process should be interpreted as projected outcomes based on the identified Lean improvement opportunities rather than empirically validated results obtained after implementation.

In addition, the Administrative Supervisor would centralize some activities, while digital tools would support communication and information flow. These improvements are projected to reduce processing time (PT) from 11.35 to 6.98 hours, waiting time (WT) from 19.25 to 19 hours, and necessary non-value-added time (NNVA) from 22.6 to 17.98 hours, resulting in a reduction of lead time from 30.6 to 25.98 hours, demonstrating the efficiency potential of the proposed changes.

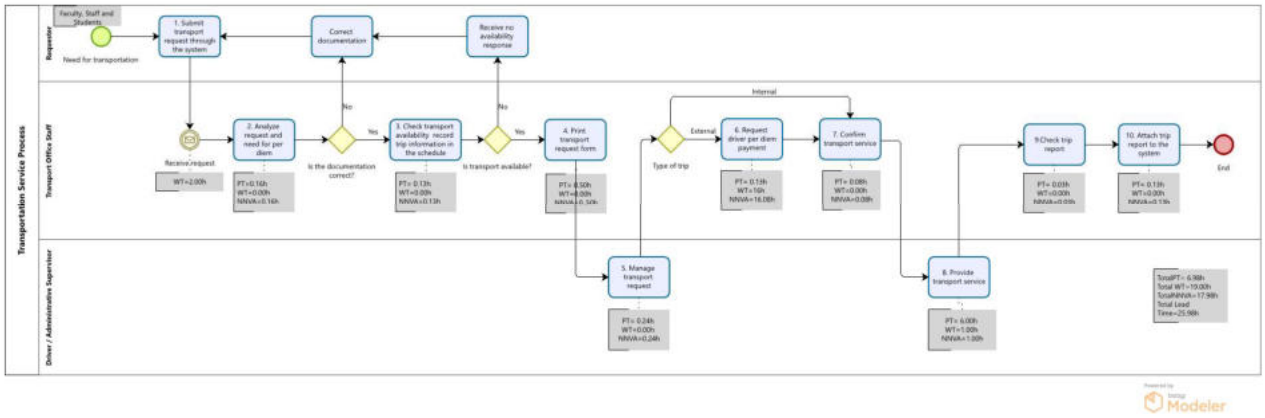


Fig. 2. Business Process Model and Notation – Value Stream Mapping (BPMN-VSM) Future State. Processing Time(PT): time required to perform the activity. Waiting Time(WT): time between activities. Necessary Non-Value-Added Time(NAV): processing time of activities that do not add value but are necessary.

D. Action plan for implementing the institutional transport service process improvements

Action plans were developed using the 5W1H technique to implement the improvement proposals identified in the previous steps. Table V presents the action plan designed to optimize the institutional transport service process in the Transport Department of UFMT.

The first action involves team training to prepare staff for the new practices and tools. From March 2025, improvement actions start to be implemented to reduce errors, eliminate redundancies, and streamline the process flow, including direct recording of information in the scheduling spreadsheet, digital document handling, and integration of activities related to driver and vehicle assignment. These initiatives aim to increase efficiency and reduce waste in the process.

TABLE V. TABLE TYPE STYLES

What	**Why**	**How**	**When**	**Who**	**Where**
1. Team training	Prepare staff for new practices and tools	Conduct training sessions and share guidelines and digital tools	After contract renewal	Technical staff	Transport Office / Drivers' room
2. Analyze transport request and verify per diem requirement	Reduce requester errors and overprocessing	Disseminate internal guidelines and integrate per diem verification into the request analysis	After contract renewal	Technical staff	Transport Office

What	Why	How	When	Who	Where
3. Check driver and vehicle availability and record trip information in the schedule	Eliminate duplication and overprocessing	Technical staff directly records trip details in the scheduling spreadsheet	After contract renewal	Technical staff	Transport Office
4. Print transport request form	Reduce unnecessary motion and processing time	Install a printer in the drivers' room and send documents digitally to the supervisor	After contract renewal	Technical staff / Administrative Supervisor	Drivers' room
5. Assign driver and vehicle	Avoid overprocessing and unnecessary movement	Share access to the scheduling spreadsheet and use vehicle inspection checklists	After contract renewal	Administrative Supervisor / Maintenance Supervisor	Drivers' room

V. Conclusions

This study investigated the application of Lean principles and tools in the transportation management process of a public higher education institution. Using a BPMN-based value stream mapping approach, the analysis identified several sources of waste affecting the efficiency of the institutional transport service process. Based on these findings, improvement proposals were developed to streamline the process and reduce unnecessary activities. The comparison between the current-state and the proposed future-state BPMN–VSM indicated the potential benefits of Lean implementation. The proposed improvements include integrating activities, reducing redundant steps, improving communication, and adopting digital tools to support information flow. These changes may reduce processing time, eliminate unnecessary movements, and simplify administrative procedures.

Beyond the potential operational improvements demonstrated in the case organization, this study contributes to knowledge creation by documenting how Lean principles can be operationalized in administrative transportation services within a public higher education institution. The findings highlight recurring sources of waste, coordination challenges, and improvement mechanisms that may also be relevant in similar public-sector service environments. Therefore, the study provides analytical insights that can support both future research and Lean implementation efforts in comparable organizational contexts.

From a theoretical perspective, this study extends the understanding of Lean applications in public higher education institutions by demonstrating how Lean principles can be adapted to administrative service processes. The findings contribute to the growing body of literature on Lean in higher education by identifying specific sources of waste and improvement mechanisms in transportation management services. Furthermore, the study illustrates the value of integrating BPMN and VSM for the analysis of administrative processes, providing methodological insights for future research in public-sector and higher education contexts.

From a managerial perspective, the results suggest that Lean practices can help public HEIs improve operational efficiency even in highly regulated administrative environments. Process mapping and structured improvement plans, such as 5W1H, provide managers with practical tools to identify inefficiencies and implement improvements using existing resources. From an academic perspective, this study contributes to the growing literature on Lean in services and higher education by exploring a process that has received limited attention in previous studies, namely transportation management in public HEIs. The findings reinforce the potential of Lean approaches to improve service processes in public organizations.

However, this study has some limitations. First, it is based on a single case study conducted in one public university campus, which may limit the generalizability of the results. Second, the future-state process represents a projection of expected improvements, as the proposed changes have not yet been fully implemented or empirically validated. Future research could expand this investigation by analyzing Lean implementation in other administrative processes within HEIs, conducting comparative studies across institutions, or evaluating the actual performance impacts after the implementation of the proposed improvements. Overall, the findings highlight that Lean can support the improvement of administrative and operational processes in public HEIs, contributing to more efficient resource use and improved service delivery.

References

Amin, N.N.H.M., Wahab, A.N.A., Elias, N.F., Jenal, R., Jambak, M.I. and Ashril, N.A.N.M. (2024) 'A Lean service conceptual model for digital transformation in the competitive service industry', *International Journal of Advanced Computer Science and Applications*, 15(1).

Antony, J. (2020) 'Challenges in the deployment of LSS in the higher education sector: viewpoints from leading academics and practitioners', in Antony, J. *Lean Six Sigma in Higher Education: A Practical Guide for Continuous Improvement Professionals in Higher Education*. Bingley: Emerald Publishing Limited. https://doi.org/10.1108/978-1-78769-929-820201005.

Antony, J., Krishan, N., Cullen, D. and Kumar, M. (2012) 'Lean Six Sigma for higher education institutions (HEIs): challenges, barriers, success factors and tools/techniques', *International Journal of Productivity and Performance Management*, 61(8), pp. 940–948.

Antony, J., Rodgers, B. and Cudney, E.A. (2017) 'Lean Six Sigma for public sector organizations: is it a myth or reality?', International Journal of Quality & Reliability Management, 34(9), pp. 1402–1411.

Antony, J., Ghadge, A., Ashby, S.A. and Cudney, E.A. (2018) 'Lean Six Sigma journey in a UK higher education institute: a case study', International Journal of Quality & Reliability Management, 35(2), pp. 510–526.

Balzer, W.K., Francis, D.E., Krehbiel, T.C. and Shea, N. (2016) 'A review and perspective on Lean in higher education', Quality Assurance in Education, 24(4), pp. 442–462.

Bhasin, S. and Burcher, P. (2006) 'Lean viewed as a philosophy', *Journal of Manufacturing Technology Management*, 17(1), pp. 56–72.

Belhadi, A., Kamble, S.S., Gunasekaran, A., Zkik, K., M, D.K. and Touriki, F.E. (2023) 'A Big Data analytics-driven Lean Six Sigma framework for enhanced green performance: a case study of a chemical company', *Production Planning & Control*, 34(9), pp. 767–790.

Biazzi, M.R., Muscat, A.R.N. and Biazzi, J.L. (2011) 'Modelo de aperfeiçoamento de processos em instituições públicas de ensino superior', *Gestão & Produção*, 18, pp. 869–880.

Bowen, D.E. and Youngdahl, W.E. (1998) 'Lean service: in defense of a production-line approach', *International Journal of Service Industry Management*, 9(3), pp. 207–225.

Cano, M., Murray, R. and Kourouklis, A. (2022) 'Can Lean management change the managerial culture in higher education?', *Studies in Higher Education*, 47(4), pp. 915–927.

Comm, C.L. and Mathaisel, D.F. (2005) 'An exploratory study of best Lean sustainability practices in higher education', *Quality Assurance in Education*, 13(3), pp. 227–240.

Cross, J., Joshi, M. and Jensen, P. (2022) 'A preliminary framework for assessing Lean implementation in higher education', *International Journal of Productivity and Performance Management*, 73(1), pp. 210–241.

Danese, P., Manfè, V. and Romano, P. (2018) 'A systematic literature review on recent Lean research: state-of-the-art and future directions', *International Journal of Management Reviews*, 20(2), pp. 579–605.

Douglas, J., Antony, J. and Douglas, A. (2015) 'Waste identification and elimination in HEIs: the role of Lean thinking', *International Journal of Quality & Reliability Management*, 32(9), pp. 970–981.

Eisenhardt, K.M. (1989) 'Building theories from case study research', *Academy of Management Review*, 14(4), pp. 532–550.

Fenner, S. and Netland, T. (2023) 'Lean service: a contingency perspective', *Operations Management Research*, 16(3), pp. 1271–1289.

Gastelum-Acosta, C. et al. (2024) 'Modeling critical success factors of Lean Six Sigma in higher education institutions', *International Journal of Lean Six Sigma*, 15(2), pp. 326–346.

Gil-Vilda, F., Yaguee-Fabra, J.A. and Sunyer, A. (2021) 'From Lean production to Lean 4.0: a systematic literature review with a historical perspective', *Applied Sciences*, 11(21), 10318.

Gómez-Molina, D.L. and Moyano-Fuentes, J. (2022) 'Lean management in universities: a systematic literature review', *International Journal of Lean Six Sigma*, 13(1), pp. 156–177.

Gupta, S., Sharma, M. and Sunder, M.V. (2016) 'Lean services: a systematic review', *International Journal of Productivity and Performance Management*, 65(8), pp. 1025–1056.

Holweg, M. (2007) 'The genealogy of Lean production', *Journal of Operations Management*, 25(2), pp. 420–437.

Klein, L.L., Moyano-Fuentes, J., Vieira, K.M. and Marçal, D.R. (2024) 'An exploratory study of the relationships between Lean practices and team performance in higher education', *International Journal of Lean Six Sigma*, 15(2), pp. 372–399.

Klein, L.L., Vieira, K.M., Marçal, D.R. and Pereira, J.R.L. (2023) 'Lean management practices perception and their influence on organizational performance in a public higher education institution', *The TQM Journal*, 35(3), pp. 673–697.

Klein, L.L., Tonetto, M.S., Avila, L.V. and Moreira, R. (2021) 'Management of Lean waste in a public higher education institution', *Journal of Cleaner Production*, 286, 125386.

Leite, H.R. and Vieira, G.E. (2015) 'Lean philosophy and its applications in the service industry: a review of the current knowledge', *Production*, 25(3), pp. 529–541.

Lima, E.S., Oliveira, U.R., Carvalho Costa, M., Fernandes, V.A. and Teodoro, P. (2023) 'Sustainability in public universities through Lean evaluation and future improvement for administrative processes', *Journal of Cleaner Production*, 382, 135318.

Lovelock, C. and Gummesson, E. (2004) 'Whither services marketing? In search of a new paradigm and fresh perspectives', *Journal of Service Research*, 7(1), pp. 20–41.

Malmbrandt, M. and Åhlström, P. (2013) 'An instrument for assessing Lean service adoption', *International Journal of Operations & Production Management*, 33(9), pp. 1131–1165.

Mundth, G., Mothersell, W.M., Motwani, J. and Moore, M.L. (2020) 'The adoption of Lean principles and practices: transforming human resources delivery system model', *International Journal of Business Excellence*, 20(1), pp. 32–50.

Neves, P., Silva, F.J.G., Ferreira, L.P., Pereira, T., Gouveia, A. and Pimentel, C. (2018) 'Implementing Lean tools in the manufacturing process of trimmings products', *Procedia Manufacturing*, 17, pp. 696–704.

Nallusamy, S. (2016) 'Lean manufacturing implementation in a gear shaft manufacturing company using value stream mapping', *International Journal of Engineering Research in Africa*, 21, pp. 231–237.

Radnor, Z. and Boaden, R. (2008) 'Lean in public services — panacea or paradox?', *Public Money & Management*, 28(1), pp. 3–7.

Santhiapillai, F.P. and Ratnayake, R.M.C. (2024) 'BPMN-VSM-based knowledge-work process analysis and improvement in police services: a case study', *International Journal of Productivity and Performance Management*, 73(7), pp. 2116–2135.

Seth, D., Seth, N. and Dhariwal, P. (2017) 'Application of value stream mapping (VSM) for Lean and cycle time reduction in complex production environments: a case study', *Production Planning & Control*, 28(5), pp. 398–419.

Sunder, M.V., Mahalingam, S. and Krishna, M.S.N. (2020) 'Improving patients' satisfaction in a mobile hospital using Lean Six Sigma: a design-thinking intervention', *Production Planning & Control*, 31(6), pp. 512–526.

Tapping, D. and Shuker, T. (2018) *Value Stream Management for the Lean Office: Eight Steps to Planning, Mapping, and Sustaining Lean Improvements in Administrative Areas*. Boca Raton: CRC Press.

Vadivel, S.M., Sequeira, A.H., Sakkariyas, R.R. and Boobalan, K. (2022) 'Impact of Lean service, workplace environment, and social practices on the operational performance of India Post service industry', *Annals of Operations Research*, 315(2), pp. 2219–2244.

Voss, C., Tsikriktsis, N. and Frohlich, M. (2002) 'Case research in operations management', *International Journal of Operations & Production Management*, 22(2), pp. 195–219.

Waterbury, T. (2010) 'Using the collaborative improvement model to improve a university hiring process', *International Journal of Productivity and Quality Management*, 5(1), pp. 75–87.

Womack, J.P. and Jones, D.T. (2003) *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*. New York: Free Press.

Yin, R.K. (2009) *Case Study Research: Design and Methods*. Thousand Oaks: Sage.

Zeithaml, V.A., Bitner, M.J. and Gremler, D.D. (2014) *Services Marketing: Integrating Customer Focus Across the Firm*. New York: McGraw-Hill.

Zighan, S. and El-Qasem, A. (2021) 'Lean thinking and higher education management: revaluing the business school programme management', *International Journal of Productivity and Performance Management*, 70(3), pp. 675–703.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

The impact of ISO 9001 implementation on transport and logistics processes

Andreea Madalina Alexandru [1] [0009-0000-1725-1621], Irina Severin 1 [0000-0002-0296-4730]
National University of Science and Technology Politehnica Bucharest, Romania
[1]andreea.alexandru98@stud.fiir.upb.ro irina.severin@upb.ro

Abstract*—Quality management has become a key factor influencing the performance of logistics and transport organizations, where the pressure of on-time delivery, quality delivery and operational consistency are critical competitive elements. Although ISO 9001 is widely implemented across the sector, there are various challenges in adapting its general requirements to the specific operational realities of transport processes. This paper explores this gap by analysing the ISO 9001:2015 framework and its relevance for logistics activities such as transport planning, resource management, operational control and performance monitoring. The paper is based on a documentary selection of selected logistics companies, including both global suppliers and European carriers, with the aim of identifying how ISO 9001 is adopted and positioned in practice. Analysis of publicly available data shows that ISO 9001 is consistently implemented alongside other management system standards, supporting integrated approaches that combine quality, environmental, safety and supply chain requirements. To complement this perspective, a small number of written responses were obtained from transport organizations. These responses show that ISO 9001 is not only used for certification, but also as a practical tool that helps organize processes, monitor performance and support continuous improvement in daily logistics operations. Overall, the paper provides an exploratory perspective on how ISO 9001 can be interpreted as an operational reference in the field of logistics and transportation, while recognizing the reliance on publicly available information and the limited scope of transport organisations input.

Keywords – quality management; ISO 9001; transport, supply chain

I. INTRODUCTION

The purpose of this paper is to analyse how the requirements of the ISO 9001:2015 standard are reflected in the transport processes and performance indicators used by organizations operating in the logistics and transport sector, based on publicly available information.

One of the research objectives is to develop an analytical and theoretical framework that correlates the requirements of the standard with transport operational processes and logistics performance indicators, providing a structured perspective on how these elements are reported and can be used in practice. The analysis is conducted within the scope of a screening of transport companies. Therefore, the empirical focus is limited to a small sample of organizations recognized as global industry leaders, selected as reference cases without the objective of achieving statistical representativeness across the entire sector. The fact that the paper is based on publicly available sources such as official company websites and corporate reports positions the study even more at a descriptive level.

Within this context, ISO 9001:2015 establishes a process-oriented management framework that requires organizations to consider risk and monitor process performance. Within logistics and transport operations, this framework influences how transport activities are organized and evaluated, even in the absence of explicit logistics-specific requirements. Additionally, ISO 9001:2015 requires organizations to determine and control externally processes, products and services that may influence their ability to deliver conforming results (Clause 8.4).This is particularly relevant for transportation organizations and logistics-specific activities, as core activities are frequently outsourced to carriers, subcontractors, suppliers, customers such as factories, which reinforces the importance of structured management of suppliers and partners within quality management systems. A further important aspect to this analysis concerns the way organizations address the requirement to monitor and evaluate the performance of their processes (Clause 9.1). For logistics and transport operations, this requirement is often operationalized through performance indicators published in public reports, such as delivery reliability or the quality of the transport performed, its optimization. In line with the descriptive scope, these performance indicators for some companies are not treated as measures of actual performance, but as observable signals of how organizations position their transport processes in relation to the requirements of the standard.[1]

According to the latest ISO Survey data powered by IAF CertSearch, there are approximately 1.47–1.48 million ISO 9001 certificates in operation globally, confirming the standard's role as a basic infrastructure for quality and operational efficiency in transport and logistics.[2]

The originality of this work lies in its analytical perspective, which involves two research areas often addressed separately in the specialized literature quality management standards and logistics performance. By connecting ISO 9001:2015 requirements with operational transport processes and logistics performance indicators, the paper can contribute to future in-depth research, data collection from different departments and the increasing success of the company. This paper is classified as a research paper because it presents an exploratory analysis based on publicly available data and presents a structured interpretation of the application of the ISO 9001 standard in logistics and transportation operations.

Confidential C

I. Literature review

Recent studies highlight the growing importance of quality management practices in logistics and transport operations, especially in the context of increased performance and competition in the market, targeting several sectors. However, these contributions remain largely fragmented, as they tend to focus either on quality standards or on logistics performance indicators, without proposing an integrated framework that explicitly links ISO 9001 requirements to transport processes and logistics key performance indicators.[3]

Recent empirical research has explored the impact of implementing ISO 9001 in logistics contexts through case studies. For example, a mixed-methods study conducted in the logistics department of Sobraga, a leading brewery in Gabon, reported improvements in service quality, reduced nonconformities, and increased customer satisfaction following the implementation of ISO 9001. Based on interviews and observations among employees, the paper highlights the role of management commitment, communication, and employee involvement in achieving improvements in service performance.[3] In contrast, the present research adopts a complementary perspective, focusing on publicly available corporate information and examining how ISO 9001 requirements are reflected and communicated through transportation processes and logistics performance indicators in several organizations. This analytical approach is at the screening level, without directly assessing internal performance outcomes. Alshahrani and Husain (2024) showed that the implementation of ISO 9001 standard contributes to increasing the level of customer satisfaction in the logistics field, by ensuring service consistency and reducing delivery delays. Their results also highlight the essential role of organizational and human factors in the effectiveness of the quality management system, emphasizing the importance of internal communication, understanding the role of quality integration in operational activities and the level of employee motivation.[4] Another author, López-Lemus (2023) also confirms that employee training plays a key role in the implementation of the ISO 9001 standard, especially by supporting awareness and internal alignment around quality objectives. These findings reinforce the same opinion.[5] According to Manders and de Vries, the successful implementation of a quality management system according to the ISO 9001 standard should be reflected in the improvement of the efficiency of internal processes, which will contribute to the improvement of the financial perspective through internal benefits, such as, for example, by increasing work productivity.[6] Empirical evidence from transport companies suggests that ISO 9001 implementation is associated with improvements in selected performance indicators. In a longitudinal study of transport firms in Slovakia, Cződörová and Gnap (2023) reported statistically significant differences in performance indicators following ISO 9001 certification.[6] Measuring service quality in logistics operations has also been addressed through evaluation models, for example, Knop (2019) applied the SERVQUAL method to evaluate the quality of services provided by a transport and logistics operator in the pharmaceutical sector, identifying some systematic gaps between customer expectations and service performance. The results reinforce the relevance of using structured performance indicators to identify weaknesses and support continuous improvement initiatives in logistics contexts.[7]

The supply chain has generally focused largely on efficient and cost-effective operations as a determinant of successful performance. Logistics practices are among the long series of actions and processes that contribute to the overall performance of the company and better serve end consumers and business customers. Omoush (2022) emphasizes that logistics practices should be analysed as an integrated system, given the interdependencies between inventory management, order processing, transportation.[8] Quality management practices extend beyond internal processes to include structured relationships with external partners, such as carriers, subcontractors, suppliers, and are directly linked to logistics processes. Given that we are talking about external parties involved in transportation operations, the coordination and monitoring of external activities have become critical elements of operational consistency and service reliability. Recent research, such as that of Daoud Ben Arab, 2021, highlights the fact that supplier-customer relationships are key quality management practices that influence process-related outcomes.[9] From a supply chain perspective, internal processes transform inputs received from suppliers and logistics partners into outputs delivered to customers. The benefits of ISO 9001 are influenced by a firm's position within its supply chain network and the structure of its relationships with suppliers and customers, rather than certification alone. This highlights the importance of examining coordination between suppliers and carriers within process-based quality management in logistics and transportation.[10] The reviewed literature confirms the relevance of ISO 9001 and quality management practices for improving service quality and operational performance in logistics and transportation contexts. However, research remains fragmented, as it predominantly focuses either on internal performance outcomes based on primary data or on isolated logistics activities and quality dimensions. This shows the need for an integrative, screening-level perspective that links quality management requirements with logistics processes and performance indicators across organizations.

There is evidence of the global adoption of ISO standards for quality management systems that illustrates their relevance in organizational practice. As shown in Table 1, ISO 9001:2015 is the most certified management system standard worldwide, both in terms of the number of certificates issued and the number of locations certified, surpassing other widely used standards such as ISO 14001 and ISO 45001. This widespread adoption suggests its use as a reference framework in several business sectors.

Table 1 Estimates from ISO Survey 2024 (published in 2025, powered by IAF CertSearch)

Standards	Number of certificates	Number of sites
ISO 9001:2015	1,474,118	2,321,640
ISO 14001:2015	676,232	1,176,389
ISO 45001:2018	542,527	941,546

Source: aifcertsearchcer.org

Confidential C

According to the analytical framework, this paper is classified as an exploratory research paper, at the screening level. It combines a structured theoretical analysis of the ISO 9001:2015 standard with an industry-wide documentary screening of publicly available corporate information. The paper is predominantly conceptual and analytical in nature, as it is not based on empirical data or in-depth internal case studies. Instead, it addresses a real practical need, namely the absence of an integrated operational framework that correlates ISO 9001 clauses with logistics-specific key performance indicators (KPIs).

A. Research Questions

1. To what extent is ISO 9001 adopted and reported by logistics and transport organizations and how does it stand in relation to quality management practices within the sector?
2. How can the ISO 9001 quality management principles be conceptually linked to logistics processes and key performance indicators (KPIs), based on publicly available information and illustrative examples from logistics practice??

II. MATERIALS AND METHODS

A. Research Design

This paper adopts an exploratory, screening level documentary analysis. The methodological approach is aligned with the research questions formulated in the previous section and focuses specifically on the applicability of the ISO 9001 standard in logistics and transport processes. The ISO 9001 standard is applicable to organizations in the logistics and transport sector, regardless of their size or type (road, rail, maritime, air, warehousing, distribution).[11] Due to its process-based approach, ISO 9001 allows for the control and optimization of specific activities in the logistics chain. The methodological component involves an industry-wide selection of transport companies, using publicly available certification data and documented descriptions of management systems. The analysis focuses on the degree of adoption of ISO 9001 and, where applicable, complementary standards such as ISO 14001, ISO 45001. The paper approach allows for a comparative understanding of how quality management systems are adopted and communicated in the logistics and transport sector. The methodology does not directly assess operational effectiveness but provides indicative guidance for practitioners and decision-makers. In this context, ISO 9001 serves as an operational reference that helps process coordination and performance monitoring in logistics and transport environments. The principles of ISO 9001 can be operationalized in logistics and transport by directly integrating them into processes that shape delivery reliability, product integrity, customer satisfaction and operational efficiency. [12]

B. Documentary Data Sources

The analysis is based on publicly available documentary sources from corporate websites, annual and sustainability reports, quality and certification information. These sources were selected because they are the primary means by which logistics and transportation organizations communicate their quality management practices, process structures, and performance monitoring approaches at the industry level. A deductive coding scheme is applied for the analysis of transport companies. Information extracted from corporate sources was organized according to predefined analytical categories, including: company profile and logistics sector, reported ISO management system certifications, and documented descriptions of how quality management systems are applied within logistics operations. This classification allows for a structured comparison between organizations. The aim of the coding process was to identify key points that link ISO 9001 requirements to logistics and transport processes, as communicated in corporate documentation. The analysis focused on the presence and structure of these links, rather than on quantifying performance outcomes.

To complet the study, a short set of open-ended questions was formulated for selected logistics organizations. The respondents were chosen from companies with direct relevance to logistics operations, including a global logistics carrier (XPO Logistics) and a European transport operator (Echemar). This set of questions adds a practical, although limited, perspective to the documentary screening and reflects how ISO 9001 is understood and applied in logistics activities. The responses were used as indicative practitioner input to support the proposed framework, without aiming at statistical generalization. This stage was not intended to generate empirical validation. The following guiding questions were used:

-How is ISO 9001 perceived in logistics operations (compliance vs operational tool)?

-Are logistics processes and KPIs aligned with quality management objectives?

-Would a framework linking ISO 9001 requirements to logistics processes be useful in practice?

III. ISO 9001 AS A PROCESS-BASED OPERATIONAL TOOL IN LOGISTICS

The ISO 9001 standard can be used as an operational reference tool based on processes such as: structuring logistics activities, defining performance indicators and continuous improvement mechanisms, it does not reflect a prescriptive implementation model.

A. Customer Orientation

In logistics and transport, customer orientation is a fundamental principle because the customer directly evaluates service quality through:

- on-time delivery,
- accuracy of shipments,
- condition of goods upon arrival,
- responsiveness to request
- clarity of communication during disruptions

ISO 9001 requires logistics companies to identify and continuously review customer requirements (clause 8.2) and ensure that these requirements are embedded in operational processes, according to ISO 9001.

Confidential C

B. Process approach

Logistics involves interdependent activities such as goods receipt, warehousing, transport planning, delivery and returns management. If the process-based approach is applied, the ISO 9001 standard allows these activities to be managed as interconnected processes, with clearly defined inputs, outputs and responsibilities. This approach facilitates the identification of critical points in the logistics chain, reduces search time, possible errors and ensures an efficient flow contributing to increasing the reliability of transport services. To illustrate how ISO 9001 operationalizes the process-based approach in transport logistics, Table 2 presents the main inputs, outputs and control mechanisms associated with three basic logistics processes: transport planning, loading/unloading operations and the delivery process. [ISO 9001:2015]

Table 2 Process approach in transport logistics

Nr.	Transport planning	Loading/unloading operations:	Delivery process:
1	Inputs: customer order	Inputs: prepared shipment	Inputs: transport plan
2	Outputs: assigned truck and route plan	Outputs: load secured	Outputs: successful delivery
3	Controls: traffic conditions, load optimization, customer time slot	Controls: weight checks, sealing procedures, scanning/tracking	Controls: GPS tracking, communication protocols between involved parties

Source: Created by the author based on ISO 9001:2015 (process approach requirements)

C. Continuous improvement

Continuous improvement is essential in logistics because it is a field characterized by frequent changes in demand and the pressure to reduce costs and improve operations is increasingly demanding. Logistics departments that apply this principle with a periodically analyzed process performance, which evaluates the results obtained and implements corrective and preventive actions can monitor performance indicators. Constant review of procedures and integration of ISO 9001 support the optimization of transport routes, reduction of delivery times and increase of operational efficiency. Table 3 summarizes the practical initiatives of continuous improvement commonly implemented in logistics organizations according to the ISO 9001 standard. [ISO 9001:2015]

Table 3 Continuous improvement of activities

Nr.	Improvement processes	Action examples
1	Route optimization	Analysis of delivery lead times and delays leads to redesigned routes or revised departure times.
2	Reduction of repeated non-conformities	Repeated damaged-goods complaints training in load securing and revised loading checklists.
3	Improvement cycles based on KPIs	The organization performs a PDCA cycle to identify root causes (planning, traffic, subcontractor issues).
4	Fuel cost and emission optimization	Periodic analysis of fuel usage leads to better vehicle allocation or eco-driving training programs.

Source: Created by the author

In order to assess the implementation of the ISO 9001 standard in the logistics and transport sector, a set of relevant performance indicators was analyzed. These indicators reflect both operational efficiency and service quality, by allowing an objective assessment of the impact of quality management practices on logistics processes. The performance indicators presented in Table 4 were selected based on their relevance to the basic principles and requirements of the ISO 9001:2015 standard and their applicability to logistics and transport services. If these indicators are monitoring, organizations are able to evaluate the process control, risk management and continuous improvement initiatives, thus ensuring an alignment between quality objectives and operational performance. To illustrate how ISO 9001 requirements are translated into measurable operational results, Table 4 presents key performance indicators (KPIs) used in logistics and transport, showing their connection to ISO 9001 and their impact on service quality. [ISO 9001:2015]

Table 4 Performance Indicators Relevant to ISO 9001 Implementation in Logistics and Transport

Performance Indicator	Relevance to ISO 9001	Impact on Logistics Quality	Example
On-time delivery rate	8.1 Process control / 9.1 Performance evaluation	Increased reliability and customer satisfaction.	Delivery completed within agreed time window (±15 min)
Operational efficiency	9.1 Monitoring and measurement	Cost reduction and resource optimization	Cost per shipment, vehicle utilisation, fuel consumption
Nonconformities	10.2 Corrective action	Reduction of operational errors and complaints	Damaged goods, incorrect documentation, missed time slots

Created by the author based on ISO 9001:2015 requirements

To support an evaluation of ISO 9001, Table 5 offers an expanded description of the selected performance indicators and shows a correlation of each KPI with specific ISO 9001 clauses and practical manifestations in transport logistics.

Table 5 Key performance indicators descriptions

Nr.	Performance Indicator	Description
1	On-time delivery rate	This indicator measures the percentage of deliveries completed within the agreed timeframe. ISO 9001 contributes to improving this indicator by standardizing transport planning processes, monitoring deviations, and implementing corrective actions. Carrier performance monitoring (clause 8.4), where subcontracted transporters with on-time rates below a set target (e.g., <92%) receive corrective actions or are reassigned to less time-critical side
2	Operational efficiency	Operational efficiency is assessed through indicators such as transport cost per shipment, vehicle utilization rate, and route optimization. The continuous improvement principle promoted by ISO 9001 supports performance monitoring and data-driven decision-making, leading to cost reduction and better resource utilization. Data-driven route optimisation, where historical lead-time deviations and fuel consumption data determine route reassignment (e.g., switching from Route A to Route B due to 12% lower fuel usage);
3	Non-conformities	Refers to damaged goods, lost shipments, incomplete documentation, temperature deviation (cold chain), or incorrect labelling. ISO 9001 requires root cause analysis and corrective actions (10.2), leading to systematic error reduction.

Created by the author based on ISO 9001:2015 requirements

Confidential C

Since the application of ISO 9001 principles aims to ensure stable and measurable logistics processes, the transition from principles to practice can take place through the PDCA cycle, a tool used in the Quality field and easily integrated into quality management activities. The following section illustrates how PDCA supports process control, KPI monitoring and corrective actions in transport operations. The PDCA quality cycle, also known as the Deming cycle or PDCA circle, was introduced by Deming, a renowned figure in quality management. The objective of the approach is to consolidate the knowledge acquired and derive corrective actions to address unresolved problems and improve the execution strategy. [13]

In the planning stage, organizations define the objectives, resources and controls necessary to execute transport operations under predictable conditions. According to official ISO guidelines, the Planning phase involves setting objectives, defining performance indicators and specifying methods for achieving results. The "DO" phase corresponds to the implementation of planned activities. ISO describes this as executing processes and applying controls to ensure compliance. This is represented in Table 6:

Table 6 Planned activities

Process category	Examples in transport logistics
Loading operation	Secure cargo Carrying out weight checks and scanning of shipments before loading. In logistics, legislation requires that the weight on the first axle on public roads should not exceed 11.5 t
Real-time execution	Use GPS tracking to follow vehicles Ensure drivers follow defined communication protocols+
Documentation and compliance	Issue loading notes, CMR, delivery instructions and safety checklists.

Created by the author based on ISO 9001:2015 requirements

This demonstrates how PDCA operationalizes ISO 9001 requirements across the transport chain, aligning strategic quality goals with daily operational performance in fig. 1:

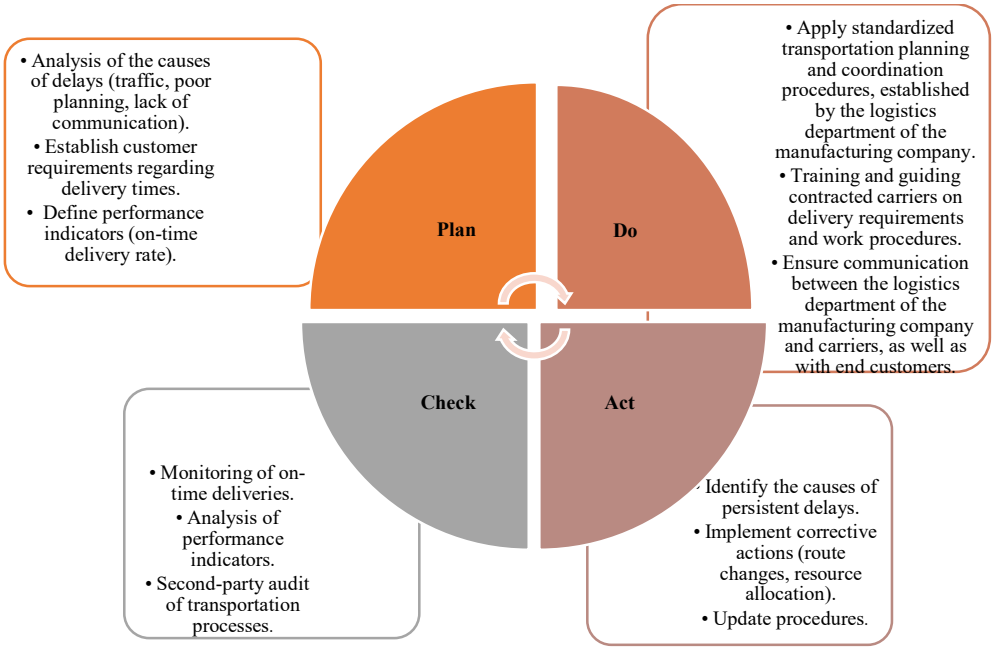


Figure 1 PDCA Plan→Do→Check→Act

Created by the author based on PDCA examples

IV. COMPARATIVE SCREENING OF LOGISTICS COMPANIES

The selection of transport companies below followed a procedure with the exploratory nature of the paper. The selection of companies followed a purposive sampling approach, in line with the exploratory nature. The organizations were chosen based on their visibility in the logistics sector and the accessibility of information on their quality management systems. In addition, the selection was supported by the author's professional experience in the logistics field, which facilitated the identification of relevant organizations representing different operational contexts, including both global logistics carriers and regionally focused transport operators.

The inclusion of both global logistics providers and smaller operators allows for a broader perspective on reported quality management practices in different organizational contexts. Based on public data, a preliminary selection was carried out to identify the certifications held by each organization and the general orientation of their quality, environmental and operational management systems. The information presented in Table 7 is based on publicly available corporate information and provides an overview of the use of ISO 9001 and other related management system certifications adopted and communicated within selected logistics and transport organizations. The table is not intended to provide a detailed assessment of internal operational

Confidential C

practices but rather serves as an indicative screening tool that highlights the widespread presence of ISO 9001 within the sector and how it is publicly positioned in relation to logistics and transport activities.[14] [15] [16][17][18]

Table 7 Preliminary screening of selected companies

Company	Main field of work/Sector	Relevant Standard/Certification	Publicly Reported Logistics and Quality Practices
DHL Freight	International logistics & transportation	ISO 9001 (quality management) ISO 45001 (occupational safety and health protection) ISO 14001 (environmental management) ISO 50001 (energy management)	Emphasizes structured quality objectives, quantifiable results and efficient resource utilization as key elements of quality management. ISO 9001 supports the evaluation and optimization of logistics processes to ensure reliable transport services. The company integrates multiple ISO and industry-specific certifications to enhance quality, operational safety and process consistency.
ROUTIER Transport	European transport & logistics	Certifications ISO 9001, ISO 14001, ISO 45001, ISO 28000, ADR, SQAS	Combines quality management, environmental standards and logistics security systems. ISO 9001 is applied alongside other certifications to support integrated operational control and ensure compliance with safety and transport requirements.
FedEx	Global logistics and transportation	ISO 9001 (quality management) Authorized Economic Operator (AEO) with Class "AA" accreditation IATA CEIV Pharma certification (for pharmaceutical logistics)	The organization supports globally standardized quality processes and long-term consistency. ISO 9001 certification has been continuously applied for over 30 years, contributing to the standardization of processes. Additional certifications, such as IATA CEIV Pharma, reflect high quality standards in specialized logistics operations.
DB Schenker	Road, rail, sea, air transport	ISO 9001 ISO 14001 ISO 50001 SQAS	Applies ISO 9001 across all transport modes; integrates environmental and energy management for efficient operations.
XPO LOGISTIC	Global logistics & transportation	ISO 28000 (supply chain security) AEO Class A (Authorized Economic Operator) Certipharm (healthcare logistics) EcoVadis Gold Medal (sustainability rating) ISO 14001 (environmental management – selected sites)	ISO 9001 is not just a certification requirement for XPO, it is a key operational tool. XPO reflects the company's publicly reported focus on quality, security, sustainability, security, sustainability, and regulatory compliance. It has renewed ISO 28000 certification for supply chain security at its Canelas site in Porto, indicating effective risk management across the logistics chain. ISO 28000 helps monitor and protect goods and reduces hazards in the supply chain. The company also holds Authorized Economic Operator (AEO) Class A status, facilitating efficient customs procedures and secure transport.
GEODIS	Global logistics & supply chain services	ISO 9001 (quality management) ISO 14001 (environmental management) ISO 45001 (occupational health & safety) AEO (Authorized Economic Operator) TAPA (Transport Asset Protection Association) ISO 13485 (medical device quality) SQAS (Safety & Quality Assessment for Sustainability) IATA CEIV Pharma (pharma logistics)	GEODIS publicly reports multiple certifications across various sites. ISO 9001 is applied across hundreds of sites, supported by environmental and safety standards (ISO 14001, ISO 45001) and industry-specific accreditations (AEO, TAPA, CEIV for pharmaceuticals), demonstrating a holistic and integrated approach to quality, security, and regulatory compliance
ECHEMAR	European road transport & logistics	ISO 9001 (quality management) ISO 27001 information security management systems. ISO 14001 (environmental management) ISO 14064-1:2018 Greenhouse Gas (GHG) reporting	In 2003, Echemar S.L.U obtained ISO 9001 certification. Our main objective is to achieve total quality in all operational processes. Operates in European road transport and logistics, combining fleet control, real-time monitoring and proprietary IT systems to support operational efficiency and traceability. ISO 9001 certification reflects a structured approach to quality management, supported by environmental and information security standards and a focus on continuous improvement of logistics processes.

Source: Author's compilation based on publicly available information from company websites and certification listings

The screening results presented in Table 7 highlight several consistent observations regarding the adoption and positioning of the ISO 9001 standard within transport organizations. First, ISO 9001 certification is present in all selected companies, regardless of their size or field of activity. This indicates that ISO 9001 is widely recognized as the reference standard for quality management in the supply chain operations. Second, ISO 9001 is often combined with additional management system standards, such as ISO14001, ISO45001 or ISO28000. This suggests that companies are integrating quality management into broader operational frameworks that also address environmental, safety and supply chain security requirements. Many organizations emphasize the integration of quality management with operational efficiency and risk control. For example, XPO and GEODIS emphasize the importance of compliance, monitoring and preventive measures within logistics operations. Thus, ISO 9001 plays a particularly important role in risk management, and logistics is an area where operational and cost risks are increasingly high.

In addition to the information published and analyzed above, additional information was obtained through written contributions from two logistics organizations representing different operational contexts. This short series of open-ended questions provides a complementary perspective to the documentary selection, providing a practical although limited perspective of the paper on how ISO 9001 is perceived and applied within logistics operations. From the preliminary responses, it follows that ISO 9001 is not perceived only as a formal certification requirement, but rather as an operational tool that supports

Confidential C

logistics processes, performance monitoring and continuous improvement activities. To offer a structured overview of this preliminary information, the responses are summarized in Table 8:

Analytical aspect	Preliminary observations from the interviewed carriers
Role of ISO 9001	The standard is perceived as an operational tool that supports process structuring, efficiency and continuous improvement.
Alignment with KPIs	Logistics processes are aligned with quality objectives through indicators such as on-time delivery, service quality and claims monitoring. For example, on-time delivery is monitored by adhering to predefined delivery time slots, allowing transport planners to adjust routes and schedules in the event of delays. Non-conformities are tracked through records of damaged goods, incorrect documentation or loading errors, which trigger corrective actions such as retraining staff or reviewing operational procedures. Customer satisfaction is assessed through feedback mechanisms such as complaints survey, service quality reviews or post-delivery evaluations, which are used to improve communication processes between the parties involved.
Customer satisfaction	Customer satisfaction is frequently assessed through complaint management systems, particularly in cases involving late deliveries, incomplete shipments, or service communication failures. In addition, third-party audits conducted in collaboration with key customers are used to assess service performance, ensure compliance with contractual quality requirements, and identify opportunities for process improvement. These audits help to strengthen transparency, accountability, and customer-oriented performance in logistics operations.

Author's compilation based on written responses from transport organizations

Overall, the combined interpretation of the selection results and preliminary practitioner contributions suggests that ISO 9001 is consistently presented as a practical framework for organizing processes, maintaining operational control, and supporting customer-oriented activities in logistics. These findings, based on corporate information and a limited perspective of the organizations, indicate that ISO 9001 is used in practice as an operational reference, rather than just as a formal certification requirement.

V. CONCLUSIONS

The results of this paper provide an exploratory perspective on the role of ISO 9001 in transport organizations, based on the analysis of publicly available information and limited contributions from practitioners. Regarding the first research question, the findings indicate that ISO 9001 is widely adopted within the companies surveyed and is consistently presented as part of integrated management systems, often combined with environmental, safety and supply chain security standards. However, this observation is derived from published corporate information and does not reflect a direct insight into internal operational processes. For example, companies such as DHL, DB Schenker and GEODIS integrate ISO 9001 with environmental and safety standards, while XPO emphasizes its linkage with supply chain security and risk management practices.

Regarding the second research question, the analysis suggests that the principles of ISO 9001 are commonly associated with process structuring, standardization and operational control in logistics processes. The principles are applied through performance indicators, such as on-time delivery, service quality and complaint monitoring, as well as operational practices, such as incident tracking, corrective actions and customer feedback mechanisms.

The contributions of transport organizations support the view that the standard is used in practice to align logistics processes with performance monitoring and continuous improvement mechanisms. It must of course be acknowledged that the study is based on a limited sample of transport organizations. Therefore, the practices identified can be interpreted as indicative examples of "best practice" approaches, rather than being representative of the entire logistics and transport sector. In conclusion, the study does not aim to provide an empirical validation of the effectiveness of ISO 9001, but rather to illustrate how the standard is positioned and perceived within logistics operations.

Future research could extend this research by including empirical data, more stakeholders involved in interviews, surveys or internal performance measurements, to further validate the observed relationships.

A. Limitations of the study

This paper has several limitations that need to be presented. First, the research is based primarily on publicly available corporate information, including company websites, certification data, and sustainability reports or annual reports. Such sources are subject to reporting errors, as they tend to highlight positive aspects of organizational practices and may not fully reflect internal operational challenges or inconsistencies. Thus, the results should be interpreted as indicators of how organizations communicate their quality management practices, rather than as direct evidence of actual implementation or performance outcomes.

Secondly, the paper is based on an exploratory selection of a limited number of organizations. As a result, the practices identified may reflect "best practice" approaches typical of well-established companies, rather than the full diversity of practices present in the transport and logistics sector, especially among smaller or less structured operators.

Third, the study does not include access to internal company data, performance indicators or operational documentation. As a result, the study does not assess the effectiveness of ISO 9001 implementation in terms of measurable outcomes, but focuses instead on how the standard is positioned and understood in transport activities. Finally, the practitioner input included in the study is based on a limited number of written responses and is used as indicative insight rather than empirical evidence. Although these responses offer a useful practical perspective, they are not sufficient to support statistical generalization or comparative analysis. Future research could address these limitations by incorporating primary empirical data, such as larger-scale surveys, structured interviews or access to internal operational indicators, in order to better explore the relationship between ISO 9001 and logistics performance.

References

Confidential C

[1]ISO (2015), *ISO 9001:2015 – Quality management systems: Requirements*, International Organization for Standardization, available at: https://www.iso.org/home.html (accessed 9 May 2026).

[2]IAF CertSearch (n.d.), *ISO Survey*, available at: https://www.iafcertsearch.org/services/iso-survey (accessed 9 May 2026).

[3]Nze Ayong, A.H. and Yan, B. (2025), "The role of ISO 9001 in improving service quality at Sobraga's logistics department", in *Proceedings of the 8th European Conference on Industrial Engineering and Operations Management*, Paris, France, 2–4 July, IEOM Society International, available at: https://doi.org/10.46254/EU08.20250362.

[4]Alshahrani, M.A. and Husain, K.S. (2024), "The effectiveness of the implementation of ISO 9001 on SMEs performance: the case of an emerging economy", *International Journal of Quality & Reliability Management*, Vol. 41 No. 1, pp. 84–106.

[5]López-Lemus, J.A. (2023), "ISO 9001 and the public service: an investigation of the effect of the QMS on the quality of public service organisations", *International Journal of Organisational Analysis*, Vol. 31 No. 4, pp. 1143–1164.

[6]Cződörová, R. and Gnap, J. (2023), "Investigation of the effectiveness of the introduction of the quality management system according to the ISO 9001 standard in transport companies: Slovakia case study", *Sustainability*, Vol. 15 No. 3, p. 2401, available at: https://doi.org/10.3390/su15032401.

[7]Knop, K. (2019), "Evaluation of quality of services provided by a transport and logistics operator from pharmaceutical industry for improvement purposes", *Transportation Research Procedia*, Vol. 40, pp. 1080–1087, available at: https://doi.org/10.1016/j.trpro.2019.07.152.

[8]Omoush, M.M. (2022), "The impact of logistic management practices on operational performance: a field study of road transport companies", *Journal of Governance and Regulation*, Vol. 11 No. 4 (Special Issue).

[9]Daoud Ben Arab, S. (2022), "Quality management practices and innovation: the moderating effect of ISO 9001 certification", *Journal of the Knowledge Economy*, Vol. 13, pp. 2177–2202.

[10]Hung-Chung, S., Ta-Wei, K. and Linderman, K. (2020), "Where in the supply chain network does ISO 9001 improve firm productivity?", *European Journal of Operational Research*, Vol. 283 No. 2, pp. 530–540, available at: https://doi.org/10.1016/j.ejor.2019.11.042.

[11]Lascu, E., Severin, I., Lascu, F.D., Gudana, R.A., Nalbitoru, G. and Ignat, N.D. (2021), "Framework on performance management in automotive industry: a case study", *Journal of Risk and Financial Management*, Vol. 14 No. 10, p. 480, available at: https://doi.org/10.3390/jrfm14100480.

[12]Farag, M.I.H. (2025), "Sustainability as a management strategy: integrating environmental, social, and governance practices into business administration", *Management Science Advances*, available at: https://doi.org/10.31181/msa31202630.

[13]Qin, X., Li, H. and Wen, Y. (n.d.), "Research on the reform of logistics and supply chain management courses based on PDCA quality circle", *International Journal of Education and Humanities*, Vol. 13 No. 3, available at: https://pdfs.semanticscholar.org/0f92/0a8434b675979a71b0c72966f8603688c79a.pdf (accessed 16 March 2026).

[14]DHL Freight (n.d.), *Homepage*, available at: https://dhl-freight-connections.com/en/ (accessed 16 March 2026).

[15]Routier Transport (n.d.), *Homepage*, available at: https://routiertransport.com/ro/acasa/ (accessed 16 March 2026).

[16]FedEx (n.d.), *Homepage*, available at: https://www.fedex.com/ro-ro/home.html (accessed 16 March 2026).

[18]XPO Logistics (n.d.), *Homepage*, available at: https://europe.xpo.com/en/ (accessed 16 March 2026).

GEODIS (n.d.), *CSR Report 2023*, available at: https://geodis.com/csrreport2023/100/ (accessed 16 March 2026).

Confidential C

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Quality Management System of forensic genetics laboratories in Brazil: a structured guideline of quality management and improvement program

Laryssa Silva de Andrade Bezerra[1,2], Advaldo Carlos Souza Neto[2], Maico Roris Severino[1], Fabiane Letícia Lizarelli[3] e Camila Fabricio Poltronieri[1,4,]

1 Universidade Federal de Goiás (UFG) – Programa de Pós-graduação em Engenharia de Produção; laryssasa@yahoo.com.br , maico_severino@ufg.br
2 Superintendência de Polícia Técnico-Científica de Goiás (SPTC-GO); acsnbio@gmail.com
3 Universidade Federal de São Carlos (UFSCar); fabiane@dep.ufscar.br
4 Universidade de São Paulo (USP); camilafp@usp.br

***Abstract*—In Brazil, the implementation of a quality management system in forensic genetics laboratories is based on its own normative, which is defined by the Steering Committee of Integrated Network of Genetic Profile Databases (RIBPG). However, a quality management system based in ISO/IEC 17025 is essential for international acceptance and accreditation of these laboratories. In conformity assessment, a quality management and improvement program (QMIP) with internal and external assessments is necessary. This work aims to describe a QMIP structured guideline for Brazilian forensic genetics laboratories. The description of this system was based on documentary research, involving an analysis of inputs, outputs, the transformation process, objectives, and the emergence of this program, detailing the role of each stakeholder in the process. Describing the structured guideline is essential to provide a context, a solid theoretical and methodological basis that guide decisions, facilitate implementation, and ensure the coherence and effectiveness of the resulting system.**

Keywords – RIBPG, credibility, performance, accreditation, Genetic Profile Databases.

I. Introduction

Forensic genetics has become an important component of contemporary criminal justice systems, supporting criminal investigations, human identification, and missing persons investigations through DNA analysis. The expansion of national genetic profile databases has further increased the role of forensic genetics by enabling the comparison and exchange of genetic profiles across jurisdictions. In this context, quality management frameworks have been increasingly adopted to support the reliability, traceability, and consistency of forensic laboratory activities.

Quality Management Systems (QMS) are widely recognised as important tools for promoting technical competence, standardisation, and confidence in forensic laboratory results. According to Pillai et al. (2022), the implementation of a QMS requires laboratories to define objectives, identify operational needs, and establish procedures aligned with applicable regulatory and technical requirements. Internationally, accreditation according to ISO/IEC 17025 has been adopted in several countries as a mechanism for demonstrating laboratory competence and promoting harmonisation among forensic institutions (Pádár et al., 2015; Nogel et al., 2019). However, studies have also highlighted operational, financial, and organisational challenges associated with implementing and maintaining these systems, particularly in public forensic laboratories and resource-constrained environments (Triverio and Márquez, 2022).

In Brazil, forensic genetics laboratories are organised through the Integrated Network of Genetic Profile Databases (RIBPG), a cooperative system involving the Ministry of Justice and Public Security, state public security agencies, and the Federal Police. Within this framework, each federative unit maintains its own local database while sharing genetic profiles through the National Genetic Profile Database (BNPG). This structure enables the exchange of forensic genetic information among Brazilian jurisdictions and with international partners through INTERPOL, supporting criminal investigations and missing persons identification (Da Silva Junior et al., 2020; RIBPG, 2022).

The Brazilian forensic genetics system presents operational and regulatory characteristics that differ from those observed in many European and North American forensic networks. While international initiatives such as ENFSI and SWGDAM/FBI provide guidance for quality assurance and accreditation, the Brazilian context involves a federative structure characterised by heterogeneous institutional capacities, different levels of QMS maturity, and decentralised operational management. In this scenario, harmonisation of quality practices and continuous monitoring mechanisms represent important challenges for the national forensic system.

To support quality assurance within the RIBPG, the Steering Committee established Resolution No. 12, which defines technical and management requirements for laboratories participating in the network. The resolution incorporates elements derived from ISO/IEC 17025, as well as forensic-specific requirements related to DNA databases and laboratory procedures (RIBPG, 2019). Nevertheless, Resolution No. 12 does not fully encompass all ISO/IEC 17025 requirements, and accreditation according to ISO/IEC 17025 remains voluntary in Brazil. Previous studies have demonstrated that Brazilian forensic genetics laboratories still face important challenges regarding the implementation of several ISO/IEC 17025 requirements, particularly those related to calibration, method validation, proficiency testing, and risk management (Bezerra et al., 2025).

Among the requirements addressed by both ISO/IEC 17025 and Resolution No. 12 are activities related to continuous monitoring, internal assessments, corrective actions, and quality improvement processes. These mechanisms may contribute to identifying nonconformities, monitoring laboratory performance, and supporting continuous improvement practices within forensic laboratories. However, despite the existence of regulatory requirements and international guidance documents, the literature still provides limited discussion on how structured quality monitoring programs can be operationalised within large federative forensic systems such as the Brazilian RIBPG.

Therefore, this study proposes a structured Quality Management and Improvement Program (QMIP) framework for Brazilian forensic genetics laboratories, integrating international quality management principles with the specific regulatory and operational characteristics of the Brazilian forensic genetics system. The proposed framework aims to support the development of harmonised monitoring practices capable of assisting laboratories in quality assessment, continuous improvement, and alignment with national and international quality requirements.

II. Methodological Procedure

This study adopted a qualitative descriptive approach based on bibliographic review and documentary analysis to develop a conceptual framework for a QMIP applicable to Brazilian forensic genetics laboratories. The methodological approach aimed to analyse regulatory, technical, and operational references related to QMS, conformity assessment procedures, and continuous quality monitoring practices in forensic genetics laboratories.

A bibliographic review was conducted using the Scopus and Web of Science databases. Searches were performed using combinations of the following terms: "forensic genetics", "forensic DNA laboratory", "quality management", "quality assurance", "ISO/IEC 17025", "forensic accreditation", "forensic laboratory auditing", and "DNA database". Publications published between 2010 and 2024 in English, Portuguese, or Spanish were considered. Inclusion criteria comprised peer-reviewed studies addressing quality management systems, accreditation processes, laboratory monitoring activities, auditing practices, or quality assurance in forensic laboratories. Studies focused exclusively on clinical or non-forensic laboratory settings were excluded.

In addition to the literature review, documentary analysis was conducted on normative and technical documents related to forensic laboratory quality systems and conformity assessment procedures. The documents were selected according to three criteria: (i) regulatory relevance to forensic genetics laboratories, (ii) international recognition within forensic quality management systems, and (iii) applicability to the Brazilian forensic context. The analysis included documents issued by the National Institute of Metrology, Quality and Technology (INMETRO), the Brazilian Ministry of Justice and Public Security (MJSP), the International Laboratory Accreditation Cooperation (ILAC), the ANSI National Accreditation Board (ANAB), the Scientific Working Group on DNA Analysis Methods (SWGDAM), and the European Network of Forensic Science Institutes (ENFSI). The standards and guidelines analysed in this study are presented in Table 1.

Table 1 – Standards and/or guidelines for the evaluation of quality management systems in forensic genetics laboratories

Document	Institution	Year of publication
Resolution No. 12 of the RIBPG Managing Committee	RIBPG Steering Committee/MJSP (Brazil)	2019
ISO/IEC 17025	ILAC/ ABNT (Brazil)	2017
DOQ-CGCRE-084	CGCRE-INMETRO (Brazil)	2016
Quality assurance standards for DNA databasing laboratories	SWGDAN/ FBI (USA)	2020
Accreditation Manual for Forensic Laboratories, Forensic Inspection Bodies, and Property and Evidence Control Units	ANAB (USA)	2023
DNA database management review and recommendations	ENFSI (Europe)	2017

source: prepared by the authors

Among the analysed documents, Resolution No. 12, ABNT NBR ISO/IEC 17025, and DOQ-CGCRE-084 are directly applicable to Brazilian forensic genetics laboratories. The remaining documents correspond to international references adopted in North American and European forensic systems and were included to support comparative analysis of quality monitoring and accreditation practices. Resolution No. 12 establishes the technical and management requirements applicable to laboratories participating in the RIBPG. These requirements define the conditions for maintaining authorization to share genetic profiles with the BNPG. ISO/IEC 17025, although not mandatory for forensic genetics laboratories in Brazil, is internationally recognised as a reference standard for laboratory competence and quality management. In the Brazilian accreditation system, conformity assessment according to ISO/IEC 17025 is conducted by the General Coordination for Accreditation (CGCRE), the accreditation body of INMETRO. The DOQ-CGCRE-084 document corresponds to the Brazilian adoption of ILAC-G19 and provides guidance regarding the application of ISO/IEC 17020 and ISO/IEC 17025 within forensic activities.

The conceptual QMIP framework was developed through iterative comparative analysis of the selected documents, focusing on recurring components related to quality monitoring, conformity assessment, auditing structures, proficiency testing activities, corrective actions, and continuous improvement mechanisms. The PDCA cycle (Plan–Do–Check–Act) was adopted as an organisational reference for structuring the framework due to its broad application in quality management systems.

To support the visual representation of the proposed framework, organisational charts and process flowcharts were developed using Lucidchart software. The conceptual model was structured according to input-process-output logic, defining the principal operational components, assessment activities, actors involved, and expected outputs associated with the proposed QMIP. Internal and external assessment mechanisms were described based on Resolution No. 12 and CGCRE accreditation guidance documents.

As this study proposes a conceptual framework derived from literature review and documentary analysis, no empirical validation or pilot implementation was conducted during the present research. This limitation should be considered when interpreting the applicability of the proposed model and highlights the need for future studies involving operational validation, expert consultation, and implementation analysis in forensic laboratories.

III. RESULTS

Brazilian forensic genetics laboratories currently present different levels of QMS implementation and maturity. To contextualize the proposed framework, publicly available information from CGCRE and the most recent report issued by the RIBPG Steering Committee were analysed to identify the current status of Brazilian forensic genetics laboratories regarding compliance with Resolution No. 12 and accreditation according to ISO/IEC 17025.

TABLE 2 – STATUS OF BRAZILIAN FORENSIC GENETICS LABORATORIES IN RELATION TO THE QUALITY MANAGEMENT SYSTEM

Status of RIBPG laboratories	Quantity
Without compliance with Resolution No. 12 (without a genetic profile database)	6
In compliance only with Resolution No. 12 (including those temporarily suspended)	21
In compliance with Resolution No. 12 of the RIBPG and accredited to ISO/IEC 17025	1
Total number of brazilian forensic genetics laboratories	**28**

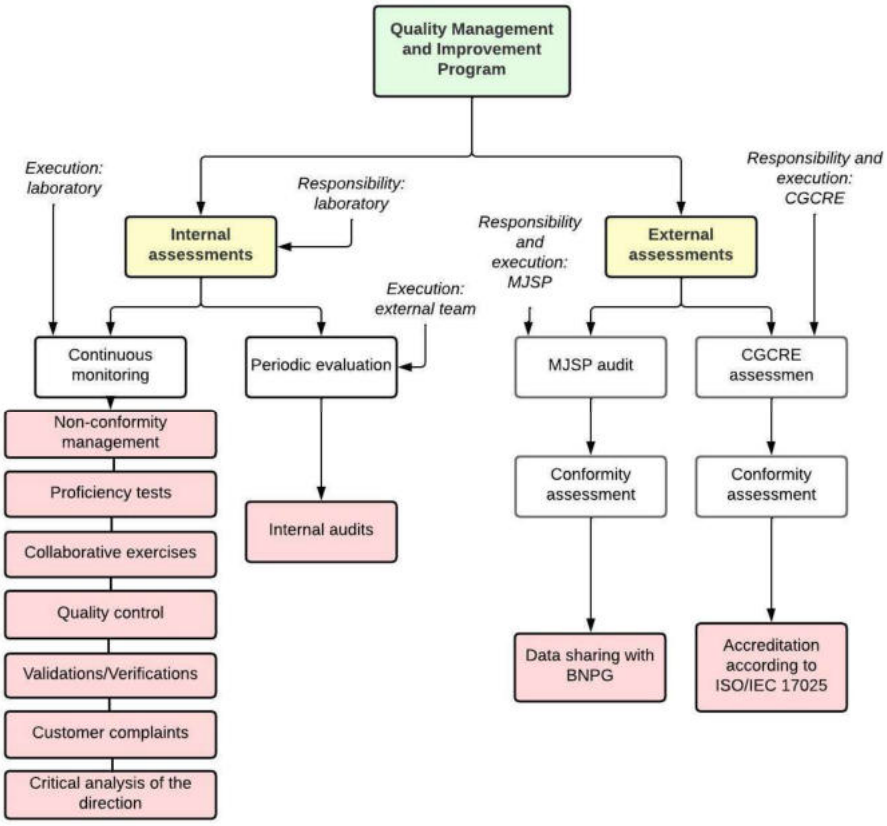
source: prepared by the authors

The data presented in Table 2 demonstrate the heterogeneity of QMS implementation among Brazilian forensic genetics laboratories. Although most laboratories comply with the minimum requirements established by Resolution No. 12, accreditation according to ISO/IEC 17025 remains limited within the Brazilian forensic genetics system. These findings reinforce the relevance of developing structured quality monitoring approaches adapted to the operational and regulatory characteristics of the Brazilian context.

Based on the analysed regulatory framework, the proposed QMIP was structured considering the complementary roles of Resolution No. 12 and ISO/IEC 17025. The framework was organised according to the PDCA cycle (Plan–Do–Check–Act) and incorporates internal and external assessment mechanisms intended to support conformity assessment, quality monitoring, and continuous improvement activities within forensic genetics laboratories.

In addition to continuous improvement principles, the proposed framework incorporates document and record control mechanisms, monitoring activities, internal assessments, proficiency testing participation, and external conformity assessment processes. Figure 1 presents the conceptual structure of the proposed QMIP framework and illustrates the interaction between internal monitoring activities, external assessments, corrective actions, and continuous improvement processes within the Brazilian forensic genetics system.

FIGURE 1 – CONCEPTUAL QMIP FRAMEWORK FOR BRAZILIAN FORENSIC GENETICS LABORATORIES



source: prepared by the authors

Based on the overall view of the program, the evaluation system is described according to the reference documents listed in Figure 1.

A. Internal assessments

Internal assessments correspond to periodic and continuous monitoring activities conducted under the responsibility of the laboratory and integrated into the routine activities of the QMS. These activities are intended to support conformity assessment, identification of nonconformities, monitoring of laboratory performance, and continuous improvement processes. Although Resolution No. 12 and ISO/IEC 17025 establish requirements related to internal monitoring and quality assurance, neither document prescribes a single operational model for implementing these activities. Consequently, laboratories may structure their internal monitoring systems according to their organisational characteristics, operational scope, and quality management maturity level.

- Continuous monitoring

Continuous monitoring activities are commonly adopted within QMS frameworks to support conformity assessment, identification of nonconformities, monitoring of analytical performance, and verification of compliance with technical and management requirements. These activities involve periodic evaluation of processes, procedures, equipment, and analytical performance indicators. Within forensic genetics laboratories, continuous monitoring activities may include the use of reference materials and quality control materials, intermediate equipment checks, replicate analyses, retesting of retained items, review of reported results, control chart monitoring, interlaboratory comparisons, and analysis of blind samples. These activities are described in ISO/IEC 17025 and in the DOQ-CGCRE-084 guidance document issued by CGCRE/INMETRO. The proposed framework incorporates these monitoring activities as part of a continuous evaluation process intended to support reliability, traceability, and consistency of laboratory operations.

- Periodic evaluation

Periodic evaluation activities within the proposed QMIP framework include internal audits and participation in proficiency testing (PT) programs. These activities contribute to evaluating conformity with QMS requirements, identifying nonconformities, monitoring laboratory performance, and supporting continuous improvement processes. Resolution No. 12 establishes the implementation of periodic internal audits within participating laboratories, while ISO/IEC 17025 requires laboratories to conduct internal audits at planned intervals to verify conformity with internal procedures and applicable accreditation requirements. In addition, ISO 19011 may be adopted as guidance for planning and conducting internal audits.

Participation in proficiency testing programs is required both by Resolution No. 12 and by ISO/IEC 17025. However, differences exist regarding periodicity, provider recognition, participation planning, and performance evaluation criteria. While Resolution No. 12 requires annual participation in proficiency testing activities recognised by the RIBPG Steering Committee, ISO/IEC 17025 adopts a risk-based approach and requires participation according to planning criteria established by accreditation guidance documents such as NIT-DICLA-026. Another relevant difference concerns provider competency assessment. Under ISO/IEC 17025, PT providers are expected to comply with ISO/IEC 17043 requirements, whereas Resolution No. 12 establishes recognition by the RIBPG Steering Committee as the primary criterion. Table 3 summarises the principal differences identified between these two frameworks regarding proficiency testing activities.

TABLE 3 – COMPARISON BETWEEN PROFICIENCY TESTING REQUIREMENTS IN RESOLUTION NO. 12 AND ISO/IEC 17025

Guidance document	Resolution No. 12	ISO/IEC 17025
Participation requirements	Defined by Resolution No. 12	Defined by NIT-DICLA-026
Periodicity	Annual	Risk-based planning (typically 2–4 years)
Provider competency assessment	Recognition by RIBPG Steering Committee	Compliance with ISO/IEC 17043
Planning	Not explicitly required	Participation plan required
Selection criteria	Not specified	Risk-based criteria
Performance monitoring	Critical review of results	ritical review, planning, and risk analysis

source: prepared by the authors

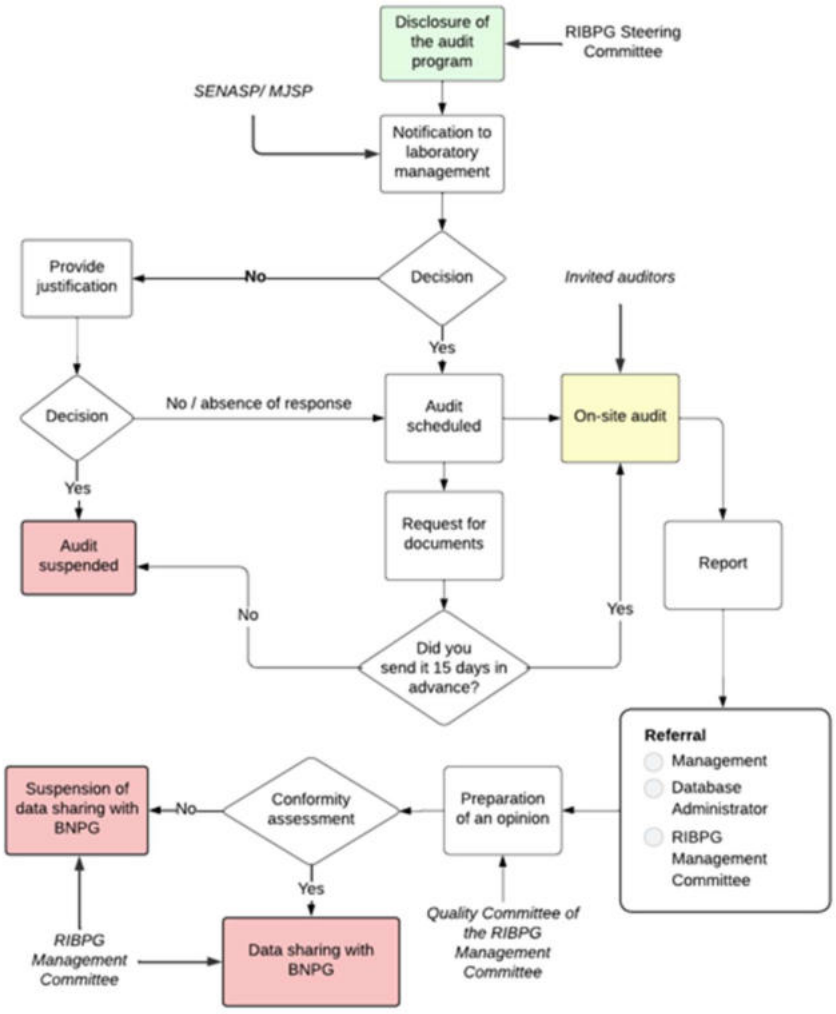
B. External assessments

External assessments correspond to conformity assessment activities conducted by organizations external to the laboratory. Within the proposed QMIP framework, two principal external assessment mechanisms applicable to Brazilian forensic genetics laboratories were identified: (i) conformity audits coordinated by the MJSP within the RIBPG framework and (ii) accreditation assessments conducted by the CGCRE, the accreditation body of INMETRO.

- MJSP conformity audit

The conformity audit process coordinated by the MJSP aims to evaluate laboratory compliance with the technical and management requirements established by Resolution No. 12. This process is directly associated with authorisation for sharing genetic profiles within the BNPG. The audit cycle is coordinated by the National Secretariat of Public Security (SENASP/MJSP) and includes document review, on-site evaluation, identification of nonconformities, preparation of technical reports, and decision-making by the RIBPG Steering Committee. Nonconformities considered capable of compromising the reliability of laboratory results may result in temporary suspension of genetic profile sharing with the BNPG. Figure 2 presents a schematic representation of the MJSP conformity audit process applicable to forensic genetics laboratories participating in the RIBPG.

FIGURE 2 – MJSP CONFORMITY AUDIT PROCESS FOR RIBPG LABORATORIES

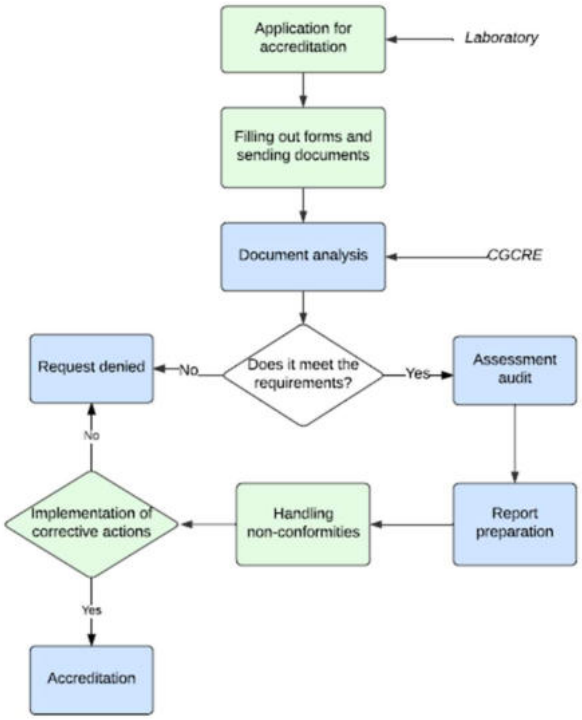


source: prepared by the authors

- CGCRE accreditation assessments

In Brazil, accreditation according to ISO/IEC 17025 is conducted by the CGCRE, the accreditation body of INMETRO. Through this process, laboratories may obtain formal recognition of their technical competence to perform specific forensic analyses within a defined scope of accreditation. Unlike the conformity audits conducted within the RIBPG framework, participation in the CGCRE accreditation process is voluntary and initiated by the laboratory itself. Although accreditation is not mandatory for forensic genetics laboratories in Brazil, it has been recognised internationally as an important mechanism for strengthening reliability, traceability, and confidence in forensic laboratory results. The accreditation assessment process includes document review, on-site assessments, evaluation of technical competence, analysis of nonconformities, implementation of corrective actions, and follow-up assessments when necessary. The process is conducted by qualified CGCRE assessors and follows the requirements established in ISO/IEC 17025 and related accreditation guidance documents. Figure 3 presents a schematic representation of the CGCRE accreditation assessment process.

FIGURE 3 – CGCRE ACCREDITATION ASSESSMENT PROCESS FOR FORENSIC LABORATORIES



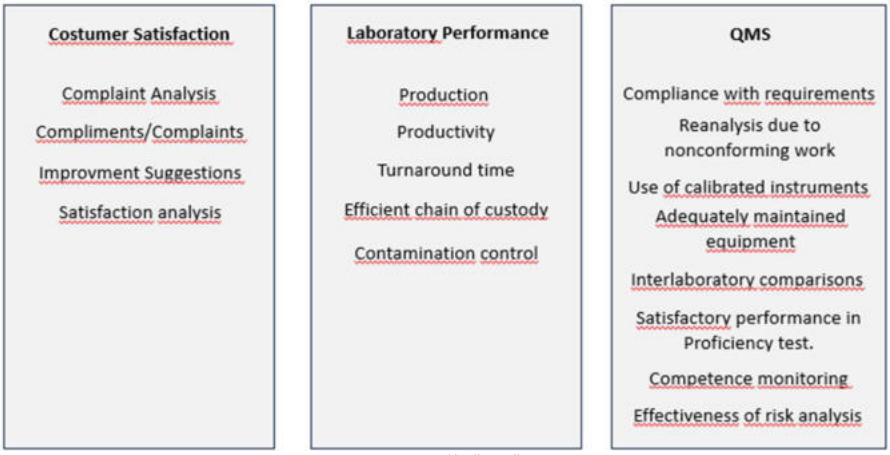
in blue: activity carried out by CGCRE. in green: activity carried out by the laboratory.

source: prepared by the authors

IV. PERFORMANCE EVALUATION

Performance indicators is an important components of QMS and may support monitoring of laboratory activities, assessment of operational performance, and identification of improvement opportunities. Within forensic genetics laboratories, performance evaluation may involve indicators related to analytical quality, operational efficiency, turnaround time, customer satisfaction, and conformity with quality objectives. The definition of performance indicators should consider the objectives of the QMS, operational characteristics of the laboratory, and the needs of stakeholders involved in criminal justice activities. In addition to analytical reliability, forensic laboratories are expected to provide timely and traceable results capable of supporting criminal investigations and judicial procedures. Figure 4 presents a conceptual representation of performance indicators applicable to forensic genetics laboratories within the proposed QMIP framework.

FIGURE 4 – SUGGEST PERFORMANCE INDICATORS FOR FORENSIC GENETICS LABORATORIES



source: prepared by the authors

The indicators presented in Figure 4 include aspects related to personnel competence, facilities and environmental conditions, equipment management, analytical processes, and reporting activities. The combined evaluation of these indicators may contribute to monitoring laboratory performance and identifying opportunities for continuous improvement.

The proposed framework does not establish fixed performance metrics or mandatory targets, since these parameters may vary according to laboratory structure, operational scope, and institutional objectives. Consequently, each laboratory should define its own monitoring criteria, evaluation frequency, and performance targets according to its operational context and quality management strategy.

V. Discussion

The increasing incorporation of DNA evidence into criminal investigations has expanded the importance of quality assurance mechanisms within forensic genetics laboratories. In Brazil, this scenario is directly associated with the expansion of the RIBPG and the growing use of the BNPG in criminal investigations. According to the most recent RIBPG report, the BNPG currently contains more than 190,000 registered genetic profiles and has contributed to thousands of investigative matches nationwide (RIBPG, 2023).

Within this context, the reliability and traceability of laboratory results become central elements for the maintenance of confidence in forensic evidence and for reducing risks associated with analytical errors or misinterpretation of results. Previous studies have highlighted that the quality management of forensic laboratories has progressively evolved through the development of professional networks, harmonization initiatives, accreditation systems, and standardization of methodologies (Heavey et al., 2023). In the Brazilian context, the RIBPG has partially fulfilled this coordinating role by establishing minimum technical and management requirements for participating laboratories.

The proposed QMIP framework was structured to integrate the Brazilian regulatory requirements established by Resolution No. 12 of the RIBPG Steering Committee with internationally recognised quality management principles derived from ISO/IEC 17025 and related forensic guidance documents. Unlike existing Brazilian approaches, which are primarily focused on compliance requirements for data sharing with the BNPG, the proposed framework seeks to organize internal and external assessment activities into a broader quality monitoring structure based on continuous improvement principles.

This approach presents similarities with quality systems adopted in other forensic contexts. In the United States, for example, forensic genetics laboratories commonly operate under ISO/IEC 17025 accreditation combined with additional requirements established by the Federal Bureau of Investigation (FBI) Quality Assurance Standards for Forensic DNA Testing Laboratories and DNA Databasing Laboratories (FBI, 2025). Similarly, European forensic networks such as ENFSI have developed recommendations focused on harmonisation of DNA database management and quality assurance practices across member laboratories (ENFSI, 2017). In both contexts, accreditation systems are associated not only with technical competence, but also with standardized monitoring procedures, corrective action systems, and continuous quality improvement mechanisms.

The Brazilian system presents some similarities with these international models, particularly regarding the use of external conformity assessments and proficiency testing activities. However, important differences remain. Unlike the American and several European systems, accreditation according to ISO/IEC 17025 is not mandatory for Brazilian forensic genetics laboratories. In addition, the Brazilian system is currently structured around two partially independent mechanisms: the conformity audits conducted within the RIBPG framework and the voluntary accreditation assessments conducted by CGCRE.

The coexistence of these two systems may contribute to duplication of assessment activities and differences in monitoring practices among laboratories. The proposed QMIP framework attempts to address this issue by organizing these mechanisms into an integrated conceptual structure that incorporates internal monitoring activities, proficiency testing participation, conformity assessments, corrective actions, and performance evaluation indicators.

Another relevant aspect identified in this study concerns the heterogeneity of QMS implementation among Brazilian forensic genetics laboratories. The data presented in Table 2 indicate that most laboratories currently operate exclusively under the requirements of Resolution No. 12, while accreditation according to ISO/IEC 17025 remains limited. Similar implementation challenges have been reported in other forensic systems, particularly regarding resource limitations, personnel training, infrastructure differences, and the operational complexity associated with accreditation processes (Pádár et al., 2015; Nogel et al., 2019).

The implementation of harmonised quality systems within the Brazilian public security structure may also face challenges associated with regional disparities in infrastructure, personnel availability, financial resources, and institutional priorities among different federative units. These factors may directly influence the feasibility and sustainability of accreditation and continuous monitoring activities. Consequently, the implementation of a nationwide QMIP framework would likely require gradual adaptation according to the operational maturity and resource availability of individual laboratories.

Performance evaluation mechanisms also represent an important component of the proposed framework. Previous studies have indicated that performance indicators may support monitoring of laboratory activities, identification of operational vulnerabilities, and evaluation of corrective actions within forensic quality systems (Heavey et al., 2023). In forensic genetics laboratories, performance evaluation may involve indicators related to analytical quality, turnaround time, contamination events, corrective actions, customer satisfaction, chain of custody management, and proficiency testing performance.

In this context, some indicators relevant to forensic activities are not explicitly detailed within ISO/IEC 17025 and may require complementary monitoring approaches adapted to forensic operational risks. Butler and Willis (2020) and Butler (2023), for example, highlight the relevance of contamination monitoring, interpretation consistency, and chain of custody management as critical aspects of forensic DNA quality systems.

This study presents some limitations that should be considered. The proposed QMIP corresponds to a conceptual framework developed through literature review and documentary analysis and was not empirically validated through pilot

implementation, expert consultation, or comparative operational assessment. Consequently, the applicability and effectiveness of the proposed structure within different forensic laboratory contexts remain to be evaluated in future studies.

Despite these limitations, the study contributes to the discussion regarding quality management in forensic genetics laboratories by systematising the Brazilian regulatory framework and organising internal and external assessment mechanisms into a structured conceptual model. In addition, the study provides a comparative perspective between the Brazilian system and international forensic quality management approaches, highlighting operational and regulatory aspects that may support future harmonisation initiatives within the Brazilian forensic genetics network.

VI. Final considerations

This study aimed to describe a structured QMIP framework for Brazilian forensic genetics laboratories based on internal and external assessment mechanisms. The proposed framework was developed through bibliographic review and documentary analysis of national and international standards, guidelines, and regulatory documents related to quality management and conformity assessment in forensic laboratories. The proposed model was structured considering the requirements established in Resolution No. 12 of the RIBPG Steering Committee and ISO/IEC 17025, integrating continuous monitoring activities, periodic internal assessments, proficiency testing participation, external conformity assessments, and performance evaluation mechanisms within a continuous improvement perspective based on the PDCA cycle.

Within the proposed framework, internal assessment activities include continuous monitoring practices and periodic evaluations conducted under the responsibility of the laboratory. These activities are intended to support conformity assessment, monitoring of analytical performance, identification of nonconformities, and continuous improvement processes. External assessments include both the conformity audits coordinated by the MJSP within the RIBPG framework and the accreditation assessments conducted by CGCRE according to ISO/IEC 17025.

The study also identified important differences between the Brazilian forensic genetics quality system and international models adopted in Europe and the US, particularly regarding the mandatory nature of accreditation systems and the level of harmonisation among laboratories. While Brazil has established minimum requirements for participation in the RIBPG, the implementation of accreditation according to ISO/IEC 17025 remains limited among Brazilian forensic genetics laboratories.

The results additionally demonstrate the heterogeneity of QMS implementation across Brazilian laboratories, suggesting that differences in infrastructure, personnel availability, operational maturity, and resource allocation may influence the implementation and maintenance of quality management systems. These factors should be considered in future initiatives aimed at harmonising quality management practices within the Brazilian forensic genetics network.

This study presents some limitations. The proposed QMIP corresponds to a conceptual framework derived from literature review and documentary analysis and was not empirically validated through pilot implementation, expert consultation, or comparative operational analysis. In addition, the study did not include direct evaluation of laboratory routines, organisational culture, or operational barriers associated with accreditation and quality monitoring activities within individual laboratories.

Future studies may contribute to the development of the proposed framework through empirical validation approaches, including case studies, expert consultation panels, comparative analyses between forensic systems, and pilot implementation in forensic laboratories. Further investigation regarding operational challenges associated with accreditation, quality monitoring, and performance evaluation in the Brazilian public security system may also support the development of more harmonised and sustainable quality management strategies for forensic genetics laboratories.

Despite these limitations, the present study contributes to the discussion regarding quality management in forensic genetics laboratories by systematising the Brazilian regulatory framework and proposing an integrated conceptual structure for quality monitoring and conformity assessment activities within the context of the RIBPG.

References

ANAB (2023). *Accreditation manual for forensic laboratories, forensic inspection bodies, and property and evidence control units*. ANSI National Accreditation Board.

Bezerra, L. S.A., Souza-Neto, A. C., Chaves, P. B., and Poltronieri, C. F. (2025). Implementation of ISO/IEC 17025 in forensic genetics laboratories in Brazil. *Forensic Science International*, Vol. *371*, pp. 112489.

BRASIL (1941). Decreto-Lei nº. 3.689, de 3 de Outubro de 1941. Código de Processo Penal.

Bittencourt, E.A. and Iwamura, E.S.M. (2019), "Brazilian DNA database—Establishment, legislation and accreditation", *Forensic Science International: Genetics Supplement Series*, Vol. 7 No. 1, pp. 422–423.

Butler, J.M. (2023), "Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019–2022", *Forensic Science International: Synergy*, Vol. 6.

Butler, J.M. and Willis, S. (2020), "Interpol review of forensic biology and forensic DNA typing 2016–2019", *Forensic Science International: Synergy*, Vol. 2, pp. 352–367.

Coordenação Geral de Acreditação. (2016), *DOQ-Cgcre-084: tradução brasileira do documento ILAC-G19:08/2014 Módulos de um processo forense*, 41 f., Instituto Nacional de Metrologia, Qualidade e Tecnologia, available at: http://www.inmetro.gov.br/credenciamento/organismos/doc_organismos.asp?tOrganismo=CalibEnsaios (accessed 10 July 2023).

CGCRE. (2023), *Manual da Qualidade da CGCRE*, Revisão 27, 22 f., Coordenação Geral de Acreditação, available at: https://www.gov.br/cgcre/manual_cgcre_rev_27 (accessed 10 July 2023).

Da Silva Junior, R.C., et al. (2020), "Development of DNA databases in Latin America", *Forensic Science International*, Vol. 316.

Dirnagl, U., Kurreck, C., Castaños-Vélez, E. and Bernard, R. (2018), "Quality management for academic laboratories: burden or boon?", *EMBO reports*, Vol. 19 No. 11.

ENFSI. (2017), *DNA database management review and recommendations*, European Network of Forensic Science Institutes, DNA Working Group.

FBI. (2025), *Quality Assurance Standards for Forensic DNA Testing Laboratories*, Federal Bureau of Investigation.

García Merlin, L.P. (2023), "La calidad en los laboratorios genéticos forenses", *Ciencia Latina Revista Científica Multidisciplinar*, Vol. 7 No. 1, pp. 5578–5598.

Heavey, A.L., Turbett, G., Houck, M. and Lewis, S. (2023), "Management and disclosure of quality issues in forensic science: A survey of current practice in Australia and New Zealand", *Forensic Science International: Synergy*, Vol. 7.

ISO/IEC (2017). ISO/IEC 17025: *General requirements for the competence of testing and calibration laboratories*. Geneva, Switzerland: ISO.

ISO/IEC (2023). ISO/IEC 17043: *Conformity assessment — General requirements for the competence of proficiency testing providers*. Geneva, Switzerland: ISO.

ISO (2018). ISO 19011: *Guidelines for auditing management systems.* Geneva, Switzerland: ISO.

INMETRO. (2023), *Sistema de Consulta aos Escopos de Acreditação dos Laboratórios de Análises Clínicas (ISO 15189:2015) e Laboratórios de Ensaio (ISO/IEC 17025:2017) Acreditados (RBLE)*, Instituto Nacional de Metrologia, Qualidade e Tecnologia, available at: http://www.inmetro.gov.br/laboratorios/rble/2023 (accessed 13 July 2023).

INMETRO. (2021), *Norma nº NIT-DICLA-026: Requisitos para a participação de laboratórios em atividades de ensaio de proficiência*, Instituto Nacional de Metrologia, Qualidade e Tecnologia, Rio de Janeiro.

ILAC. (2014), *ILAC G19:08/2014 Modules in a Forensic Science Process*, International Laboratory Accreditation Cooperation.

INTERPOL. (2015), *Best Practice Principles: Recommendations for the Establishment of a National DNA Database*, DNA Monitoring Expert Group, International Criminal Police Organization, available at: https://www.interpol.int (accessed 25 July 2023).

Karthiyayini, N. and Rajendran, C. (2017), "Critical factors and performance indicators: accreditation of testing- and calibration-laboratories", *Benchmarking*, Vol. 24 No. 7, pp. 1814–1833.

Nogel, M., Czebe, A., Kovács, G. and Pádár, Z. (2019), "A work in progress - accreditation of forensic DNA laboratories as a part of the European Forensic Science Area 2020 (EFSA 2020) concept", *Forensic Science International: Genetics Supplement Series*, Vol. 7 No. 1, pp. 836–837.

Pádár, Z., Nogel, M. and Kovács, G. (2015), "Accreditation of forensic laboratories as a part of the 'European Forensic Science 2020' concept in countries of the Visegrad Group", *Forensic Science International: Genetics Supplement Series*, Vol. 5, pp. e412–e413.

Pillai, S., Calvert, J. and Fox, E. (2022), "Practical considerations for laboratories: Implementing a holistic quality management system", *Frontiers in Bioengineering and Biotechnology*, Frontiers Media S.A.

Prodanov, C.C. and Freitas, E.C. (2013), *Metodologia do trabalho científico e técnicas da pesquisa e do trabalho acadêmico*, 2nd ed., Universidade Fevale, Novo Hamburgo.

RIBPG. (2019), *Resolução n. 12, de 09 de agosto de 2019. Estabelece os requisitos técnicos para a realização de auditorias*, available at: https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/ribpg/resolucoes (accessed 25 July 2023).

RIBPG. (2022), *Manual de Procedimentos Operacionais (versão 5) – Resolução n. 17*, Comitê Gestor da RIBPG, Brasília, available at: https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/ribpg/manual (accessed 25 July 2022).

RIBPG. (2023), *XVIII Relatório da Rede Integrada de Bancos de Perfis Genéticos (Maio/2023)*, Comitê Gestor da RIBPG, Brasília, available at: https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/ribpg/relatorio (accessed 25 July 2023).

Sala, D. (2018), "A perícia criminal: evidências, profissional perito e nulidade pericial – uma revisão literária", *Revista Brasileira de Criminalística*, Vol. 7 No. 3, pp. 28–31.

Triverio, S.C. and Márquez, M.C. (2022), "The need for cross-border exchange of genetic data for criminal investigation purposes in Latin America: implementation challenges", *Revista Espanola de Medicina Legal*, Ediciones Doyma, S.L.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

The Integration of Big Data Analytics and Lean Six Sigma: A Systematic Literature Review

Rafaela Gaeta [1], Antonio Almeita Junior [1], João Alfineto [1] and Fabiane Lizarelli [1]

[1] Federal University of São Carlos, Brazil; fabiane@dep.ufscar.br

Abstract*— The integration of Big Data Analytics (BDA) with Lean Six Sigma (LSS) has gained increasing attention in the context of Industry 4.0, as organizations seek more data-driven approaches to improve operational performance and support continuous improvement initiatives. Despite the growing interest in this topic, the literature remains fragmented and lacks a structured synthesis of how BDA supports traditional LSS methodologies. In response, this study conducts a systematic literature review to examine the relationship between Big Data Analytics and Lean Six Sigma and to identify the mechanisms through which analytics capabilities enhance continuous improvement practices. The review followed a structured protocol based on the methodology proposed by Tranfield et al., using the Web of Science and Scopus databases. After applying predefined inclusion and exclusion criteria and following the PRISMA screening process, 19 articles were selected for in-depth analysis. The results reveal that BDA significantly strengthens Lean Six Sigma by expanding the analytical capabilities of traditional improvement tools and enabling their evolution toward digitally enabled solutions such as MSA 4.0, dynamic value stream mapping, and predictive maintenance systems. In addition, BDA enhances the different phases of the DMAIC cycle by enabling automated data collection, advanced root-cause analysis through machine learning, and real-time monitoring of improvement initiatives. The findings also indicate that BDA reinforces both technical and social Lean practices by improving process monitoring, supplier and customer integration, and data-driven decision making. Overall, the integration between BDA and LSS contributes to improved operational performance, greater process visibility, and more proactive continuous improvement systems, highlighting the role of analytics as a key enabler in the evolution toward LSS 4.0.

Keywords – Lean Six Sigma, Big Data Analytics, Systematic Literature Review, Integration

I. Introduction

The integration of Big Data Analytics (BDA) and Lean Six Sigma (LSS) has become increasingly relevant in the context of Industry 4.0, as it promotes a structural evolution in continuous improvement methodologies. In today's highly competitive business environment, shaped by rapid technological development and globalization, organizations are under increasing pressure to improve productivity, quality, and operational efficiency (Schumacher et al., 2016; Pagliosa et al., 2019). In this context, LSS has gained prominence as a well-established process improvement methodology capable of delivering improved outcomes such as cost reduction, quality enhancement, and higher customer satisfaction (Macias-Aguayo et al., 2022). LSS combines Lean management principles, which focus on efficiency and the elimination of waste and non-value-added activities, with Six Sigma's statistical and analytical methods for defect reduction and systematic problem solving (Snee, 2010; Cherrafi et al., 2017; Belhadi et al., 2020, 2023; Skalli et al., 2024).

While traditional LSS has long been recognized for its effectiveness in reducing process variability and operational waste, its implementation often relies on manual data collection and time-consuming analyses that may limit responsiveness in highly dynamic environments (Pongboonchai-Empl et al., 2024; Sony et al., 2019). In this context, Big Data and Big Data Analytics (BDA) have been highlighted in the literature as key enabling technologies capable of expanding the analytical capabilities of LSS (Gupta et al., 2019; Belhadi et al., 2020; Tay and Loh, 2021; Zulfiqar et al., 2024; Wankhede et al., 2025). BDA provides advanced analytical techniques, such as anomaly detection, risk assessment, trend analysis, and predictive analytics, allowing organizations to process large volumes of heterogeneous data and generate more robust evidence-based insights (Gupta et al., 2020; Khattab, 2025; Maia et al., 2023).

The integration between BDA and LSS therefore has the potential to significantly enhance continuous improvement initiatives. By enabling real-time monitoring, predictive maintenance, and event-based inspections, BDA can strengthen the analytical foundation of LSS projects and improve decision-making across different organizational processes (Belhadi et al., 2020, 2023; Fayyaz et al., 2024; Tissir et al., 2024). Moreover, the combination of BDA and LSS allows organizations to identify patterns, trends, and opportunities for improvement that would be difficult to detect using traditional statistical tools alone (Belhadi et al., 2023). As a result, this integration may improve operational performance while also supporting broader sustainability objectives, including resource efficiency and environmental performance (Belhadi et al., 2020; Ibrahim and Kumar, 2024).

Despite the strategic relevance of this integration, the current literature still presents important gaps. Several studies highlight the growing interest in combining digital technologies with continuous improvement methodologies; however, empirical evidence demonstrating how this integration operates in practice remains limited (Pongboonchai-Empl et al., 2024; Sordan et al., 2025). Much of the existing research remains conceptual and provides only high-level discussions of Lean Six Sigma 4.0, often without detailing which Industry 4.0 technologies support specific LSS activities or how they can be operationalized across the DMAIC cycle (Maia et al., 2023; Zulfiqar et al., 2023). In addition, many traditional Six Sigma statistical tools were originally designed for small and structured datasets and may face limitations when applied in Big Data environments, highlighting the

need for new analytical approaches capable of supporting LSS in data-intensive contexts (Escobar et al., 2023; Koppel and Chang, 2021).

In this context, synthesizing and structuring the knowledge on the relationship between BDA and LSS becomes essential to support both researchers and practitioners in the digital transformation of continuous improvement programs (Gupta et al., 2020; Sordan et al., 2022). Such efforts allow the identification of emerging critical success factors for LSS implementation in digitally enabled environments, including BDA capabilities, agile manufacturing, and innovation-oriented practices (Sousa et al., 2023). By consolidating the existing evidence on how Big Data capabilities influence Lean technical and social practices, this research contributes to clarifying the mechanisms through which digital technologies enhance continuous improvement initiatives and support sustainable operational excellence (Clancy et al., 2023; Mesquita et al., 2023).

Although several studies recognize the potential of BDA to enhance Lean Six Sigma initiatives, the existing literature remains fragmented and lacks a structured synthesis of how BDA supports LSS tools, DMAIC phases, organizational practices, and the benefits associated with this integration. In response to this gap, this study aims to systematically review and synthesize the literature on the relationship between Big Data Analytics and Lean Six Sigma. Specifically, the study identifies how BDA enhances traditional Lean Six Sigma tools, examines how digital analytics support the different phases of the DMAIC cycle, analyses how BDA strengthens Lean technical and social practices, and synthesizes the main operational and organizational benefits associated with this integration. By structuring the existing knowledge in these dimensions, this research contributes to clarifying the mechanisms through which Big Data capabilities enable the evolution of Lean Six Sigma toward more data-driven and digitally enabled continuous improvement systems.

II. Theoretical Background

A. Lean Six Sigma

LSS is recognized as a holistic methodology that integrates two complementary approaches to continuous improvement: Lean and Six Sigma (Cherrafi et al., 2017; Sordan et al., 2020; Belhadi et al., 2023). Lean principles focus on the elimination of waste and the improvement of process flow, whereas Six Sigma emphasizes statistical rigor to reduce variation and defects (Cherrafi et al., 2017; Srijithesh et al., 2024). In practice, Lean promotes efficiency, while Six Sigma enhances quality through data-driven problem solving (Singh and Rathi, 2019; Sim et al., 2022; Asif, 2021). The integration of these two philosophies results in a powerful methodology for improving competitiveness and organizational performance (Singh and Rathi, 2019; Sunder et al., 2018), which explains its continued relevance across different sectors (Asnaashari and Khodabandehlou, 2023).

LSS contributes to operational excellence by systematically eliminating non-value-added activities and defects while promoting speed, accuracy, and continuous improvement (Cherrafi et al., 2017; Belhadi et al., 2020, 2023; Tissir et al., 2024). As a data-driven methodology, its effectiveness depends on the availability of reliable and timely data, as well as on the application of robust analytical tools (Arcidiacono and Pieroni, 2018; Skalli et al., 2024; Kumar et al., 2024).

By reducing variation and improving process efficiency, LSS generates tangible organizational benefits, including cost reduction, higher quality levels, improved customer satisfaction, and increased operational capacity (Abliwi et al., 2015; Pongboonchai-Empl et al., 2023; Citybabu and Yamini, 2024). However, the growing volume and complexity of data in modern organizations introduce new challenges for LSS implementation, requiring adaptations in traditional practices to effectively manage and analyze large datasets (Pongboonchai-Empl et al., 2023).

B. Integration between LSS and BDA

Big Data refers to large amounts of data collected from multiple sources and characterized by high volume, variety, velocity, and uncertainty, which must be transformed into actionable insights through BDA (Belhadi et al., 2019, 2023). BDA encompasses analytical methods such as statistics, simulations, and optimization techniques to support decision-making processes (Arunachalam et al., 2018; Deniswara et al., 2020; Belhadi et al., 2019, 2023). It enables the identification of patterns and trends beyond the reach of traditional analytical approaches (Gupta et al., 2020), improving resource utilization and supporting strategic decision-making (Belhadi et al., 2020). BDA capability refers to an organization's ability to transform data into strategic insights through the effective deployment of tangible resources (e.g., infrastructure), human resources (e.g., technical skills), and intangible resources (e.g., data-driven culture and organizational learning) (Akter et al., 2016; Gupta and George, 2016).

LSS is inherently data-driven (Zwetsloot et al., 2018; Skalli et al., 2024; Kumar et al., 2024), requiring data collection and analysis throughout the DMAIC cycle. Consequently, the ability to manage and analyze large volumes of data becomes increasingly important (Gupta et al., 2019). Integrating BDA into LSS practices modernizes traditional improvement approaches and provides a stronger analytical foundation for more informed decision-making (Belhadi et al., 2020). For example, BDA can automate problem identification and significantly reduce analysis time in LSS projects (Koppel and Chang, 2021; Pongboonchai-Empl et al., 2024).

LSS can benefit substantially from BDA, and in some cases this relationship becomes mutually reinforcing (Pongboonchai-Empl et al., 2024). The DMAIC methodology can provide the structured framework and analytical tools required for the effective application of BDA within improvement initiatives (Shivajee et al., 2019). When combined with appropriate infrastructure, large volumes of data enable more reliable, robust, and predictive analyses, making BDA an important element for strengthening LSS practices (Zwetsloot et al., 2018; Gupta et al., 2019).

This synergy not only improves operational efficiency but also facilitates the implementation of continuous improvement programs, contributing to cost reduction and sustainable organizational growth (Belhadi et al., 2020). The positive and significant impact of BDA on LSS strengthens companies' ability to adapt and respond to contemporary operational challenges (Rejikumar et al., 2020). Despite the growing relevance of BDA, there is still a considerable lack of empirical studies

examining its relationship with LSS (Rejikumar et al., 2018; Belhadi et al., 2020; Fayyaz et al., 2024), particularly in developing countries, which limits the broader understanding of its impact in these contexts (Belhadi et al., 2020).

III. METHOD

This paper conducts a Systematic Literature Review (SLR) on studies addressing the integration of BDA and LSS. The main objective of the SLR is to establish a theoretical and scientific foundation by identifying the state of the art on the topic (Levy and Ellis, 2006). According to Walsham (2006), literature reviews should receive careful attention and be conducted in a systematic and rigorous manner to ensure the reliability and robustness of the research process. In this sense, SLR enables the identification of research gaps, the consolidation of existing knowledge, and the generation of insights into potential future research directions (Thomé et al., 2018).

To achieve the objectives of the review, a structured protocol was developed and followed throughout the research process. The identification and analysis of the relevant documents were conducted according to the three-stage approach proposed by Tranfield et al. (2003) and Tranfield et al. (2009). The first stage involved planning the review, including the development of the review protocol, the definition of research questions, the selection of keywords and databases, and the establishment of inclusion and exclusion criteria. The second stage consisted of conducting the review, which included the search, screening, and refinement of the documents, as well as the reporting of the number of studies identified and retained at each step. Finally, the third stage focused on reporting the results of the review, presenting and discussing the findings derived from the selected studies.

A. Planning

The Web of Science (WoS) and Scopus databases were selected as the primary sources for identifying the documents analyzed in this research. These databases were chosen due to their relevance and extensive coverage of high-quality scientific publications (Buer et al., 2018). WoS is a multidisciplinary database that provides access to high-impact research across several fields, including social sciences, arts, and humanities (Pranckutė, 2021). Furthermore, Scopus has evolved as a complementary database to WoS, expanding the coverage of scientific publications and improving access to peer-reviewed research (Modak et al., 2020). The research questions addressed in this study were introduced in the Introduction section.

To support the identification of keywords and the formulation of the search strategy, a preliminary review of the literature on Big Data Analytics (Belhadi et al., 2019; 2023; Deniswara et al., 2020) and Lean Six Sigma (Antony, 2011; Sony et al., 2020) was conducted. Based on this exploratory analysis, a search string was developed to ensure that the results were sufficiently comprehensive while remaining focused on the research topic (Thomé et al., 2016). The final search string applied in the databases was: ("Big Data Analytics" OR "BDA") AND ("Lean Six Sigma" OR "Lean Sigma" OR "LSS" OR "Lean 6 Sigma"). No time restriction was applied, and studies published up to 2025 were considered in the analysis.

To address this objective, the following research questions were defined:

RQ1: How has the scientific literature on the integration between Big Data Analytics and Lean Six Sigma evolved over time and across publication sources?

RQ2: How does Big Data Analytics support traditional Lean Six Sigma tools used in quality improvement initiatives?

RQ3: How does Big Data Analytics enhance the different phases of the DMAIC cycle in Lean Six Sigma projects?

RQ4: How does Big Data Analytics strengthen Lean Six Sigma technical and social practices within organizations?

RQ5: What benefits are associated with the integration between Big Data Analytics and Lean Six Sigma?

As illustrated in Fig. 1, one of the key steps in the planning phase of the SLR is the definition of inclusion and exclusion criteria for the studies to be analyzed. The criteria adopted in this study are presented in Table 1. This study included a quality assessment of the selected articles, focusing on the clarity and structure of the research methods reported in each study. The evaluation considered whether the methodological procedures were adequately described and whether the research design presented a coherent sequence of steps that would allow understanding and potential replication of the study.

TABLE I. INCLUSION AND EXCLUSION CRITERIA

Topic	***Inclusion Criteria***	***Exclusion Criteria***
Duplication	Is not duplicate paper.	Is a duplicate paper.
Language	Be written in English.	Not be written in English.
Source	Journal and conference papers.	Other type of document (e.g., books chapters, editorial).
Theme	The document relate to LSS and BDA	The document is not related to LS and BDA
Focus	The paper deals with the relationships between BDA and LSS	The paper does not address the relationships between BDA and LSS

B. Conducting

The search string was applied to the two predefined databases using the selected keywords in titles, abstracts, and keywords. This process initially identified 50 records. Following the PRISMA approach (Fig. 1), records were screened according to the predefined inclusion and exclusion criteria. First, duplicate records were then removed, resulting in 32 articles for further

assessment. Second, documents other than journal articles, such as conference papers, books, and other non-article publications, were excluded (n=5).

After this stage, articles not published in English were excluded (n=1). Next, the remaining studies were screened based on their titles, abstracts, and keywords in order to verify their alignment with the objective of this review. Studies that did not specifically address the relationship between Big Data Analytics and Lean Six Sigma were excluded. The full texts of the potentially relevant articles were then assessed for eligibility, considering their adherence to the review scope and the consistency of their methodological description. At the end of this process, 19 articles were included in the final sample for analysis. The study selection process was conducted in a transparent and systematic way, following the logic recommended by the PRISMA framework, which supports the identification, screening, eligibility assessment, and inclusion of studies in literature reviews.

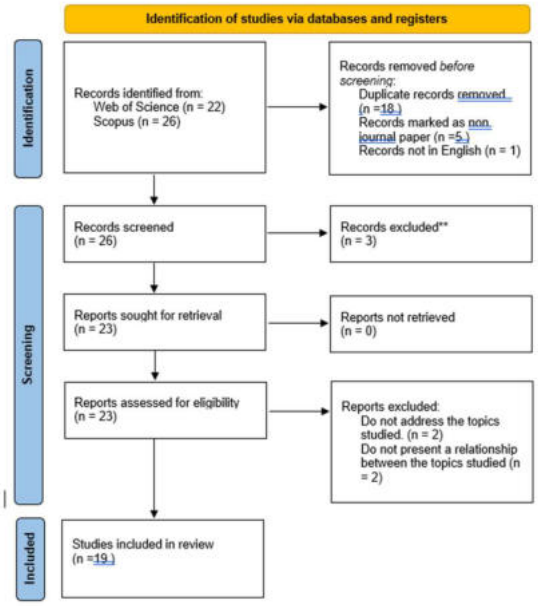


Fig. 1. Methodological Protocol of the Systematic Literature Review

C. Results

The outcomes are presented in Section 4 through a descriptive analysis addressing the initial research question regarding the characterization of the literature on the relationship between BDA and LSS. This analysis considers aspects such as the volume of publications, leading authors, and other relevant characteristics of the selected studies.

In addition to the descriptive analysis, a thematic analysis was conducted through an in-depth examination of the selected articles. Thematic analysis enables researchers to identify, analyze, and report patterns within qualitative data (Snyder, 2019). This approach aimed to provide insights into the research questions and to better understand the relationship between Big Data Analytics and Lean Six Sigma. The selected articles were read multiple times to identify recurring meanings and patterns. Qualitative data were extracted individually from each article and organized in an Excel spreadsheet, with each finding recorded in a separate cell. The spreadsheet included information such as the article title, year of publication, journal, authors, article type, and other relevant characteristics related to the implications of LSS and BDA (Tranfield et al., 2003).

IV. Results

The results of the systematic literature review are organized into four main analytical dimensions. First, a descriptive analysis is presented to characterize the evolution of the literature and the main publication sources related to the integration between BDA and LSS. Subsequently, the review examines how BDA supports traditional Lean Six Sigma tools, highlighting the technological mechanisms that enhance their analytical and operational capabilities. The analysis then explores how BDA contributes to the different phases of the DMAIC cycle, demonstrating how digital technologies transform traditional improvement processes into more data-driven and predictive systems. Finally, the results synthesize how BDA reinforces Lean Six Sigma organizational practices and the main benefits associated with this integration, including improvements in operational performance, decision-making, and sustainability outcomes.

A. Descriptive Analysis

Publications over time are presented in Fig. 2. The distribution of articles across the years shows a noticeable increase from 2023 onward, indicating the growing academic interest in the integration between BDA and LSS. This trend suggests that the topic is emerging and gaining relevance, highlighting a favourable context for further research and future publications in this area.

The distribution of publication sources is presented in Fig. 3. The *International Journal of Lean Six Sigma* stands out as the journal with the highest number of publications on the topic. In addition, multiple contributions were identified in several other

journals, including *Engineering Management Journal*, *Journal of Cleaner Production*, *Production Planning & Control*, and *The TQM Journal*. The remaining studies are distributed across a variety of journals, indicating that research on the integration between BDA and LSS spans different fields within operations management, quality management, and manufacturing research.

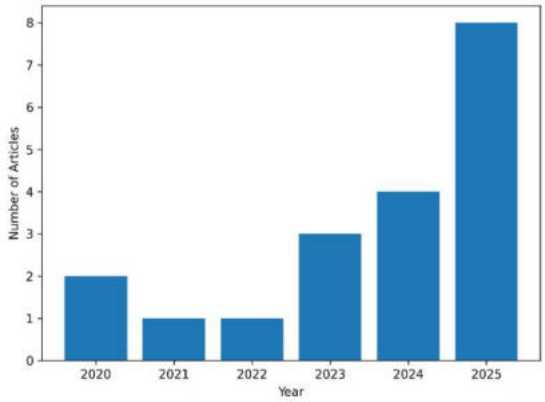


Fig. 2. Articles published over the years.

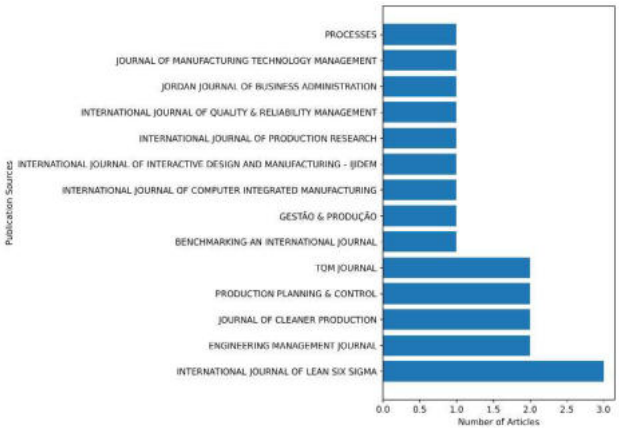


Fig. 3. Distribution of publications sources

B. BDA support for Lean Six Sigma tools

To answer the second Research Question (RQ2), Table II summarizes how BDA) enhances traditional LSS tools by integrating advanced digital technologies and data-driven analytical capabilities into continuous improvement practices. The literature indicates that the incorporation of BDA transforms conventional LSS tools into more dynamic, automated, and predictive systems capable of processing large volumes of structured and unstructured data. Technologies such as artificial intelligence, machine learning, IoT sensors, process mining, and advanced simulation enable real-time monitoring, predictive analysis, and automated decision support across several improvement tools. As a result, traditional approaches such as Measurement System Analysis, Value Stream Mapping, and Poka-Yoke evolve into digitally enabled versions often referred to as LSS 4.0 tools. These developments allow organizations to improve the identification of process deviations, optimize operational parameters, enhance maintenance strategies, and strengthen customer-focused quality analysis. Overall, the integration of BDA significantly expands the analytical capabilities of Lean Six Sigma tools, enabling more accurate, scalable, and real-time continuous improvement initiatives in data-intensive environments.

TABLE II. BDA SUPPORT FOR LEAN SIX SIGMA TOOLS

Lean Six Sigma Tool	***How Big Data Analytics Supports the Tool***	***Key References***
Measurement System Analysis (MSA 4.0)	Artificial Neural Networks (ANN) are used to assess data reliability in Big Data environments where traditional statistical methods (e.g., ANOVA) become ineffective due to extremely large sample sizes.	Pongboonchai-Empl et al. (2024)
Poka-Yoke (Error Proofing 4.0)	AI algorithms, IoT sensors, computer vision, and machine learning enable real-time error detection and automatic blocking of defective parts during the production process.	Pongboonchai-Empl et al. (2024); Sordan et al. (2025)
Value Stream Mapping (Dynamic VSM / VSM 4.0)	BDA, RFID, and simulation tools enable dynamic value stream mapping using real-time data, improving bottleneck identification and lead-time prediction compared to traditional static mapping.	Sordan et al. (2022; 2025); Pongboonchai-Empl et al. (2024)
Cause-and-Effect Tools (Ishikawa, Pareto, Cause-Effect Matrix)	Data mining algorithms such as decision trees (CART/CHAID) automatically identify statistically significant factors influencing defects using historical datasets.	Pongboonchai-Empl et al. (2024)

Lean Six Sigma Tool	*How Big Data Analytics Supports the Tool*	*Key References*
Total Productive Maintenance (TPM)	BDA enables predictive maintenance ("e-maintenance") by analyzing sensor data (e.g., vibration and temperature) to predict equipment failures and improve Overall Equipment Effectiveness (OEE).	Sordan et al. (2022); Khattab (2025)
Design of Experiments (DOE)	Advanced analytics and simulation allow the evaluation of thousands of variable combinations, identifying optimal process parameters faster and at lower cost than traditional physical experiments.	Pongboonchai-Empl et al. (2024)
Kanban / Pull Systems (e-Kanban)	Integration of ERP/MES systems with digital technologies (e.g., QR codes and sensors) enables automated replenishment and dynamic inventory management based on real-time data.	Sordan et al. (2025); Mesquita et al. (2023)
Quality Function Deployment (QFD)	BDA allows the analysis of large volumes of customer data (e.g., social media and online reviews) to identify customer requirements and critical-to-quality (CTQ) attributes for the House of Quality.	Sordan et al. (2022)
SIPOC and Flowcharts	Process mining techniques automatically discover and map processes from system event logs, enabling more accurate identification of suppliers, inputs, processes, outputs, and customers.	Pongboonchai-Empl et al. (2024); Sordan et al. (2022)
Digital Andon	Digital dashboards and mobile notifications provide real-time visualization of production status and anomalies, enabling faster responses and paperless monitoring.	Sordan et al. (2025)
Spaghetti Diagram	Micro-location technologies allow the automated tracking of movements and traffic flows within production environments, enabling more accurate layout analysis and optimization.	Wielki and Kozioł (2019) cited by Sordan et al. (2022)

C. BDA support for DMAIC phases

Fig. 4 summarizes how BDA enhances the different phases of the LSS DMAIC cycle, addressing RQ3 of this study. The literature indicates that the integration of BDA introduces greater analytical depth, automation, and real-time responsiveness across all stages of the improvement process. In the Define phase, advanced analytics techniques support the identification and prioritization of improvement opportunities through the analysis of large datasets and digital customer feedback. During the Measure phase, digital technologies such as IoT sensors and automated data architectures enable real-time data acquisition and improve the reliability and scalability of measurement systems. In the Analyze phase, machine learning algorithms and process mining techniques allow organizations to uncover complex root causes and analyze non-linear relationships that traditional statistical tools may not capture. The Improve phase benefits from predictive and prescriptive analytics, digital twins, and optimization methods that enable organizations to test and implement improvement solutions more efficiently. Finally, in the Control phase, BDA enables continuous monitoring and predictive maintenance through real-time dashboards and advanced analytics, transforming traditional statistical process control into a dynamic and data-driven monitoring system. Overall, the evidence suggests that BDA significantly strengthens the DMAIC methodology by expanding its analytical capabilities and enabling more proactive and intelligent continuous improvement processes.

Fig. 4 presents the proposed BDA-Enhanced DMAIC Framework, which synthesizes the main mechanisms through which Big Data Analytics strengthens Lean Six Sigma practices across the DMAIC cycle. The framework illustrates how BDA, enabling technologies, and analytical methods support each DMAIC phase by increasing automation, analytical depth, predictive capabilities, and real-time responsiveness.

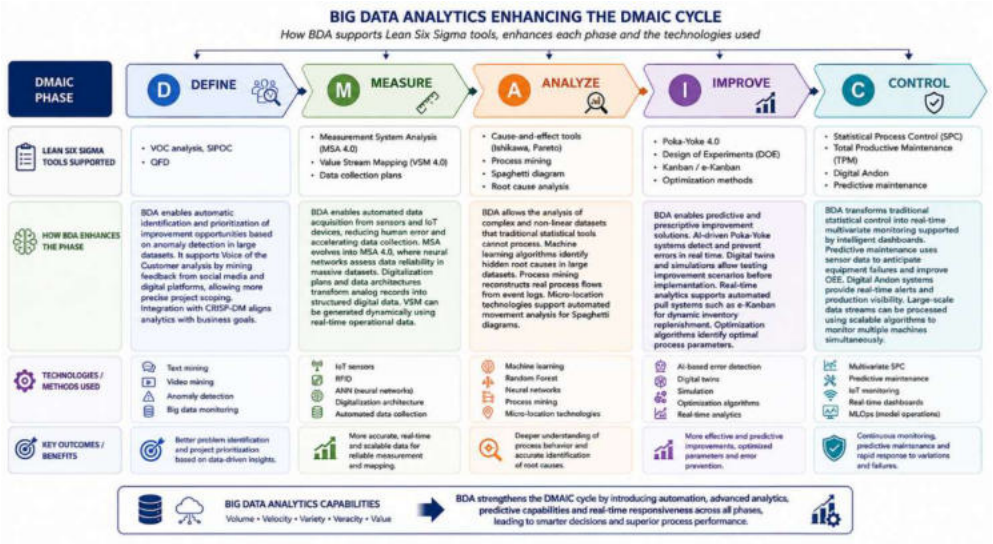


Fig. 4. Conceptual Framework of BDA Integration Across DMAIC Phases

The framework demonstrates that the role of BDA extends beyond traditional data support functions, actively transforming the DMAIC cycle into a more dynamic, intelligent, and data-driven continuous improvement system. While the Define and Measure phases benefit primarily from data acquisition and customer insight generation, the Analyze, Improve, and Control

phases increasingly rely on machine learning, predictive analytics, digital twins, and real-time monitoring technologies to optimize decision-making and operational performance.

D. Integration between BDA and LSS Practices

The integration between BDA and LSS practices (RQ4) enhances both technical (hard) and social (soft) Lean practices, transforming traditional operational practices into dynamic and predictive processes. In the framework proposed by Mesquita et al. (2023), BDA capabilities strengthen Lean practices by enabling organizations to process and analyse large volumes of structured and unstructured data, thereby improving decision-making and operational responsiveness. Within the technical dimension, BDA supports the evolution of traditional Lean Six Sigma practices identified by Costa et al. (2020), such as Statistical Process Control (SPC) and Total Productive Maintenance (TPM).

Advanced analytics enables real-time monitoring, multivariate analysis, and predictive maintenance through the integration of sensor and IoT data, allowing organizations to anticipate deviations and equipment failures before they occur. BDA also enhances practices related to flow management, including continuous flow and pull systems, by enabling dynamic value stream mapping and digital kanban systems based on real-time data. In the social dimension, BDA facilitates stronger customer and supplier integration through advanced data analysis techniques that capture the voice of the customer from digital sources and improve supply chain coordination. Furthermore, analytics-driven decision making strengthens employee involvement and supports more evidence-based continuous improvement initiatives. As a result, the traditional improvement infrastructure of Lean Six Sigma, including structured improvement procedures and performance measurement systems, becomes increasingly data-driven, requiring professionals capable of combining operational excellence expertise with data analytics capabilities.

To clarify how Big Data Analytics enhances Lean Six Sigma practices, Table III summarizes the main LSS constructs proposed by Costa et al. (2020) and the contributions of BDA identified in the literature.

TABLE III. BIG DATA ANALYTICS CONTRIBUTIONS TO LEAN SIX SIGMA PRACTICES

LSS Practices (Costa et al., 2020)	***LSS Dimension***	***Contribution of Big Data Analytics***	***Key References***
Statistical Process Control	Lean Technical Practices	BDA enables real-time multivariate monitoring and predictive quality control, allowing organizations to anticipate deviations and prevent defects rather than reacting to them.	Sordan et al. (2022); Pongboonchai-Empl et al. (2024)
Total Productive Maintenance	Lean Technical Practices	Predictive analytics and sensor data enable condition-based maintenance, allowing early identification of equipment failures and improved equipment availability.	Khattab (2025); Sordan et al. (2022)
Setup Time Reduction	Lean Technical Practices	Machine learning algorithms analyse operational data to optimise machine configurations and reduce changeover times.	Sordan et al. (2025)
Continuous Flow	Lean Technical Practices	BDA enables real-time monitoring of production flows and supports dynamic value stream mapping, improving process synchronization and waste reduction.	Mesquita et al. (2023); Sordan et al. (2025)
Pull Production	Lean Technical Practices	Digital kanban systems and data-driven demand signals improve just-in-time coordination and reduce inventory levels.	Mesquita et al. (2023); Romero et al. (2018)
Supplier Involvement	Lean Social Practices	Data sharing platforms and machine-to-machine communication enhance collaboration and synchronization across supply chains.	Mesquita et al. (2023)
Customer Involvement	Lean Social Practices	Advanced analytics techniques such as text mining and video mining enable deeper analysis of customer feedback and the development of Voice of the Customer 4.0.	Pongboonchai-Empl et al. (2024); Gupta et al. (2020)
Employee Involvement	Lean Social Practices	Data-driven decision making increases transparency and trust in improvement initiatives, strengthening employee engagement and participation.	Belhadi et al. (2020); Khattab (2025)
Six Sigma Structured Improvement	Improvement Infrastructure	Continuous monitoring of large-scale datasets allows organizations to identify improvement opportunities and prioritize projects based on real-time insights.	Koppel & Chang (2021)
Six Sigma Focus on Metrics (SSMET)	Improvement Infrastructure	BDA improves measurement accuracy and enables large-scale performance monitoring through advanced analytics.	Koppel & Chang (2021); Pongboonchai-Empl et al. (2024)
Six Sigma Role Structure (SSROLE)	Improvement Infrastructure	The increasing use of analytics requires Lean Six Sigma professionals to combine operational excellence expertise with data science capabilities.	Maia et al. (2023); Pongboonchai-Empl et al. (2024)

[a.] Sample of a Table footnote. (*Table footnote*)

E. Benefits of the integration between BDA and LSS Practices

The integration of BDA and LSS represents a shift from reactive and manually driven improvement practices toward a proactive, predictive, and data-driven paradigm commonly referred to as LSS 4.0. BDA strengthens traditional LSS methodologies by enhancing analytical capabilities, enabling real-time process monitoring, and supporting more informed decision-making. Based on the literature, its contributions can be grouped into three main mechanisms: the enhancement of traditional improvement tools, the transformation of the DMAIC cycle, and the reinforcement of organizational practices.

First, BDA enhances traditional LSS tools by transforming static analytical methods into dynamic and intelligent systems. Examples include MSA 4.0, which uses artificial neural networks to improve measurement reliability, and Poka-Yoke 4.0, which integrates artificial intelligence and predictive algorithms to prevent defects in real time (Pongboonchai-Empl et al., 2024; Sordan et al., 2025). Likewise, Dynamic Value Stream Mapping combines BDA and IoT technologies to continuously monitor process performance, while predictive maintenance systems use sensor data to anticipate equipment failures (Sordan et al.,

2022, 2025; Khattab, 2025). BDA also strengthens Quality Function Deployment by enabling the analysis of large volumes of customer feedback from digital platforms and social media (Li et al., 2015; Sordan et al., 2022).

Second, BDA transforms the DMAIC cycle by increasing automation, analytical depth, and responsiveness. In the Define phase, data-driven approaches such as MDAIC allow improvement opportunities to be identified through monitoring systems before project definition (Koppel and Chang, 2021). In the Measure phase, digital technologies automate data collection and integration (Clancy et al., 2023). During Analyze, machine learning techniques support root-cause analysis using large volumes of operational data (Gupta et al., 2020; Belhadi et al., 2020). In Improve and Control, digital twins, simulation models, and advanced process monitoring systems enable virtual testing of solutions and continuous control of process performance (Sanchez-Marquez and Jabaloyes Vivas, 2020; Sordan et al., 2025).

Finally, BDA reinforces the organizational dimension of LSS by supporting data-driven cultures and sociotechnical practices. Big Data Analytics Capability facilitates both Lean technical and social practices, including leadership, training, organizational learning, and fact-based decision-making (Mesquita et al., 2023; Belhadi et al., 2020). Recent studies also identify BDA as a critical success factor for LSS implementation, particularly in digitally transforming environments, where it supports change management, sustainability initiatives, and cross-functional collaboration among improvement specialists and data scientists (Sousa et al., 2023; Khattab, 2025; Mehmood et al., 2025).

F. Barriers and Contradictory Evidence in the Integration Between BDA and LSS

Although the literature highlights the potential synergies between BDA and LSS, their successful integration is not guaranteed and remains associated with several organizational, technical, and strategic barriers. Key challenges include poor understanding of data analytics, lack of top management support, high implementation costs, financial constraints, and shortages of qualified personnel, particularly in SMEs (Costa et al., 2020; Sharma and Singh, 2023; Zulfiqar et al., 2023). These findings suggest that successful implementation depends not only on technological infrastructure but also on organizational readiness, managerial commitment, and workforce capabilities.

The literature also presents conflicting evidence regarding the benefits of BDA for Lean and Six Sigma. While many studies report positive effects, Khattab (2025) argued that BDA may negatively affect green lean and quality assurance practices when implemented without an appropriate strategic roadmap. Additional concerns relate to methodological and technical limitations. Traditional statistical methods may become unreliable in Big Data environments due to increased risks of false discoveries (Pongboonchai-Empl et al., 2024), while Six Sigma may be less suitable for highly dynamic and innovation-driven settings (Clancy et al., 2023). Furthermore, digital transformation initiatives often fail when technological adoption is not aligned with fundamental operational improvement principles (Sordan et al., 2025). Data security, cloud reliability, and the limited empirical validation of many LSS 4.0 frameworks remain additional concerns (Gupta et al., 2020; Pongboonchai-Empl et al., 2024). Overall, the literature suggests that realizing the benefits of BDA-LSS integration requires careful alignment between analytical capabilities, operational realities, organizational culture, and strategic objectives.

V. Conclusions

This study systematically reviewed the literature on the integration between BDA and LSS in order to understand how digital analytics capabilities support continuous improvement methodologies in the context of Industry 4.0. By synthesizing the existing body of knowledge, the study addressed five research questions related to the evolution of the literature, the role of BDA in supporting LSS tools, its contribution across the DMAIC phases, its influence on Lean Six Sigma practices, and the benefits associated with this integration. This study contributes by consolidating dispersed evidence into an integrated framework and identifying key barriers, contradictions, and research gaps that remain underexplored in the BDA–LSS literature.

The results indicate that the integration of BDA significantly enhances traditional Lean Six Sigma methodologies. First, BDA expands the analytical capabilities of several LSS tools by transforming conventional methods into more dynamic and data-driven systems. Tools such as Measurement System Analysis, Value Stream Mapping, Poka-Yoke, Kanban systems, and Total Productive Maintenance evolve into digitally enabled versions (e.g., MSA 4.0, VSM 4.0, and predictive maintenance) supported by technologies such as artificial intelligence, IoT sensors, process mining, and advanced analytics. These technologies enable organizations to process large volumes of operational data and improve the identification of process deviations and improvement opportunities.

Second, the findings show that BDA strengthens the entire DMAIC cycle by increasing analytical depth and enabling more automated and predictive decision-making. While advanced analytics supports the identification of improvement opportunities during the Define phase, IoT-enabled data collection enhances the Measure phase by improving data reliability and availability. Machine learning and process mining techniques strengthen the Analyze phase by uncovering complex root causes that traditional statistical tools may fail to detect. In the Improve and Control phases, technologies such as digital twins, predictive maintenance, and real-time monitoring systems allow organizations to implement and monitor improvement solutions more effectively.

Third, the integration between BDA and LSS reinforces both technical and social Lean practices. Advanced analytics improves technical practices such as process monitoring, maintenance management, and flow optimization, while simultaneously strengthening social practices related to customer integration, supplier collaboration, and employee involvement. As a result, continuous improvement initiatives become increasingly data-driven and require new competencies that combine operational excellence expertise with data analytics capabilities.

Overall, the synthesis of the literature suggests that the integration between BDA and LSS contributes to improved operational performance by enabling faster decision-making, higher process transparency, enhanced predictive capabilities, and more effective continuous improvement initiatives. These findings highlight the role of BDA as a key enabler in the evolution of Lean Six Sigma toward what is increasingly described in the literature as Lean Six Sigma 4.0. Despite the growing interest in

BDA-LSS integration, the literature still presents important gaps, including the limited number of empirical validation studies, the predominance of conceptual frameworks, the lack of longitudinal investigations, and the scarcity of evidence regarding implementation in SMEs and real-time industrial environments. However, the review also identified important barriers and contradictory findings regarding the benefits of BDA-LSS integration, indicating that successful implementation depends on the alignment of technological, organizational, and managerial factors

From a theoretical perspective, this study contributes by consolidating fragmented knowledge on the relationship between Big Data Analytics and Lean Six Sigma and by structuring the main mechanisms through which BDA enhances LSS tools, DMAIC phases, and organizational practices. By organizing these contributions into a coherent analytical framework, this research advances the understanding of how digital technologies reshape traditional continuous improvement methodologies. However, the field is still evolving and supported by limited empirical evidence, as many identified studies remain conceptual or exploratory.

From a managerial perspective, the results provide practical insights for organizations seeking to integrate advanced analytics into their improvement initiatives. Managers should recognize that the effective adoption of BDA in Lean Six Sigma programs requires not only investments in digital technologies but also the development of data infrastructures, analytical competencies, and a data-driven organizational culture. The integration between improvement specialists and data science teams becomes increasingly important to fully leverage the potential of BDA in operational environments.

Despite its contributions, this study presents some limitations. The review was limited to articles indexed in two major databases and resulted in a final sample of 19 publications, reflecting the emerging and fragmented nature of the research field. In addition, the study relied on secondary data from the literature and therefore did not empirically test the relationships identified in the conceptual synthesis. Consequently, caution should be exercised when generalizing the findings related to the evolution toward Lean Six Sigma 4.0.

Future research should focus on empirical studies that examine the implementation of BDA-enabled Lean Six Sigma initiatives in organizational contexts. Longitudinal studies could provide deeper insights into how the integration between analytics capabilities and continuous improvement practices evolves over time. Additionally, further research is needed to explore applications in service sectors and digital ecosystems, as well as to develop quantitative models capable of testing the impact of BDA capabilities on Lean Six Sigma performance outcomes.

References

Albliwi, S., Antony, J., Abdul Halim Lim, S. and van der Wiele, T. (2014) 'Critical failure factors of Lean Six Sigma: a systematic literature review', *International Journal of Quality & Reliability Management*, 31(9), pp. 1012–1030.

Akter, S., Wamba, S.F., Gunasekaran, A., Dubey, R. and Childe, S.J. (2016) 'How to improve firm performance using big data analytics capability and business strategy alignment?', *International Journal of Production Economics*, 182, pp. 113–131.

Antony, J. (2011) 'Six Sigma vs Lean: Some perspectives from leading academics and practitioners', *International Journal of Productivity and Performance Management*, 60(2), pp. 185–190.

Arcidiacono, G. and Pieroni, A. (2018) 'The revolution Lean Six Sigma 4.0', *International Journal on Advanced Science, Engineering and Information Technology*, 8(1), pp. 141–149.

Asif, M. (2021) 'Lean Six Sigma institutionalization and knowledge creation: towards developing theory', *Total Quality Management & Business Excellence*, 32(7–8), pp. 811–828.

Asnaashari, H. and Khodabandehlou, F. (2024) 'Lean Six Sigma, effectiveness, and efficiency of internal auditing', *International Journal of Lean Six Sigma*, 15(4), pp. 787–816.

Belhadi, A., Kamble, S.S., Gunasekaran, A., Zkik, K., M, D.K. and Touriki, F.E. (2023) 'A Big Data Analytics-driven Lean Six Sigma framework for enhanced green performance: a case study of chemical company', *Production Planning & Control*, 34(9), pp. 767–790.

Belhadi, A., Mani, V., Kamble, S.S., Khan, S.A.R. and Verma, S. (2024) 'Artificial intelligence-driven innovation for enhancing supply chain resilience and performance under the effect of supply chain dynamism', *Annals of Operations Research*, 333(2), pp. 627–652.

Belhadi, A., Kamble, S.S., Zkik, K., Cherrafi, A. and Touriki, F.E. (2020) 'The integrated effect of Big Data Analytics, Lean Six Sigma and Green Manufacturing on the environmental performance of manufacturing companies: The case of North Africa', *Journal of Cleaner Production*, 252, 119903.

Buer, S.V., Strandhagen, J.O. and Chan, F.T. (2018) 'The link between Industry 4.0 and lean manufacturing: mapping current research and establishing a research agenda', *International Journal of Production Research*, 56(8), pp. 2924–2940.

Cherrafi, A., Elfezazi, S., Govindan, K., Garza-Reyes, J.A., Benhida, K. and Mokhlis, A. (2017) 'A framework for the integration of Green and Lean Six Sigma for superior sustainability performance', *International Journal of Production Research*, 55(15), pp. 4481–4515.

Citybabu, G. and Yamini, S. (2024) 'Lean Six Sigma 4.0 – a framework and review for lean six sigma practices in the digital era', *Benchmarking: An International Journal*, 31(9), pp. 3288–3326.

Clancy, R., O'Sullivan, D. and Bruton, K. (2023) 'Data-driven quality improvement approach to reducing waste in manufacturing', *The TQM Journal*, 35(1), pp. 51–72.

Deniswara, K., Handoko, B.L. and Mulyawan, A.N. (2020) 'Big data analytics: literature study on how big data works towards accountant millennial generation', *International Journal of Management*, 11(5).

Elmarzouki, M. and Jiuhe, W. (2026) 'Hybrid innovation models for productivity growth: the role of Lean, Six Sigma and Industry 4.0 integration', *International Journal of Lean Six Sigma*, 17(1), pp. 71–117.

Fayyaz, A., Liu, C., Xu, Y., Khan, F. and Ahmed, S. (2025) 'Untangling the cumulative impact of big data analytics, green lean six sigma and sustainable supply chain management on the economic performance of manufacturing organisations', *Production Planning & Control*, 36(8), pp. 1137–1154.

Gupta, M. and George, J.F. (2016) 'Toward the development of a big data analytics capability', *Information & Management*, 53(8), pp. 1049–1064.

Gupta, S., Modgil, S. and Gunasekaran, A. (2020) 'Big data in lean six sigma: a review and further research directions', *International Journal of Production Research*, 58(3), pp. 947–969.

Ibrahim, A. and Kumar, G. (2024) 'Selection of Industry 4.0 technologies for Lean Six Sigma integration using fuzzy DEMATEL approach', *International Journal of Lean Six Sigma*, 15(5), pp. 1025–1042.

Khattab, S.A. (2025) 'The impact of big-data analytics on environmental performance: the mediating roles of Lean Six Sigma and green production', *Jordan Journal of Business Administration*, 21(4), pp. 539–562.

Koppel, S. and Chang, S. (2021) 'MDAIC – a Six Sigma implementation strategy in big data environments', *International Journal of Lean Six Sigma*, 12(2), pp. 432–449.

Kumar, P., Bhamu, J., Goel, S. and Singh, D. (2024) 'Interpretive structural modeling of Lean Six Sigma critical success factors in perspective of Industry 4.0 for Indian manufacturing industries', *International Journal of System Assurance Engineering and Management*, 15(8), pp. 3776–3793.

Macias-Aguayo, J., Garcia-Castro, L., Barcia, K.F., McFarlane, D. and Abad-Moran, J. (2022) 'Industry 4.0 and Lean Six Sigma integration: a systematic review of barriers and enablers', *Applied Sciences*, 12(22), 11321.

Maia, D.D.R.P., Lizarelli, F.L. and Gambi, L.D.N. (2024) 'Industry 4.0 and Six Sigma: a systematic review of the literature and research agenda proposal', *Benchmarking: An International Journal*, 31(3), pp. 1009–1037.

Mehmood, K., Kautish, P., Rashid, M., Joshi, Y. and Iftikhar, Y. (2025) 'Digitalization in the circular economy: synergistic impact of big data analytics, green internet of things, and ambidextrous green innovation', *Journal of Cleaner Production*, 509, 145610.

Mendonça, R.S., Medeiros, R.L.P., Silva, L.E.S., Silva, R.G.G., Santos, L.G.S. and de Lucena, V.F. Jr. (2025) 'Enabling technologies of Industry 4.0 for the modernization of an industrial process', *Processes*, 13, 2488.

Mesquita, L.L., Lizarelli, F.L. and Duarte, S. (2023) 'Big data analytics and lean practices: impact on sustainability performance', *Production Planning & Control*. DOI: 10.1080/09537287.2023.2267512.

Pagliosa, M., Tortorella, G. and Ferreira, J.C.E. (2021) 'Industry 4.0 and Lean Manufacturing: a systematic literature review and future research directions', *Journal of Manufacturing Technology Management*, 32(3), pp. 543–569.

Pongboonchai-Empl, T., Antony, J., Garza-Reyes, J.A., Komkowski, T. and Tortorella, G.L. (2024) 'Integration of Industry 4.0 technologies into Lean Six Sigma DMAIC: a systematic review', *Production Planning & Control*, 35(12), pp. 1403–1428.

Pongboonchai-Empl, T., Antony, J., Neustock, J., Stemann, D., Komkowski, T. and Sony, M. (2025) 'DMAIC 4.0 in action: introducing MSA 4.0 and Poka Yoke 4.0', *International Journal of Computer Integrated Manufacturing*. DOI: 10.1080/0951192X.2025.2599558.

Pranckutė, R. (2021) 'Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world', *Publications*, 9(1), 12.

Rejikumar, G., Aswathy Asokan, A. and Sreedharan, V.R. (2020) 'Impact of data-driven decision-making in Lean Six Sigma: an empirical analysis', *Total Quality Management & Business Excellence*, 31(3–4), pp. 279–296.

Schumacher, A., Erol, S. and Sihn, W. (2016) 'A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises', *Procedia CIRP*, 52, pp. 161–166.

Skalli, D., Cherrafi, A., Charkaoui, A., Chiarini, A., Elbaz, J. and Hamani, N. (2024) 'Select a winning Lean Six Sigma 4.0 project: Best Worst Method based decision making approach', *Total Quality Management & Business Excellence*, 35(5–6), pp. 503–528.

Snee, R.D. (2010) 'Lean Six Sigma – getting better all the time', *International Journal of Lean Six Sigma*, 1(1), pp. 9–29.

Sony, M. (2020) 'Design of cyber physical system architecture for Industry 4.0 through Lean Six Sigma: conceptual foundations and research issues', *Production & Manufacturing Research*, 8(1), pp. 158–181.

Snyder, H. (2019) 'Literature review as a research methodology: an overview and guidelines', *Journal of Business Research*, 104, pp. 333–339.

Sordan, J.E., Oprime, P.C., Pimenta, M.L., Silva, S.L. and González, M.O.A. (2022) 'Contact points between Lean Six Sigma and Industry 4.0: a systematic review and conceptual framework', *International Journal of Quality & Reliability Management*, 39(9), pp. 2155–2183.

Sordan, J.E. *et al.* (2025) 'Lean Six Sigma practices supported by Industry 4.0 technologies: evidence from heavy vehicle manufacturers', *Journal of Manufacturing Technology Management*, 36(2), pp. 504–533.

Sousa, E.L., Andretti, F.V. and Castro, M.T.G. (2023) 'Overview of stages of change of Lean Six Sigma programs in organizations from 2005 to 2021', *Gestão & Produção*, 30, e6522.

Srijithesh, P.R. *et al.* (2025) 'Leveraging Lean Six Sigma principles in an Indian tertiary care hospital: a case study', *International Journal of Quality & Reliability Management*, 42(2), pp. 600–630.

Sunder, M.V. and Mahalingam, S. (2018) 'An empirical investigation of implementing Lean Six Sigma in higher education institutions', *International Journal of Quality & Reliability Management*, 35(10), pp. 2157–2180.

Tay, Z.Y. *et al.* (2021) 'Big data analytics and machine learning of harbour craft vessels to achieve fuel efficiency: a review', *Journal of Marine Science and Engineering*, 9(12), 1351.

Thomé, A.M.T., Scavarda, L.F. and Scavarda, A.J. (2016) 'Conducting systematic literature review in operations management', *Production Planning & Control*, 27(5), pp. 408–420.

Tissir, S., Cherrafi, A., Chiarini, A., Elfezazi, S., Shokri, A. and Antony, J. (2024) 'Lean Six Sigma and Industry 4.0 implementation framework for operational excellence: a case study', *Total Quality Management & Business Excellence*, 35(9–10), pp. 1015–1053.

Tranfield, D., Denyer, D. and Smart, P. (2003) 'Towards a methodology for developing evidence-informed management knowledge', *British Journal of Management*, 14(3), pp. 207–222.

Virmani, N. *et al.* (2025) 'Driving operational excellence: the role of technology-organization-environment framework in Lean Six Sigma Integrated Industry 4.0 adoption', *Engineering Management Journal*. DOI: 10.1080/10429247.2025.2465073.

Zulfiqar, M. *et al.* (2023) 'Unlocking the potential: empirical analysis of enablers, barriers, benefits and technologies for integrating Industry 4.0 and Lean Six Sigma in manufacturing organisations', *The TQM Journal*, 36(8), pp. 2360–23.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Performance Assessment and Competencies in Vocational Education: A Bibliometric Analysis

Cleiton Abadia Duarte[1], Lillian do Nascimento Gambi[2], Raquel Mello[3] and Camila Fabricio Poltronieri[1,3]

1 Universidade Federal de Goiás (UFG) – Programa de Pós-graduação em Engenharia de Produção; cleitonduarte@discente.ufg.br
2 Universidade Federal de Viçosa (UFV); lillian.gambi@ufv.br
3 Universidade de São Paulo (USP); raquel.mello@usp.br, camilafp@usp.br

Abstract: Student performance assessment in vocational education has assumed a strategic role, considering the need for competency-based training aligned with current industrial demands. In this context, it becomes essential to understand how scientific production has addressed performance analysis and skills development within the training process. This study aims to analyze the evolution of scientific production on performance assessment and competencies in vocational education through bibliometric analysis. To this end, a search was conducted in the Scopus database, resulting in a set of 241 documents published between 1968 and 2025. The analyses were carried out using Bibliometrix, enabling the examination of growth indicators, scientific collaboration networks, leading sources, authors, institutions, countries, and the thematic structure of the investigated field. The results show a progressive increase in publications, with a predominance of studies in the health field and a strong geographical concentration in countries such as the United States and Australia. A high dispersion among sources and the absence of dominant authors were also observed, indicating the field's multidisciplinary and collaborative nature. The keyword analysis and thematic map revealed the consolidation of conceptual axes related to competency-based assessment and academic performance, as well as the existence of themes that can guide future investigations. Thus, this study contributes to understanding the field's structure and trends, offering a theoretical basis for the development of assessment instruments applicable to technical and vocational education.

Keywords: Vocational Education, Performance Assessment, Competence, Training Process, Bibliometric Analysis, Quality Management in Education.

I. Introduction

The discussion on quality in vocational education has gained increasing prominence in recent decades, mainly due to the structure of technical training, which is understood as a strategic element for productive development. In addition to expanding course offerings, institutions have also faced a more complex challenge: demonstrating, through consistent evidence, that students develop the expected competencies throughout their training. In this context, the assessment process is no longer seen solely as a pedagogical element and instead assumes a structuring role in relation to the organization of teaching and the alignment of training activities within the management of educational performance (Kubrushko et al., 2018; Mohamed et al., 2021; Yamada et al., 2018).

The competency-based approach has changed the way academic performance is understood. Assessment has come to seek evidence of practical application, integration of knowledge, and mastery of technical skills, rather than merely verifying content acquisition. However, this transition has not been smooth, and there remain difficulties related to the clear definition of competencies, the construction of reliable instruments, and the articulation between diagnostic and formative assessment (Yusop et al., 2022; Bramley et al., 2021; Casanova et al., 2021). Other analyses indicate that assessment processes require theoretical frameworks and attention to the contexts in which students are situated, with a view to producing more consistent results (Pierre and Okstad, 2021).

From a strategic management perspective, this methodological fragility has a direct impact on the monitoring of results aimed at promoting improvement. The absence of structured instruments compromises the standardization of assessment processes and the generation of indicators over time. Implementing assessment systems requires alignment among criteria, teacher training, and institutional objectives, with a focus on decision-making (Mohamed et al., 2021; Shah et al., 2023). Studies such as Dickson et al. (2021) reinforce the idea that a clear definition of criteria can become a key condition for transforming assessment into a management tool.

In addition to technical aspects, motivational factors significantly influence student performance. Autonomy and engagement in active learning methodologies are associated with better academic outcomes (Expósito-lópez et al., 2023; Ng and Lo, 2022). Addimando (2019) further argues that organizational climate and support actions directly influence pedagogical practices and engagement in the learning process, which is reflected in assessment outcomes.

The incorporation of technological resources into technical training has further increased this complexity. Hybrid models supported by Virtual Learning Environments (VLEs) require new monitoring and assessment formats, creating a need for greater institutional preparedness and stronger student engagement (Shakeel et al., 2023). Abdigapbarova and Zhiyenbayeva (2023)

highlight an important issue related to digital transformation and the revision of current instruments, while Perkins et al. (2025) point to the reorganization of assessment formats, criteria, and evidence of learning in a broader context.

Another relevant element in this scenario concerns students' preparation for the professional context. Historical factors may influence academic engagement and student commitment (Wang et al., 2022). Achuodho and Piko (2025) observe that social, motivational, and cultural factors also influence the educational trajectory, underscoring the importance of assessment instruments that address the multiple dimensions of knowledge. Challenges such as these are directly linked to the management of productive systems. As in industrial environments, the quality of the educational product depends on the clear definition of processes, indicators, and control mechanisms. In this sense, assessment can be understood as a form of performance measurement capable of providing data for analysis and continuous improvement (Bramley et al., 2021). Merino-Soto et al. (2024) emphasize that well-structured instruments contribute to greater reliability in interpreting results.

In this context, performance assessment can be understood as part of quality management in professional education, as it allows monitoring whether formative processes are achieving the expected outcomes. By transforming learning evidence into indicators, assessment contributes to monitoring training, identifying weaknesses, and defining improvement actions (Aid et al., 2024; Androshchuk et al., 2023). In professional education, this process becomes even more relevant, since training must integrate curriculum, pedagogical practices, teaching competencies, infrastructure, and the demands of the world of work (Ingason and Jónsdóttir, 2017).

This connection between educational assessment and quality management can also be understood through the literature on quality competencies. Martin et al. (2019) highlight that quality management has been moving from a focus solely on control and compliance toward a more strategic environment, related to continuous improvement and innovation. Martin et al. (2021) indicate that quality management competencies involve human and conceptual dimensions related to methods and processes, reinforcing the need for assessments capable of observing different aspects of professional training. In addition, Santos et al. (2021) point out that quality requires new competencies, such as communication, critical thinking, leadership, teamwork, openness to change, use of digital technologies, and data analysis for decision-making.

Given this scenario, a systematic review of the literature is relevant to understand how scientific production has addressed vocational education, particularly in relation to technical education, assessment, and performance. Bibliometrics is an appropriate tool for this purpose, as it enables the analysis of different types of patterns and trends within a research field (Ellegaard and Wallin, 2015). Despite the growth of research related to assessment in vocational education, studies on scientific production focused on performance processes and the development of competencies in technical training remain limited. As a result, there is still difficulty in identifying trends, theoretical gaps, and methodological directions necessary for the development of assessment instruments (Yusop et al., 2022).

Thus, this study aims to analyze, through a bibliometric approach, the scientific production related to assessment and competency development in vocational education, especially in the technical modality, considering its evolution over time, published documents, sources, authors, institutions, countries, and the most frequently used keywords. By understanding how this field has been organized, the study seeks to provide theoretical and methodological support for the development of a multidisciplinary assessment instrument that may assist in monitoring student performance and contribute to the improvement of educational quality.

II. RESEARCH METHOD

This study conducted bibliometric research aimed at analyzing the scientific production related to performance assessment and competencies in technical and vocational education. Bibliometrics involves the application of statistical and mathematical methods to examine structural patterns in the scientific literature, enabling the identification of trends, collaboration networks, temporal evolution, and thematic gaps (Ellegaard and Wallin, 2015). By adopting this approach, the study seeks to understand how the field has been organized over time, what its main lines of discussion are, and how relationships among authors, institutions, and themes are configured.

To operationalize the analyses, the Bibliometrix package, developed in the R environment, was used. This tool is widely recognized for its ability to integrate descriptive analyses, bibliometric indicators, and scientific mapping transparently and efficiently (Aria and Cuccurullo, 2017). The choice of this software is justified by its transparent analytical procedures and its broad use in scientific mapping studies.

Data collection was conducted on January 31, 2026, in the Scopus database, selected for its widespread use in bibliometric studies, multidisciplinary coverage, and continuous updating (ARIA; CUCCURULLO, 2017). This combination enabled the capture of publications that articulate technical training with assessment processes and performance measurement. The search was conducted in the title, abstract, and keywords, using the filters Article / Review / Journal / English. The search string resulted in 241 documents, as shown in Table 1 below:

Table 1 – Search String

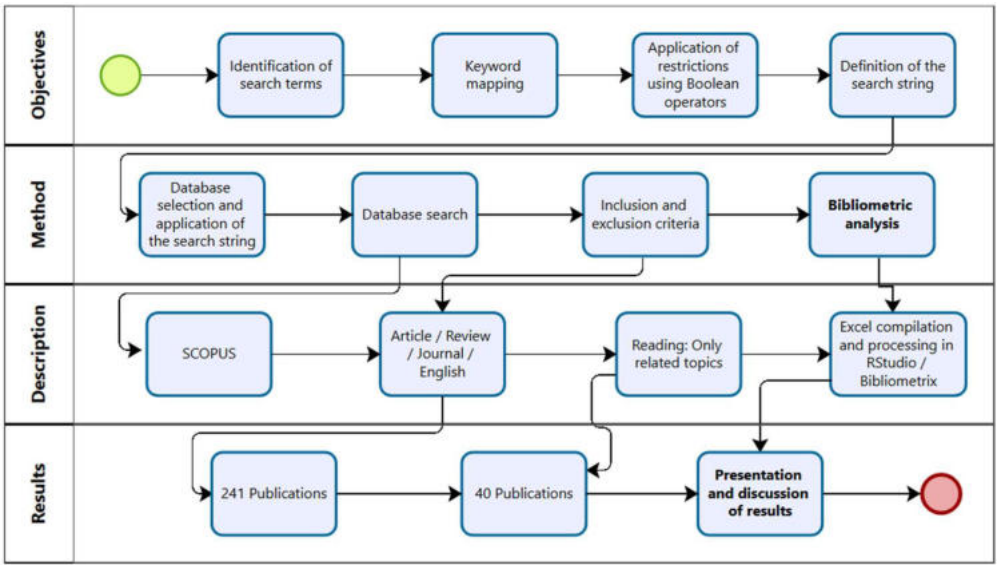
Database	String	Total
Scopus	(("Technical and Vocational Education and Training" OR "Technical Education" OR "Professional Education" OR "Professional Training" OR "Technical Training") AND ("Competency-Based Assessment" OR "Skills Assessment" OR "Academic Achievement" OR "Educational Assessment" OR "Multidisciplinary Assessment"))	241

Source – Authors

In addition to the analyses, a complementary reading of the main thematic axes identified was carried out, aiming to relate the results obtained to discussions on quality management, performance assessment, and competency development in professional education. This stage was intended to support the discussion of the bibliometric findings based on recent studies aligned with the topic, especially those addressing quality management in professional training, quality control in educational institutions, and the competencies required in quality-related contexts.

The results and the inclusion and exclusion criteria were organized in a spreadsheet and analyzed using RStudio (version 2026.01.0). Figure 1 presents the general research flowchart.

Figure 1 – Flowchart



Source – Authors

Forty publications related to the context of the present study were selected. The reading was carried out to assess their coherence and to effectively support the results of the bibliometric analysis.

III. RESULTS

The initial analysis enabled an overall picture of the scientific production related to the investigated topic. The field spans 1968 to 2025, revealing a long historical trajectory, though with a more pronounced intensification in recent decades. A total of 241 publications across 158 different sources were identified, demonstrating thematic dispersion and a diverse set of journals involved in the discussion.

The average age of the publications was 14.3 years, suggesting the presence of consolidated contributions over time and continuous renewal in the field. The annual growth rate of 4.98% reinforces the progressive expansion of research, suggesting the gradual consolidation of the topic. Regarding impact, the average of 13.77 citations per document indicates a consistent level of academic relevance across the publications analyzed.

Regarding authorship, 879 authors were identified across the publications, with a predominance of collaborative work. Only 46 documents were produced by a single authorship, showing that a high level of scientific cooperation characterizes the field. The international co-authorship rate of 14.11% indicates a still limited level of collaboration among researchers from different countries, with room for expansion. Finally, 676 author keywords were recorded, revealing conceptual diversity within the field under investigation. Table 2 presents the general summary of the database.

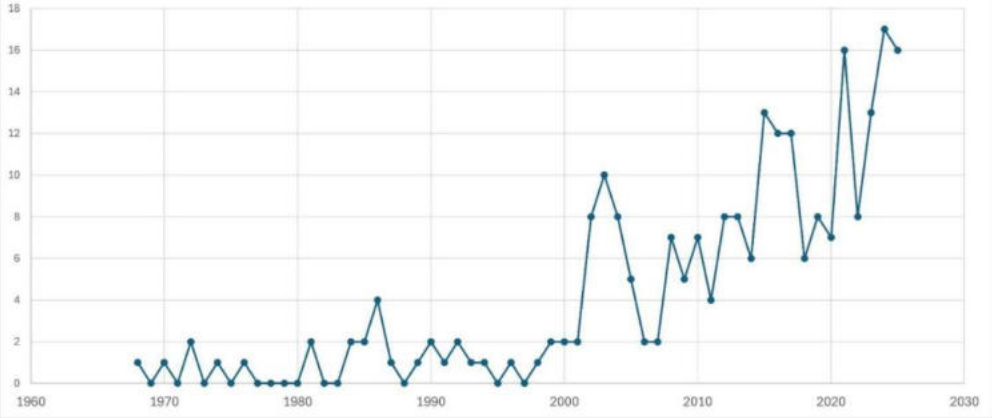
Table 2 – General Summary of the Database

Main Information	Results
Main Information	1968:2025
Sources	185
Documents	241
Average age of documents	14,3
Annual growth rate	4,98%
Average citations per document	13,77
Authors	879
Single-authored documents	46
International co-authorship	14,11%
Authors' keywords	676

Source – Authors

Figure 2 presents the annual evolution of scientific production, with the first publication appearing in 1968. During the first decades, production remained relatively stable, with no clear growth trend. This initial pattern suggests that the topic had not yet occupied a central position in the scientific field. A significant increase in the number of publications is observed, especially from 2001 onward, reaching the highest volumes in the series analyzed. This advance may be related to the growing interest in topics such as performance assessment, competency-based training, and quality in vocational education, in a context marked by technological transformations and new demands placed on training processes. Studies such as those by Yamada et al. (2018) and Mohamed et al. (2021) show that this context has expanded discussions on assessment and competency development. In this sense, Yusop et al. (2022) highlight that competency-based assessment has come to occupy a central place in more recent research in the field. Thus, the evolution of scientific production in this study reinforces the topic's relevance and its connection to the article's purpose.

Figure 2 – Annual Scientific Production



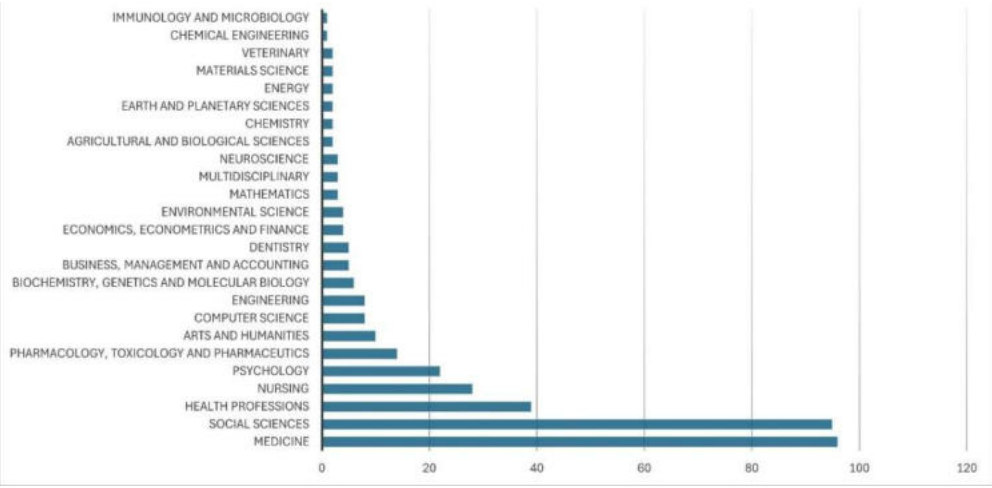
Source – Authors

In addition to the general temporal analysis of scientific production, this study also sought to understand how the documents are distributed across different fields of knowledge in the database. This analysis enables the identification of which areas have engaged more intensively in the discussion of assessment, performance, and competencies in vocational education, as shown in Figure 3.

In general, scientific production is concentrated in the areas of *Medicine* and *Social Sciences*, followed by *Health Professions*, *Nursing, and Psychology*. This configuration indicates that the debate has a strong presence in the health field, while also maintaining consistent dialogue with the *Social Sciences*, especially on topics related to educational policies, training processes, and the management of educational quality.

When analyzing the evolution of the three largest areas over time, it is also evident that a large proportion of the documents classified under Social Sciences are thematically related to the health field, particularly when they address professional training. This aspect is relevant because it shows that the presence of the *Social Sciences* ultimately reinforces the centrality of health within the field under analysis. In other words, even when the debate expands into social, educational, and organizational dimensions, it remains strongly anchored in contexts of health training and practice. These results help reinforce the relevance of professional training across different fields, such as operations management, since professional training influences organizational practices, the quality with which they are developed, and, consequently, organizational performance.

Figure 3 – Documents by Area

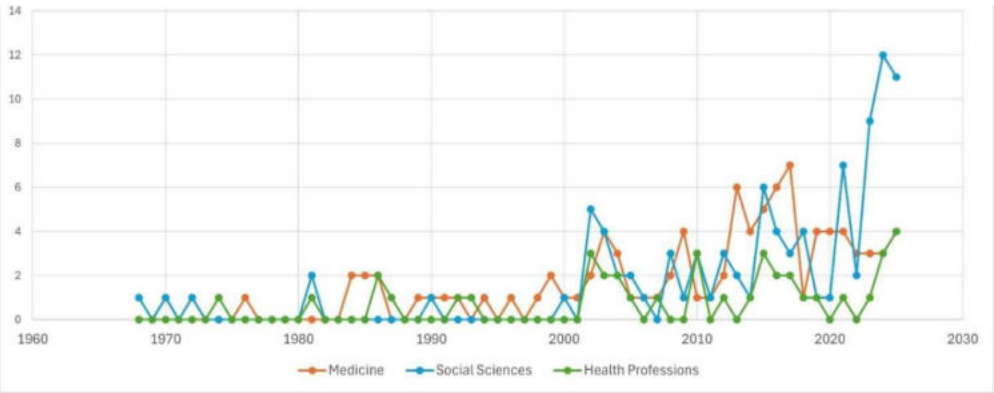


Source – Authors

Figure 4 shows that the growth in scientific production, already observed from 2001 onward, did not occur uniformly across the main areas. In the early years, the records were few and relatively evenly distributed. Over time, *Medicine* and *Social Sciences* came to concentrate a greater number of publications, indicating that the topic gained prominence mainly in these two fields. In recent years, this trend has become even more evident, especially in the *Social Sciences*, which shows a sharper increase and reaches the highest volumes in 2024 and 2025. The field of *Medicine* also maintained an important presence throughout the period, despite some fluctuations. *Health Professions*, in turn, showed a smaller share, though it grew gradually in the final years of the series. These results indicate that, although the debate remains strongly linked to the health field, it has also been expanding and reinforces the multidisciplinary nature of the field analyzed.

Despite this expansion, the results also indicate that the relationship between performance assessment, competency development, and quality management in professional training still appears only implicitly in the analyzed literature.

Figure 4 – Documents per Year



Source – Authors

In this context, an analysis based on scientific sources enables the identification of where studies are being published and which journals exert the greatest intellectual influence in the field under investigation. By examining the most cited sources, their local impact, and their production over time, it becomes possible to visualize the core structure of the field.

In Tables 3, 4 and 5, colors were used to highlight sources that appear in more than one indicator, bringing together the number of publications, impact, and citation count. Among the results, *BMC Medical Education* stands out with 6 publications, 36 citations, and an h-index of 5, demonstrating a relevant presence in the analyzed database. The *Journal of Allied Health* also stands out, with 10 publications and an h-index of 5, indicating expressive participation in the body of studies. *Nurse Education Today*, in turn, presents 5 articles, 36 citations, and an h-index of 5, showing a balance between productivity and impact. Overall, the data reinforce the observation that the discussion on performance assessment and competencies in professional training has been concentrated mainly in journals from the health field and health education, which helps to explain the theoretical basis supporting the investigated topic. This result also reveals an important gap regarding the scope of quality management. Although there are relevant journals addressing health education, competencies, and academic performance, there is a lower presence of sources directly associated with quality management, organizational excellence, or continuous improvement.

Table 3 – Sources

Most Relevant Sources

Sources	Articles
JOURNAL OF ALLIED HEALTH	10
BMC MEDICAL EDUCATION	6
NURSE EDUCATION TODAY	5
PHARMACY EDUCATION	5
THE JOURNAL OF HEALTH ADMINISTRATION EDUCATION	5
PHYSICAL THERAPY	4
AUSTRALIAN OCCUPATIONAL THERAPY JOURNAL	3
FRONTIERS IN PSYCHOLOGY	3
MEDICAL TEACHER	3
THE AMERICAN JOURNAL OF OCCUPATIONAL THERAPY	3

Source – Authors

Table 4 – Sources

Sources Local Impact

Sources	H Index
BMC MEDICAL EDUCATION	5
JOURNAL OF ALLIED HEALTH	5
NURSE EDUCATION TODAY	5
PHARMACY EDUCATION	4
PHYSICAL THERAPY	4
AUSTRALIAN OCCUPATIONAL THERAPY JOURNAL	3
MEDICAL TEACHER	3
THE AMERICAN JOURNAL OF OCCUPATIONAL THERAPY	3
ADVANCES IN HEALTH SCIENCES EDUCATION	2
CANADIAN JOURNAL OF OCCUPATIONAL THERAPY	2

Source – Authors

Table 5 – Sources

Most Local Cited Sources

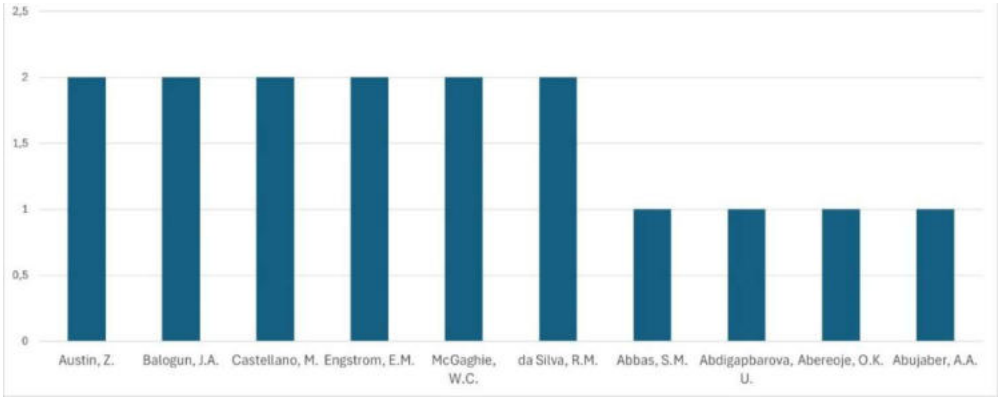
Sources	Articles
ACAD MED	60
MEDICAL EDUCATION	59
MED EDUC	47
ACADEMIC MEDICINE	39
MED TEACH	39
JOURNAL OF ADVANCED NURSING	38
BMC MED EDUC	36
NURSE EDUCATION TODAY	36
MEDICAL TEACHER	32
BMJ	30

Source – Authors

Regarding the authors, the analysis enables understanding of the researchers who have supported the field's development and how their contributions have been distributed over time. Unlike source-based analysis, which reveals the editorial spaces where the debate has been consolidated, authorship analysis highlights the intellectual leadership and research clusters that sustain the discussion's continuity.

According to Figure 5, the absence of a dominant author suggests that the field has been built collaboratively and progressively, with multiple researchers contributing to its conceptual and methodological development. Thus, it is possible to observe that no author has a publication volume substantially higher than the others, since the leading names appear with only two publications each, reflecting the absence of isolated leadership. This finding indicates a distributed production, possibly related to the multidisciplinary nature of the topic, which involves education, health, professional training, and management.

Figure 5 – Documents by Author

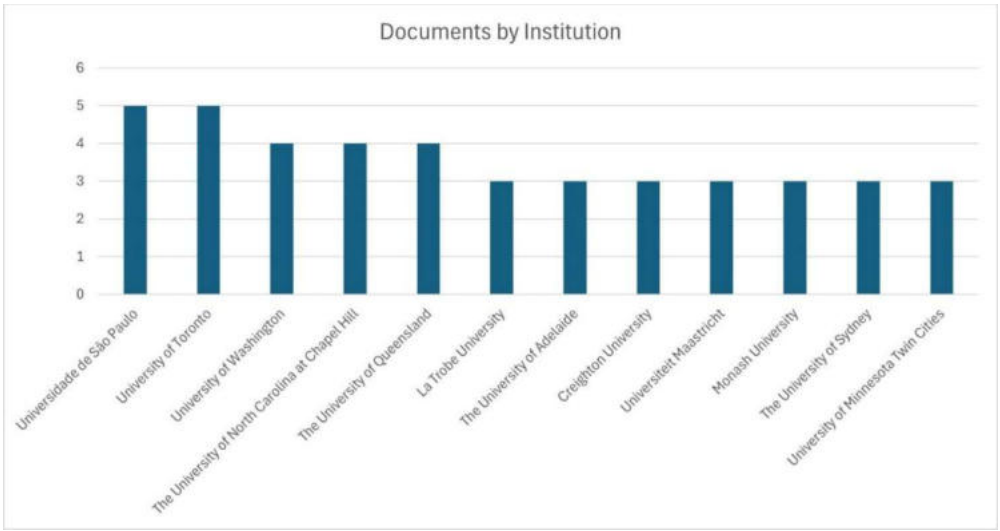


Source – Authors

From an institutional perspective, the study enables identification of where the debate has been taking place, and which academic centers have contributed most consistently to the field's consolidation. Unlike the perspective focused on individual authors, this analysis highlights research centers and areas of thematic concentration.

Figure 6 shows that the institutions with the highest number of publications, such as the *University of São Paulo* and the *University of Toronto*, each account for only five documents. They are followed by institutions such as the *University of Washington, the University of North Carolina at Chapel Hill,* and the *University of Queensland*, with four publications each, as well as other universities with three publications each. This result shows that production is distributed across different institutions, with no strong predominance of a single research center, which reinforces the more dispersed nature of the field under analysis.

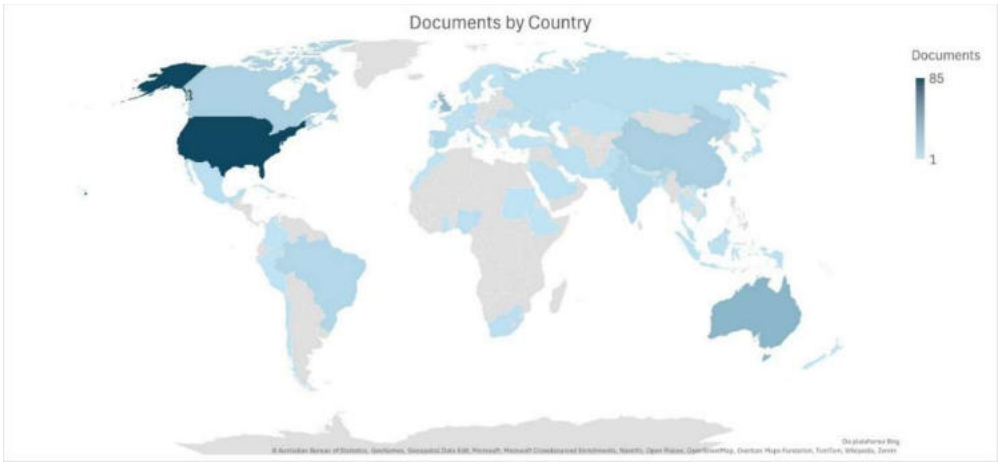
Figure 6 – Main Institutions



Source – Authors

The geographical context of scientific production, as shown in Figure 7, adds an important layer to the understanding of the investigated field. By examining the countries of affiliation of the corresponding authors, the global distribution of production, and its evolution over time, it is possible to identify where the debate has developed more intensively, and which national contexts exert greater intellectual influence.

Figure 7 – Scientific Production by Country



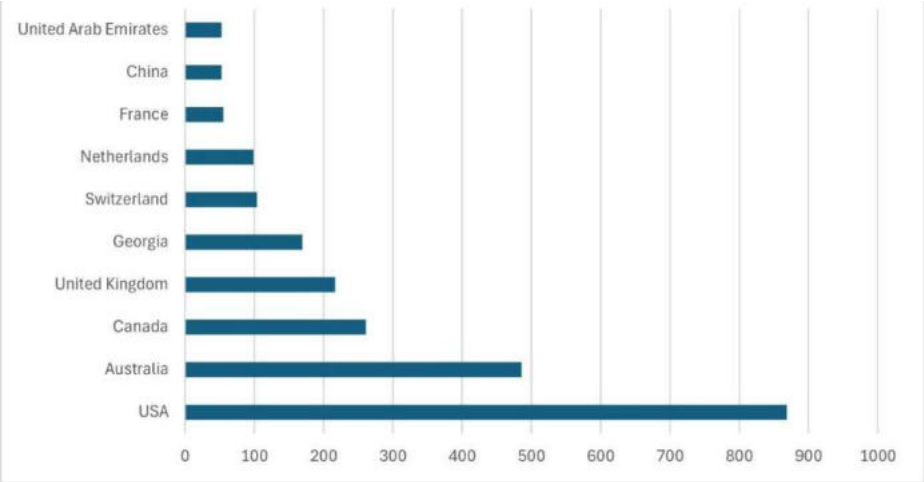
Source – Authors

The data show a predominance of the United States, both in the number of publications and in citation volume, according to Figure 8. This leading position suggests that a large part of the discussions in the field originates from American institutions, which is consistent with the previous results that indicated a strong presence of universities and journals based in that country.

Australia, the United Kingdom, and Canada also stand out, although on a smaller scale, indicating that other countries have also participated in a relevant way in scientific production on the topic. Despite the concentration in the United States, the distribution suggests that the field is not restricted to a single national context but has incorporated contributions from different countries.

Thus, the distribution of scientific production confirms not only the existence of consolidated research hubs but also a gradual process of expansion of the debate on an international scale, which is consistent with the growth and diversification observed in the other analyses of this study. The concentration in the United States and Australia may shape the field's conceptual frameworks and limit the visibility of experiences from other professional education contexts.

Figure 8 – Most Cited Countries



Source – Authors

Keyword analysis provides a complementary perspective to production and citation indicators, as it allows the observation of the most recurrent terms in the investigated field. More than simply identifying terms, this type of analysis reveals conceptual trends and theoretical connections within the scientific debate. In this way, it becomes possible to follow both the accumulated frequency of the main keywords and their relative representativeness across the publications. The word cloud shown in Figure 9 visually synthesizes the terms that structure the scientific output, highlighting the concepts that have become increasingly central to the field's development.

The most recurrent terms in the dataset are centered on *academic achievement, education, students, graduate, curriculum, clinical competence,* and *vocational education*, indicating that the literature analyzed has focused on the relationship between training, academic performance, and competency development. This interpretation is consistent with Yusop et al. (2022), who show that, in the field of vocational education, learning outcomes have often been assessed through approaches that prioritize the development of knowledge, skills, and attitudes relevant to the world of work. In this sense, Mohamed et al. (2021) reinforce that technical training is oriented toward the development of skills required by industry.

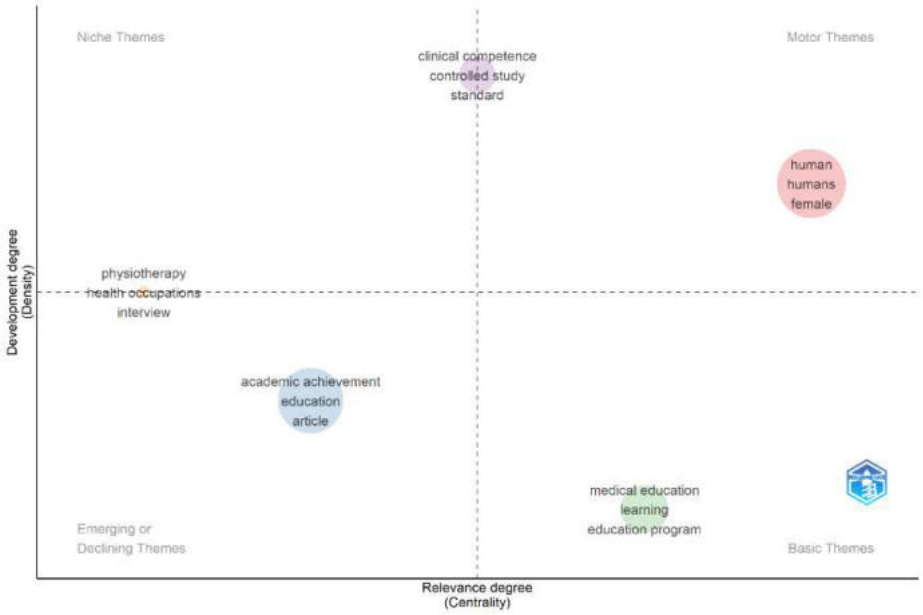
The presence of *academic achievement* is also supported by Achuodho and Piko (2025), who discuss academic performance among *Technical and Vocational Education and Training (TVET)* students based on factors such as motivation, identity, and professional preparation. In addition, the recurrence of terms related to *medical education* and *clinical competence* suggests the strong presence of studies situated in the health field, an aspect that also appears in research on professional training and the assessment of clinical performance, such as the work of Bramley et al. (2021), which focuses on professional competence, clinical practice, and factors associated with student performance. Thus, the word cloud reinforces that the investigated field is primarily organized around academic performance, competency-based training, and professional preparation, with a strong presence of studies in health education.

Figure 9 – Word Cloud

Source – Bibliometrix

With a view to concluding the analysis of the results, Figure 10 presents the thematic map, which facilitates a more interpretative reading of the study. Unlike descriptive analyses based on frequency or citation counts, this representation organizes themes according to two complementary axes: centrality (the degree of relevance to the field as a whole) and density (the level of internal development of the theme). This distribution makes it possible to understand not only which topics are more recurrent, but also how they are interconnected and what role they play in the consolidation and possible transformations of the field, offering a structured overview of its foundations, specializations, and research directions.

Figure 10 – Thematic Map



Source – Bibliometrix

In this scenario, it can be observed that the analyzed field is not organized in a homogeneous way, but rather around thematic clusters with different levels of centrality and development. Among the **basic themes**, *medical education, learning, and education program* stand out, suggesting that the discussion has been sustained by studies focused on training, teaching processes, and learning strategies, especially in the context of health education. In another axis, the cluster *clinical competence, controlled study,* and *standard* appears as a **niche theme**, indicating a more specialized and internally developed core, strongly related to the measurement of professional competence and the use of assessment standards, an aspect also observed in studies applied to clinical training and competency-based structures in the health field. The presence of these terms reinforces the relationship between competency assessment and quality management and indicates that part of the literature has addressed professional training based on standards, criteria, and performance evidence, which is close to the logic of quality systems (Aid et al., 2024; Ingason and Jónsdóttir, 2017).

In turn, the cluster *academic achievement, education* and *article* appears in a more peripheral position, suggesting that although academic performance is recurrent in the literature, it does not always occupy a central place in the articulation of the field. This result is consistent with Achuodho and Piko (2025), who treat academic performance as an important outcome in TVET, but one associated with broader variables such as motivation, identity, and training context. Finally, the *human*, *humans*, and *female* cluster, although it appears as a **motor theme** in the map, should be interpreted with caution. These terms probably reflect an effect of the Scopus database and do not necessarily represent a thematic core related to performance assessment or competencies. Thus, its position in the motor themes quadrant seems to be more associated with the strong presence of studies in the health field than with the existence of a specific conceptual framework. For this reason, the interpretation of the map should prioritize the clusters more directly related to the article's object. Taken together, the map suggests that the field has been structured mainly around discussions of health education, learning, and competency assessment, while themes related to academic performance appear in a more peripheral, less integrated way within the field's core.

Thus, the bibliometric results indicate that performance assessment can be analyzed beyond the pedagogical process. It can also be understood as a management resource, as it allows organizing information on learning, competencies, and professional training. This perspective is close to quality management, as it supports process monitoring, results analysis, and the definition of improvements in the formative context (Aid et al., 2024; Martin et al., 2019; Santos et al., 2021).

IV. Conclusions

This study sought to map, through a bibliometric approach, how scientific production has addressed the relationship between vocational education, assessment, and performance, with special attention to competency-based training and student development in the technical modality. Based on the results, it can be considered that this objective was achieved, since it was possible to map the evolution of the literature over time, identify the main authors, sources, institutions, and countries, and understand how the central themes of the field have been organized.

The results reveal an expanding field, with positive annual growth and a progressive increase in the number of authors, sources, and countries involved. This movement points to greater institutional and geographical diversification of scientific production. However, there is no evidence of consolidated leadership by a specific author, university, or journal, since production appears to be distributed across multiple actors. Thus, the field presents a more dispersed configuration, which reinforces its multidisciplinary character and suggests that the debate remains open to new research perspectives.

The analysis of keywords and the thematic map enabled a clearer understanding of how the literature has been organized internally. In general, scientific production was found to be concentrated mainly on themes related to health education, learning, and competency development, especially in training contexts oriented toward professional practice. Based on these analyses, it is possible to understand that scientific production on performance assessment and competencies in vocational education has been mainly supported by studies in the health field, with greater emphasis on competency-based training and learning, while other approaches related to student performance and assessment management remain less integrated.

From the perspective of quality management, the results reinforce that student performance assessment can be understood as a mechanism for monitoring the quality of professional training. By bringing together indicators related to competencies, learning, and academic performance, assessment provides relevant information to monitor processes, guide decisions, and propose improvements. Although there are studies on student performance, the use of technologies, and the monitoring of results, engineering appears only timidly in the context of competency-based training and educational assessment. This suggests the need to deepen research on the subject and to use the assessment process not only as a pedagogical practice, but also as a tool for quality management, monitoring student progress, and supporting institutional decision-making. In a context marked by digital transformation, the expansion of hybrid environments, and the growing use of educational technologies, understanding how the scientific field has been addressing these issues becomes even more important. Thus, the study contributes to the field of management by highlighting trends, gaps, and possibilities for applying assessment in different educational contexts. As a future direction, it is relevant to expand this debate to other areas of knowledge, deepen the thematic analyses identified, and pursue applied studies that may provide a basis for developing and validating assessment instruments for vocational education.

Overall, the study shows that the analyzed field matured over time and become more complex. The articulation among academic performance, educational processes, and human dimensions remains a central focus of the literature, while connections with related areas are also observed, strengthening the interdisciplinary character of the research. The results provide valuable support for researchers interested in understanding the field's historical trajectory, identifying research gaps, and guiding future studies.

Another important issue identified in the study is the limited articulation between performance assessment, competency development, and quality management in professional training. Although the analyzed literature presents relevant contributions on academic performance, learning, and competencies, these themes still appear only weakly connected to a quality management perspective.

As for limitations, the study relied exclusively on the Scopus database and specific language and document-type filters, which may have influenced the analysis and favored the inclusion of certain areas, especially health. In this sense, future research may broaden the investigation to other databases, compare different thematic scopes, and deepen the qualitative analysis of the most representative articles. It would also be relevant to advance toward systematic reviews and applied studies that contribute to the development and validation of assessment instruments aimed at vocational education, especially in underexplored areas.

References

Abdigapbarova, U. e Zhiyenbayeva, N. (2023), "Organization of student-centered learning within the professional training of a future teacher in a digital environment", *Education and Information Technologies*, Vol. 28, pp. 647–661.

Achuodho, H.O. e Piko, B.F. (2025), "Identity processing styles and academic achievement: focus on academic motivation as a moderator among technical and vocational education and training (TVET) students in Kenya", *Cogent Education*, Vol. 12 No. 1, p. 2584873.

Addimando, L. (2019), "The effect of positive working conditions on work engagement and teaching classroom practices: a large cross-sectional study in Switzerland", *Frontiers in Psychology*, Vol. 10, p. 2129.

Aid, L., Rouaski, K. e Moulfi, Z. (2024), "Evaluation of the quality of professional training and education under the standards of total quality management", *Business Ethics and Leadership*, Vol. 8 No. 1, pp. 45–56.

Androshchuk, I., Sergeieva, L., Stoychik, T., Kravchenko, L. e Martynenko, K. (2023), "Quality control modelling of competitive professionals' training at vocational education institutions", *The New Educational Review*, v. 72, n. 2.

Aria, M. e Cuccurullo, C. (2017), "Bibliometrix: an R-tool for comprehensive science mapping analysis", *Journal of Informetrics*, Vol. 11 No. 4, pp. 959–975.

Bramley, A.L., Thomas, C.J., McKenna, L. e Itsiopoulos, C. (2021), "E-portfolios and Entrustable Professional Activities to support competency-based education in dietetics", *Nursing & Health Sciences*, Vol. 23, pp. 148–156.

Casanova, J.R., Vasconcelos, R., Bernardo, A.B. e Almeida, L.S. (2021), "University dropout in engineering: motives and student trajectories", *Psicothema*, Vol. 33 No. 4, pp. 595–601.

Dickson, T., Deschenes, B., Gleeson, P. e Zafereo, J. (2021), "Impact of faculty and programmatic resources on the proportion of academic doctoral degrees in professional physical therapist education programs", *Physical Therapy*, Vol. 101 No. 4, p. pzab030.

Ellegaard, O. e Wallin, J.A. (2015), "The bibliometric analysis of scholarly production: how great is the impact?", *Scientometrics*, Vol. 105 No. 3, pp. 1809–1831.

Expósito-López, J., Parejo-Jiménez, N., Olmedo-Moreno, E.M. e Chacón-Cuberos, R. (2023), "Type of school and origin as conditioners of motivational development in post-compulsory secondary education", *Cogent Education*, Vol. 10 No. 1, p. 2188800.

Ingason, H.Þ. e Jónsdóttir, E.R. (2017), "The house of competence of the quality manager", *Cogent Business & Management*, Vol. 4 No. 1, p. 1345050.

Kubrushko, P., Shishov, S., Kalnei, V., Scaramanga, V., Shafazhinskaya, N. e Rabadanova, R. (2018), "Perception of educational information in the process of learning of construction and humanitarian universities students: comparative analysis", *International Journal of Civil Engineering and Technology (IJCIET)*, Vol. 9 No. 11, pp. 2331–2337.

Martin, J., Elg, M. e Gremyr, I. (2019), "Fit for purpose? Exploring competence in quality management", *International Journal of Quality and Service Sciences*, Vol. 11 No. 3, pp. 317–333.

Martin, J., Elg, M., Gremyr, I. e Wallo, A. (2021), "Towards a quality management competence framework: exploring needed competencies in quality management", *Total Quality Management & Business Excellence*, Vol. 32 No. 3-4, pp. 359–378.

Merino-Soto, C., Angulo-Ramos, M., Llaja-Rojas, V. e Chans, G.M. (2024), "Academic performance, emotional intelligence, and academic burnout: a cross-sectional study of a mediational effect in nursing students", *Nurse Education Today*, Vol. 139, p. 106221.

Mohamed, H., Mohd Puad, M.H., Rashid, A.M. e Jamaluddin, R. (2021), "Workplace skills and teacher competency from culinary arts students' perspectives", *Pertanika Journal of Social Sciences & Humanities*, Vol. 29 No. 1, pp. 107–125.

Ng, L.-K. e Lo, C.-K. (2022), "Online flipped and gamification classroom: risks and opportunities for the academic achievement of adult sustainable learning during COVID-19 pandemic", *Sustainability*, Vol. 14 No. 19, p. 12396.

Perkins, M., Roe, J. e Furze, L. (2025), "How (not) to use the AI assessment scale", *Journal of Applied Learning & Teaching*, Vol. 8 No. 2, pp. 14–23

Pierre, D.E. e Okstad, J. (2021), "Utilizing leadership assessment tools in graduate education", *New Directions for Student Leadership*, Vol. 2021 No. 170, pp. 87–95.

Santos, G., Sá, J.C., Félix, M.J., Barreto, L., Carvalho, F., Doiro, M., Zgodavová, K. e Stefanović, M. (2021), "New needed quality management skills for quality managers 4.0", *Sustainability*, Vol. 13 No. 11, p. 6149.

Shah, S.K.A., Butt, M., Saleem, A. e Rafique, M.U. (2023), "Exploring the challenges in classroom assessment: a mixed-method study of secondary schools in Pakistan", *International Journal of Evaluation and Research in Education (IJERE)*, Vol. 12 No. 4, pp. 2093–2100.

Shakeel, S.I., Haolader, M.F.A. e Sultana, M.S. (2023), "Exploring dimensions of blended learning readiness: validation of scale and assessing blended learning readiness in the context of TVET Bangladesh", *Heliyon*, Vol. 9, p. e12766.

Wang, X., Leng, J., Cong, X. e Fan, Z. (2022), "Development of junior high school students to secondary vocational students stigma scale", *Psychology Research and Behavior Management*, Vol. 15, pp. 2957–2970.

Yamada, S., Otchia, C.S. e Taniguchi, K. (2018), "Explaining differing perceptions of employees' skill needs: the case of garment workers in Ethiopia", *International Journal of Training and Development*, Vol. 22 No. 1, pp. 51–68.

Yusop, S.R.M., Rasul, M.S., Yasin, R.M., Hashim, H.U. e Jalaludin, N.A. (2022), "An assessment approaches and learning outcomes in technical and vocational education: a systematic review using PRISMA", *Sustainability*, Vol. 14 No. 9, p. 5225.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

A Dual Perspective on Service Quality: Drivers of Choice and Satisfaction in Hotel Event Services

Sofia LOPES [1], Filipa BRANDÃO[1], Marlene AMORIM [1] and Cláudia SILVA [2]

[1] GOVCOPP & DEGEIT– Governance Competitiveness and Public Policies & Department of Economics, Management, Industrial Engineering and Tourism, University of Aveiro; sofiaglopes@ua.pt;filipa.brandao@ua.pt; mamorim@ua.pt
[2]GOVCOPP& ESTGA - – Governance Competitiveness and Public Policies, Águeda School of Technology and Management, University of Aveiro; claudia.margarida@ua.pt

***Abstract*— Service quality in hospitality has traditionally been examined as a multidimensional construct, yet limited attention has been given to how it is evaluated across different stages of the service process. This study explores the drivers of venue selection and satisfaction in the context of hotel-based events, adopting the perspective of event organizers. Drawing on data collected through a questionnaire administered to 131 organizers, the analysis compares the importance attributed to service attributes during the selection phase with the satisfaction levels reported after the event. The results show that venue choice is strongly influenced by tangible and observable attributes, such as location, facilities, and price, while satisfaction reflects a broader combination of physical environment and interaction-based dimensions, including staff performance, responsiveness, and service execution. Although a high level of alignment between importance and performance is observed, the findings reveal subtle differences in how attributes are valued across stages, supporting the existence of a dual logic of service evaluation. A focused Importance–Performance Analysis further highlights specific areas for managerial attention, particularly in relation to problem-solving and communication.The study contributes to the service quality literature by emphasizing the temporal and processual nature of service evaluation and by extending the analysis to a business-to-business context. From a managerial perspective, the results underline the importance of aligning pre-service signals with actual service delivery to enhance customer satisfaction and long-term relationships.**

Keywords – Service quality; Event management; Hospitality; Customer satisfaction; Decision-making; Importance–Performance Analysis

I. INTRODUCTION

Service quality has long been recognized as a critical determinant of competitiveness in hospitality services. A particular case of services delivered by hotels concerns the events hosted within hotel settings. Events including conferences, weddings, and diverse cultural programs, can have a significant weight in activity and operations, and they have become central to hotel strategies for stimulating demand, enhancing differentiation, and generating supplementary revenue streams (Chandra, S., Bhalla, et. Al. 2022). Unlike standardized hospitality services, event typically delivery requires the coordination of several entities and service attribute including tangible elements such as facilities and equipment, alongside intangible dimensions including staff performance, responsiveness, and overall service experience. Despite this complexity, much of the existing literature on service quality in hospitality has predominantly focused on the perceptions of end-users or participants, often adopting a business-to-consumer (B2C) perspective. However, in the context of event management, a significant proportion of service evaluation is mediated by event organizers, who act as key decision-makers and intermediaries between the service provider and final participants. As highlighted in prior research, limited attention has been given to understanding how these actors evaluate service quality, particularly in terms of how they select venues and assess service performance.

This gap becomes particularly relevant when considering that service evaluation does not occur at a single point in time, but rather unfolds across distinct phases. On the one hand, the selection of a venue is typically based on ex-ante criteria, where decision-makers rely on observable and comparable attributes such as location, capacity, and price. On the other hand, satisfaction emerges ex-post, shaped by the actual service experience, including staff behavior, problem-solving capability, and service responsiveness. Evidence suggests that these dimensions are not equally weighted across phases, raising questions about the alignment between what drives choice and what ultimately determines perceived service quality. Building on this premise, this paper proposes that service quality in event management should be understood through a dual logic of evaluation, distinguishing between drivers of choice and drivers of satisfaction. Using empirical evidence from a case study conducted at a hotel event setting, based on responses from 131 event organizers, this study aims to identify and compare the factors influencing venue selection and those shaping satisfaction with service delivery.

The findings reveal a structural distinction between the two dimensions: while choice is primarily driven by tangible and observable attributes, satisfaction is largely influenced by intangible and relational aspects of service delivery. This misalignment suggests that service providers may be optimizing for factors that facilitate selection, while underestimating those that determine long-term satisfaction and loyalty. By articulating this dual perspective, the

paper contributes to the service quality literature in three ways. First, it extends the analysis to a business-to-business (B2B) context, emphasizing the role of event organizers as key evaluators of service quality. Second, it introduces a conceptual distinction between ex-ante and ex-post evaluation processes, offering a more nuanced understanding of service quality assessment. Third, it provides managerial insights for hospitality providers, highlighting the need to better align service design, communication, and delivery processes with the factors that truly drive customer satisfaction.

II. CONCEPTUAL BACKGROUND

A. Service Quality as a Multidimensional Construct

Service quality is widely recognized as a multidimensional construct, encompassing a combination of tangible and intangible elements that jointly shape customer perceptions. Among the most influential frameworks, SERVQUAL (Parasuraman et al., 1988) proposes that service quality emerges from the gap between customer expectations and perceptions of actual service performance. The model identifies five core dimensions: tangibles, reliability, responsiveness, assurance, and empathy. Its simplicity and broad applicability have made SERVQUAL one of the most widely adopted instruments for service quality assessment across industries, including hospitality and tourism.

Despite its popularity, SERVQUAL has also attracted criticism. Researchers have questioned the conceptual validity of measuring expectation–perception gaps and argued that customer evaluations are often better explained by performance perceptions alone. In response, Cronin and Taylor (1992) developed SERVPERF, a performance-based approach that assesses service quality exclusively through perceived service performance. Several empirical studies have reported that SERVPERF demonstrates stronger explanatory power and greater predictive validity regarding customer satisfaction and behavioral intentions.

In hospitality contexts, this multidimensionality is particularly pronounced, as service delivery involves both physical resources—such as facilities, equipment, and spatial layout—and interaction-based components, including staff behavior, responsiveness, and communication (Prakash, 2019). Traditional approaches to service quality emphasize dimensions such as reliability, responsiveness, assurance, empathy, and tangibles, highlighting the coexistence of objective and subjective elements in the evaluation process. While tangible attributes are typically easier to observe and compare across service providers, intangible aspects are more closely tied to the service experience and are often co-created through interactions between customers and service personnel (Ramya et al. 2019, Sipe et al., 2018). Within hospitality and tourism contexts, both SERVQUAL and SERVPERF have been extensively adapted to accommodate sector-specific characteristics. Studies have incorporated additional dimensions associated with physical environments, customer experiences, service encounters, and event-specific factors. More recently, customer experience research has expanded the scope of service quality assessment by emphasizing emotional, relational, and experiential aspects of service encounters, recognizing that customer evaluations extend beyond functional service delivery.

In the context of hotel-based events, service quality becomes even more complex due to the integration of multiple service components. Event delivery requires the coordination of venue characteristics, technical infrastructure, catering services, and real-time interactions between staff and clients. As a result, service quality is not determined by a single attribute but emerges from the interaction of several dimensions that collectively shape the overall experience (Boo & Busser, 2018, Alnawas, & Hemsley-Brown, 2019). The literature reflects this complexity by identifying a broad set of dimensions that influence service quality in hospitality and event settings (Ali et al. 2023, Kusuma et al., 2021). These include both tangible attributes—such as location, facilities, and price—and intangible aspects—such as staff performance, responsiveness, empathy, and communication (Kulu and Bola, 2023). While these dimensions are often studied together, they represent distinct aspects of service delivery and may play different roles in the evaluation process.

The literature on service quality in hospitality and event management identifies a broad range of attributes influencing customer evaluations. However, these attributes are often operationalized differently across studies and service contexts, resulting in overlapping terminology and fragmented conceptual structures. To improve clarity and support subsequent analysis, this study synthesized dimensions reported across prior service quality, hospitality, and event literature and consolidated related constructs into broader analytical categories. The resulting dimensions preserve conceptual alignment with established frameworks such as SERVQUAL and SERVPERF while incorporating additional elements relevant to hotel event services. In particular, dimensions associated with tangible aspects (e.g., facilities, accessibility, and venue characteristics) reflect factors traditionally linked to service expectations and provider selection, whereas interaction-based dimensions (e.g., responsiveness, communication, empathy, and service execution) reflect experiential elements that become more salient during service delivery. Table 1 presents the resulting framework used to organize service quality dimensions in the context of hotel event services.

TABLE I. QUALITY DIMENSIONS

Dimensions	Description
Facilities & Physical Environment	Quality, condition, and adequacy of event spaces, equipment, and infrastructure
Location & Accessibility	Ease of access, proximity to key areas, transport, and parking availability
Capacity & Event Fit	Suitability of space in terms of size, layout, and adaptability to event type
Price & Value for Money	Cost level and perceived fairness relative to service delivered
Staff Performance	Professionalism, friendliness, competence, and overall staff interaction quality
Responsiveness	Ability to respond quickly to requests, solve problems, and manage unexpected situations
Empathy & Personalization	Individualized attention and capacity to adapt services to client needs

Communication & Coordination	Clarity of information, coordination between actors, and effectiveness of communication
Food & Beverage Quality	Quality, variety, and adequacy of catering services provided during events
Reliability & Service Execution	Ability to deliver services as promised and ensure smooth event execution
Flexibility & Adaptability	Capacity to adjust to changes and unforeseen circumstances during events

The dimensions presented highlight the dual nature of service quality in event contexts. On the one hand, tangible attributes—such as facilities, location, and price—provide objective and comparable criteria that support evaluation. On the other hand, intangible dimensions—such as staff performance, responsiveness, and empathy—are closely tied to the service experience and are often only fully assessed during or after service delivery. This distinction suggests that service quality should not be treated as a uniform construct, as different dimensions may assume varying levels of importance depending on when and how the service is evaluated. In this study service evaluation is addressed as a process unfolding across different stages. Although the dimensions presented in Table 1 were adapted to the specific context of hotel event services, they maintain conceptual links with established service quality frameworks, particularly SERVQUAL (Parasuraman et al., 1988) and subsequent performance-based approaches such as SERVPERF (Cronin & Taylor, 1992). Several dimensions identified in this study correspond directly to traditional service quality constructs. For example, Facilities & Physical Environment reflects the tangible dimension; Reliability & Service Execution aligns with reliability; Responsiveness captures the ability to react promptly to requests and unexpected situations; and Staff Performance and Communication & Coordination relate to assurance and service interaction dimensions. Similarly, Empathy & Personalization reflects individualized attention and adaptation to customer needs. However, the event context introduces additional considerations that are less explicitly represented in traditional models, including event fit, operational coordination, flexibility, and venue-related decision criteria such as accessibility and pricing. Consequently, rather than replacing existing frameworks, the dimensions proposed in this study extend prior service quality approaches by incorporating attributes that reflect both the operational complexity of hotel event services and the different stages through which service evaluation occurs.

B. Events in Hotel Contexts: Operational Complexity and Quality Challenges

Events in hotels (meetings, weddings, conferences, banquets) are for many providers a core strategic tool for revenue, branding, and guest experience in hospitality. Research has been highlighting how hotels organize, manage, and leverage events to gain competitive advantage and customer satisfaction (Cheruon, and Korir, 2023). Hotels have increasingly expanded their service offerings to include event-related activities such as conferences, weddings, and corporate gatherings. These events have become important drivers of demand, enabling hotels to diversify revenue streams, increase occupancy, and differentiate their value proposition in competitive markets. Beyond direct revenue, events also generate indirect benefits, including increased visibility, cross-selling opportunities, and enhanced customer engagement (Nain, 2018).

However, the integration of events into hotel operations introduces significant managerial and operational complexity. Unlike standard hospitality services, events are characterized by high levels of customization, temporal constraints, and the need for coordination across multiple functional areas, including accommodation, food and beverage, technical support, and front-office services. As a result, event delivery often requires the simultaneous management of multiple service processes within a shared physical and organizational space (Hutson, 2022). One of the main challenges arises from the coexistence of event-related activities with regular hotel operations. The presence of large groups of event participants may create pressure on shared resources such as common areas, parking, restaurants, and staff availability, potentially affecting the experience of other guests. This overlap increases the risk of service disruptions, delays, and inconsistencies in service quality. Furthermore, event services are inherently dynamic and subject to variability. Last-minute changes, unexpected requests, and coordination issues are common, requiring a high degree of responsiveness and adaptability from hotel staff. In this context, service quality becomes highly dependent on real-time decision-making, effective communication, and the ability to manage interactions between multiple actors, including organizers, participants, and internal teams. These characteristics highlight that managing service quality in hotel-based events extends beyond the provision of physical infrastructure and involves addressing a set of operational challenges that can directly impact the consistency and reliability of service delivery. Table 2 summarizes the main challenges associated with event management in hotel contexts and their implications for service quality. While Table 1 identifies the principal dimensions used to evaluate service quality, hotel event settings introduce operational conditions that influence how these dimensions are delivered and perceived. Understanding these contextual challenges is particularly relevant because operational complexity may affect the relative importance of different service quality dimensions across the service process.

TABLE II. CHALLENGES ASSOCIATED WITH EVENT MANAGEMENT IN HOTELS

Challenge	Description	Implications for Service Quality
Resource Conflicts	Competition for shared resources (space, staff, equipment) between events and regular guests	Increased risk of delays, reduced service efficiency, and lower perceived quality
Simultaneity of Services	Concurrent delivery of multiple services (events, accommodation, dining)	Greater difficulty in maintaining consistency and service standards
High Customization	Each event requires specific configurations and tailored services	Increased complexity and risk of service errors
Demand Peaks & Variability	Sudden increases in demand during events and fluctuating workloads	Pressure on staff and potential service breakdowns
Coordination Across Departments	Need for synchronization between multiple units (e.g. reception, technical teams)	Communication failures may lead to service inconsistencies
Uncertainty & Last-Minute Changes	Frequent changes in schedules, requirements, or participant numbers	Requires high responsiveness; failures negatively impact satisfaction

Customer Heterogeneity	Different expectations between event participants and regular guests	Difficulty in balancing service quality across customer segments
Time Constraints	Strict schedules and limited setup/delivery time for events	Reduced margin for error and increased operational pressure
Dependence on Human Performance	Strong reliance on staff behavior and interaction quality	Service quality becomes highly variable and difficult to standardize

C. Service Evaluation Across Phases: A Dual Logic Perspective

Building on the multidimensional nature of service quality and the operational complexity associated with event management in hotel contexts, it becomes evident that service evaluation is not a static or uniform process. Instead, it unfolds across different stages, each characterized by distinct evaluation criteria and decision logics.

In the context of hotel-based events, service evaluation can be understood as occurring across two main phases: a pre-consumption phase, associated with the selection of the service provider, and a post-consumption phase, related to the assessment of the service experience. These phases differ not only in timing but also in the type of information available and the criteria used for evaluation.

During the pre-consumption phase, event organizers act as decision-makers who must select a venue based on available information. At this stage, the evaluation process is primarily driven by the need to reduce uncertainty and compare alternatives. Consequently, greater emphasis is placed on attributes that are observable, measurable, and easily comparable across options, such as location, facilities, capacity, and price. These dimensions, largely tangible in nature, serve as proxies for expected service quality and support rational decision-making (Ambrecht, 2021).

In contrast, the post-consumption phase is grounded in the actual service experience. Once the event takes place, the evaluation of service quality becomes increasingly dependent on interaction-based and experiential factors. Attributes such as staff professionalism, responsiveness to unexpected situations, communication effectiveness, and the ability to deliver a seamless experience play a central role in shaping satisfaction. These dimensions are inherently intangible and often only fully assessable during or after service delivery.

This distinction suggests that service quality in hotel event contexts is governed by a dual logic of evaluation. On the one hand, a selection logic dominates the pre-consumption phase, where tangible attributes guide decision-making. On the other hand, an experience logic characterizes the post-consumption phase, where intangible and relational factors determine perceived service quality and satisfaction.

Importantly, these two logics are not necessarily aligned. The dimensions that influence the initial choice of a service provider may differ significantly from those that ultimately determine satisfaction with the service experience. As evidenced in prior research, while tangible factors such as location and facilities are critical in the selection process, intangible dimensions—such as staff performance, empathy, and responsiveness—are more strongly associated with satisfaction outcomes . This misalignment suggests that service providers may prioritize attributes that facilitate selection while underestimating those that drive long-term satisfaction and loyalty (Berners and Martin, 2022).

In the specific case of hotel events, this duality is further reinforced by the role of event organizers, who operate across both phases. These actors are not only responsible for selecting the venue but also actively interact with the service delivery process, influencing and evaluating quality in real time. As such, they occupy a hybrid position between customer and co-producer, providing a particularly relevant perspective for understanding how service quality is assessed across stages.

Understanding this dual logic is critical for both theory and practice. From a theoretical perspective, it highlights the need to move beyond static models of service quality and consider the temporal and processual nature of evaluation. From a managerial standpoint, it underscores the importance of aligning pre-service signals with actual service delivery, ensuring that the attributes used to attract customers are consistent with those that ultimately shape satisfaction.

III. CONTEXT AND DATA FOR THE STUDY

A. Research Context and Data Collection Overview

The empirical analysis is based on a case study conducted in a hotel with a substantial volume of event services, located in Portugal. This unit operates in a segment where accommodation services are closely integrated with event-related activities, including corporate meetings, conferences, and social events. The hotel provides a relevant context for the study of service quality in event settings, as it combines traditional hospitality operations with a strong focus on event management. The coexistence of these activities creates a complex service environment, characterized by the need to coordinate multiple resources and stakeholders while maintaining consistent service quality. In particular, the hotel regularly hosts events of different types and scales, requiring the integration of facilities, catering, and service interactions within a shared operational structure. This makes it an appropriate setting to examine how service quality is perceived and evaluated in contexts where both tangible and intangible dimensions are critical.

Data were collected through a structured questionnaire administered to event organizers who had previously held events at the hotel. The focus on organizers reflects their central role as decision-makers and evaluators of service quality in event contexts. The questionnaire was distributed via email to a database of contacts provided by the hotel, covering a period from June to September 2025. A total of approximately 800 invitations were sent, resulting in 173 responses, of which 131 were considered valid for analysis after excluding incomplete questionnaires. The sample

includes different types of organizers, such as private individuals, company representatives, event agencies, and tour operators, allowing for a diverse perspective on service evaluation. Additionally, the dataset covers both corporate and social events, enabling comparisons across event types.

The selection of event organizers as respondents was intentional and aligned with the objectives of the study. Unlike final event participants, organizers simultaneously assume the role of service selector and service evaluator, making them particularly suitable for examining how service quality is assessed across different stages. Their decisions involve evaluating observable venue characteristics before service delivery and assessing operational and interaction-related dimensions during and after event execution. The sampling process followed a non-probabilistic convenience approach based on access to previous event clients provided by the participating hotel. Approximately 800 invitations were distributed electronically, resulting in 173 responses and a response rate of approximately 21.6%. After excluding incomplete questionnaires, 131 valid responses were retained for analysis. Although the sample does not allow statistical generalization, it provides relevant exploratory insights into how professional and non-professional organizers evaluate hotel event services. Given the diversity of respondents included in the study, the analysis sought to capture variation across event contexts rather than represent a homogeneous customer profile. This heterogeneity was considered valuable because different organizer profiles may emphasize different service quality dimensions during both venue selection and service evaluation.

B. Questionnaire Design and Measurement Structure

The questionnaire was designed to capture three main dimensions of service evaluation: (i) factors influencing the choice of the event venue, (ii) factors related to satisfaction with the service experience, and (iii) overall behavioral intentions, including the likelihood of returning and recommending the venue.

To operationalize these dimensions, respondents were asked to evaluate a set of service attributes using Likert-type scales. Specifically, the questionnaire included items measuring both the importance of different attributes in the selection process and the perceived performance of those same attributes in the context of the event experience. This dual measurement approach enables a direct comparison between expectations and perceived service delivery. The set of attributes included both tangible dimensions—such as facilities, location, capacity, and price—and intangible dimensions—such as staff performance, responsiveness, communication, and empathy—reflecting the multidimensional nature of service quality identified in the literature. Table 3 summarizes the main sections of the questionnaire, the type of variables collected, and example items used in each section.

TABLE III. QUESTIONNAIRE

Section	Objective	Type of Variables	Example Items
Section A: Event Characterization	Describe the event and organizer profile	Categorical	Type of event, number of participants, type of organizer
Section B: Factors Influencing Choice	Identify key criteria for venue selection	Importance ratings	Location, facilities, capacity, price, accessibility
Section C: Satisfaction with Service	Assess perceived performance of service attributes	Satisfaction ratings	Staff professionalism, responsiveness, communication, food quality
Section D: Behavioral Intentions	Capture post-event evaluation and loyalty	Intent measures	Likelihood to return, likelihood to recommend
Section E: Open Feedback	Collect additional qualitative insights	Open-ended	Suggestions, comments, additional feedback

Questionnaire items were developed based on dimensions identified in the service quality, hospitality, and event management literature and subsequently adapted to the context of hotel event services. Particular attention was given to ensuring that the same underlying dimensions could be evaluated across both stages of service evaluation, allowing comparison between factors influencing venue selection and satisfaction with service delivery. To improve content validity, questionnaire structure and wording were reviewed prior to distribution to ensure clarity and contextual relevance. Most service attributes were measured using Likert-type scales capturing either the importance attributed to selection criteria or satisfaction with perceived service performance. This dual measurement approach supports direct comparison between ex-ante and ex-post evaluations and aligns with prior approaches used in service quality assessment.

C. Data Analysis Approach

Data analysis was conducted in three main stages, combining descriptive and comparative techniques to examine the drivers of service evaluation across different phases.

First, descriptive statistics were used to identify the most relevant attributes influencing both the selection of the event venue and satisfaction with the service experience. Mean values and ranking of attributes were analyzed separately for importance (choice phase) and performance (satisfaction phase), allowing for an initial comparison between the two dimensions. Second, a comparative analysis was performed to examine differences between the criteria associated with venue selection and those driving satisfaction. This step aimed to identify potential divergences in the relative importance of service quality dimensions across the pre-consumption and post-consumption phases. Third, an Importance–Performance Analysis (IPA) was conducted to assess the alignment between the perceived

importance of service attributes and their evaluated performance. This technique enables the identification of critical areas for managerial attention by classifying attributes into four categories: (i) high importance/high performance, (ii) high importance/low performance, (iii) low importance/high performance, and (iv) low importance/low performance. The combination of these analytical approaches allows for a structured examination of service evaluation, supporting the identification of patterns associated with the dual logic of service quality—namely, the distinction between attributes that drive choice and those that shape satisfaction.

To improve interpretability, service attributes were grouped according to the dimensions identified in the conceptual framework. Comparative analysis focused not only on absolute scores but also on identifying differences in relative importance across evaluation stages. The Importance–Performance Analysis was selected because of its suitability for identifying managerial priorities by simultaneously considering perceived importance and perceived service performance.

IV. OVERVIEW AND DISCUSSION OF RESULTS

A. Understanding the Drivers of Venue Choice

Results are presented following the dual evaluation logic proposed in the conceptual framework. First, attributes associated with venue selection are examined to understand pre-service decision criteria. Second, satisfaction-related attributes are analyzed to identify factors shaping post-service evaluation. Finally, a comparative assessment is conducted to explore areas of convergence and divergence between the two stages. The analysis of factors influencing venue selection reveals a clear emphasis on tangible and operational attributes. The highest-rated criteria include location, facilities, capacity, and price, all of which provide concrete and comparable information for decision-making.

Location and accessibility emerge as particularly relevant, reflecting the importance of logistical convenience for both organizers and participants. Similarly, the availability of adequate facilities and appropriate event spaces plays a central role in the selection process, as these attributes directly determine the feasibility of hosting the event. Price and perceived value for money also rank among the most important factors, indicating the relevance of economic considerations in the decision-making process. These findings suggest that organizers prioritize attributes that can be evaluated in advance and compared across alternatives, reinforcing the role of tangible dimensions in the pre-consumption phase.

TABLE IV. INFLUENCE OF SERVICE ATTRIBUTES IN VENUE CHOICE

Service Quality Dimension	Attribute	Level of Influence
Location & Accessibility	Location	+++
	Accessibility (reduced mobility)	++
	Parking	++
Facilities & Physical Environment	Information on spaces and services	+++
	Hotel condition (maintenance)	+++
	Decoration and aesthetics	++
	Hotel size/dimension	++
	Surrounding environment (landscape)	++
Price & Value for Money	Room price	+++
	Equipment price	++
	Food & beverage pricing	++
Reputation & Reviews (e-WOM)	Online reviews	++
	Recommendations (friends, colleagues)	++
	Hotel classification (stars)	++
Event-Specific Fit	Capacity / adequacy of spaces	+++

B. Drivers of Satisfacton with Service

In contrast to the selection phase, satisfaction with the service experience is primarily driven by intangible and interaction-based dimensions. The results highlight the importance of staff-related attributes, particularly professionalism, friendliness, and the ability to respond effectively to client needs.

Responsiveness emerges as a critical factor, especially in the context of event services, where unexpected changes and real-time adjustments are common. Effective communication and coordination also play a key role in shaping the overall experience, contributing to a perception of reliability and service quality.

Additionally, elements such as food and beverage quality contribute to satisfaction, although their impact appears secondary compared to interaction-based factors. Overall, the findings indicate that satisfaction is closely linked to the quality of service delivery during the event, rather than to the physical attributes that influenced the initial choice.

TABLE V. INFLUENCE OF SERVICE ATTRIBUTES IN EVENT SATISFACTION

Service Quality Dimension	Attribute	Level of Influence
Service Quality (Staff Performance)	Staff professionalism	+++
	Friendliness and attitude	+++

	Availability of staff	++
Responsiveness	Ability to respond to requests	+++
	Problem-solving capability	+++
	Flexibility during the event	++
Communication & Coordination	Clarity of communication	+++
	Coordination between teams	++
	Information provided during the event	++
Empathy & Personalization	Attention to client needs	++
	Personalization of service	++
Food & Beverage Quality	Quality of food and beverages	++
	Adequacy of catering services	++
Reliability & Service Execution	Event runs as planned	+++
	Compliance with expectations	+++
Facilities & Physical Environment	Condition of spaces during the event	++
	Comfort of facilities	++
Price & Value for Money	Perceived value for money	++

C. Misalignment Between Choice and Satisfaction

The comparison between importance and satisfaction scores reveals that service evaluation follows different logics across stages. Attributes associated with venue selection are predominantly linked to tangible and observable elements, whereas post-event satisfaction is more strongly influenced by interaction quality and service execution. This pattern supports the argument that service quality should not be interpreted as a static construct but rather as a process in which evaluation criteria evolve throughout service delivery.

This distinction also aligns with previous service quality literature. While selection-related dimensions resemble traditional expectation-based evaluations associated with observable service cues, satisfaction outcomes appear more closely linked to performance and experience-oriented approaches. The findings therefore suggest that hotel event providers should manage not only the attractiveness of pre-service signals but also the consistency of operational and relational performance during delivery. This divergence indicates that the criteria used by organizers to select a venue do not fully reflect the factors that determine the quality of the service experience. In other words, decision-makers rely on observable and comparable attributes during the selection phase but evaluate the service based on experiential factors that only become evident during delivery.

The results therefore provide empirical support for the existence of a dual logic of service evaluation, as proposed in the conceptual framework. The distinction between selection and experience logics highlights a potential gap in how service providers design and communicate their value proposition.

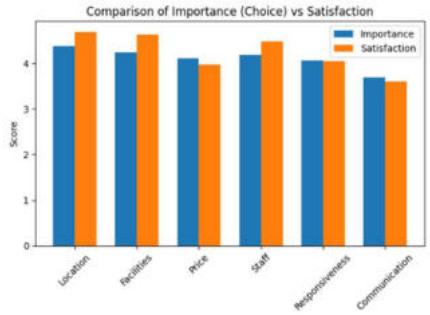


Fig. 1. Comparison of Imprtance od Service Atributes, and Satisfaction with Service Attributes

The results show a general alignment across most attributes, with location and facilities maintaining high scores in both stages. However, some differences are observed, particularly in attributes such as price and communication, where satisfaction does not fully match their perceived importance. Conversely, staff-related dimensions tend to perform slightly above their importance levels, indicating the strong role of service interaction in shaping the overall experience. From a managerial perspective, this misalignment suggests that hotels may be optimizing attributes that facilitate selection while underestimating those that drive satisfaction and long-term customer relationships.

TABLE VI. GAP ANALYSIS FOR IMPORTANCE AND SATISFACTION SERVICE ATTRIBUTES

Attribute	Importance (M)	Satisfaction (M)	Gap
Location	4.38	4.69	+
Facilities (condition)	4.24	4.63	+
Staff performance	4.19	4.48	+
Problem-solving capability	4.06	4.05	~
Information / communication	3.69	3.60	−
Landscape / environment	4.08	4.68	+
Decoration & aesthetics	4.17	4.63	+
Parking	4.02	4.60	+

A focused Importance–Performance Analysis was conducted on selected attributes displaying greater variation between importance and satisfaction scores. The results indicate a generally high level of alignment between expectations and performance, with most attributes falling into a 'keep up the good work' profile. In particular, location, facilities, and staff-related dimensions exhibit strong performance relative to their importance. However, some differences emerge. The ability to respond to problems and unexpected situations presents a more balanced profile, suggesting that performance only meets, rather than exceeds, expectations. Additionally, aspects related to information and communication show slightly lower satisfaction compared to their importance, indicating a potential area for improvement. Conversely, several attributes related to the physical environment—such as landscape, decoration, and parking—appear to perform above their relative importance, suggesting possible overinvestment in these areas. Overall, the results highlight a well-performing service system with specific opportunities for improvement in interaction and coordination processes. Although the comparative analysis provides useful insights regarding differences between selection and satisfaction criteria, results should be interpreted considering the exploratory nature of the study and the context-specific characteristics of the sample.

V. CONCLUSION AND FINAL INSIGHTS

This study set out to examine how service quality is evaluated in the context of hotel-based events, focusing on the distinction between the drivers of venue selection and those shaping satisfaction with the service experience. Drawing on data collected from event organizers, the analysis explored the relative importance of different service attributes across these two stages.

The results confirm that service evaluation is not uniform, but unfolds across phases characterized by different configurations of importance. While the selection of a venue is strongly influenced by tangible and observable attributes—such as location, facilities, and price—satisfaction reflects a broader combination of physical environment and interaction-based dimensions, including staff performance, responsiveness, and service execution. These findings provide empirical support for the existence of a dual logic of service evaluation, where expectations formed prior to service delivery are complemented and, in some cases, reweighted through the actual service experience.

Importantly, the analysis also reveals a generally high level of alignment between importance and performance, indicating that the service system under study performs well across key dimensions. However, the results highlight specific areas requiring attention, particularly those related to problem-solving and communication, where performance does not fully exceed expectations. At the same time, certain physical environment attributes appear to perform above their relative importance, suggesting potential overinvestment. Overall, the study contributes to the service quality literature by emphasizing the temporal and multidimensional nature of service evaluation, particularly in complex service settings such as hotel events. By focusing on the perspective of event organizers, it also extends existing research beyond traditional end-user approaches, offering a more nuanced understanding of how service quality is assessed in business-to-business contexts.

REFERENCES

Ali, F. M., Anuar, J., Saien, S., Karim, N. A., & Shattar, N. A. (2023), "Exploring hotel facilities and event promotion in influencing customer satisfaction with event design: Cases of 4 star hotel in Terengganu", International Journal of Business and Technology Management, 5(S1), 126–135.

Alnawas, I., & Hemsley-Brown, J. (2019), "Examining the key dimensions of customer experience quality in the hotel industry". *Journal of Hospitality Marketing & Management*, *28*(7), 833-861.

Armbrecht, J. (2021), "Event quality, perceived value, satisfaction and behavioural intentions in an event context", *Scandinavian Journal of Hospitality and Tourism*, *21*(2), 169-191.

Boo, S., & Busser, J. A. (2018), "Tourists' hotel event experience and satisfaction: an integrative approach", *Journal of Travel & Tourism Marketing*, *35*(7), 895-908.

Berners, P., & Martin, A. (2022), "*The practical guide to achieving customer satisfaction in events and hotels*". Routledge.

Chandra, S., Bhalla, R., & Agarwal, S. (2022), "Business events and hospitality", In The Routledge handbook of business events (pp. 263-272). Routledge.

Cheruon, R. C., & Korir, J. C. (2023), "Dynamic capability and competitive advantage of event management ventures in Kenya", Research in Hospitality Management, 13(3), 161-165.

Cronin Jr, J. J., & Taylor, S. A. (1992). "Measuring service quality: a reexamination and extension.", Journal of marketing, 56(3), 55-68.

Hutson, N. (2022), "Strategic event planning in hospitality: The need to recommit, post COVID", *Journal of Global Hospitality and Tourism*, *1*(1), 88-89.

Kulu, B. H., & Bola, I. O. (2023), "Event service quality and customer satisfaction in the hospitality industry in Bida, Nigeria", International Journal Social Science Humanities Research, 6(2), 35–48.

Kusuma, P. Y. A., Aryana, N. R., & Sari, G. A. M. K. K. (2021), "The application of hotel strategies based on factor analysis to improve the number of events at The Trans Resort Bali", Journal of Applied Sciences in Travel and Hospitality, 4(1), 27–34.

Nain, A. (2018), "A study on major challenges faced by hotel industry globally", *International Journal of Creative Research Thoughts*, *6*(1), 561-571.

Parasuraman, A. B. L. L., Zeithaml, V. A., & Berry, L. (1988). "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality.", *1988*, *64*(1), 12-40.

Prakash, G. (2019), "Understanding service quality: insights from the literature", Journal of Advances in Management Research, 16(1), 64-90.

Ramya, N., Kowsalya, A., & Dharanipriya, K. (2019), "Service quality and its dimensions", *EPRA International Journal of Research & Development*, *4*(2), 38-41.

Sipe, L. J., & Testa, M. R. (2018), "From satisfied to memorable: An empirical study of service and experience dimensions on guest outcomes in the hospitality industry", *Journal of Hospitality Marketing & Management*, *27*(2), 178-195.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

A Quality by Design Framework for Mobile-Based Household Food Waste Reduction

Alberto Piovano[1]*, Rossella Palermo[1], Elisa Verna[1], Maurizio Galetto[1]

[1] Politecnico di Torino, Department of Management and Production Engineering, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
*Corresponding author (mail alberto.piovano@polito.it)

***Abstract*— Household food waste remains one of the most damaging waste streams in environmental and economic terms, and progress toward SDG 12.3 at the consumer stage is still limited. Household food management can be framed as a multi-stage process (planning, purchasing, storage, cooking, consumption and leftovers) in which waste emerges as a quality defect driven by weak process control. Digital technologies—mobile applications, AI and smart tracking—can function as quality tools by enabling measurement, decision support and feedback across this journey. However, current market offerings are typically narrow in scope and designed around isolated stages, resulting in fragmented support and limited behavioural fit. This paper combines (i) a structured review of behavioural drivers and barriers, (ii) a benchmarking of existing mobile solutions and (iii) a user survey to prioritise design requirements. The evidence is synthesised into an integrated Quality-by-Design framework that maps digital controls to each stage of household food management through a measurement–control–feedback logic. A conceptual prototype (FoodFlow) illustrates the framework through meal planning, inventory and expiry management, AI-assisted recipe suggestions, engagement mechanisms and a sustainability-impact dashboard. The contribution is a replicable, quality-driven design route for digital interventions addressing household food waste, with actionable guidance for developers, researchers and policymakers.**

Keywords – household food waste, mobile application design, quality in design, sustainable consumption, digital quality

I. Introduction

Household food waste (HFW) remains an ethical and ecological challenge, shaped less by a lack of awareness than by the everyday organisation of domestic food management. Recent literature increasingly conceptualises consumption-stage waste as the cumulative outcome of breakdowns across planning, purchasing, storage, preparation, consumption, and leftover management, through which small failures compound into recurrent losses (Roodhuyzen *et al.*, 2017; Schanes *et al.*, 2018). Recent categorisation studies further suggest that effective prevention requires linking what is wasted to how it becomes waste, thereby shifting attention from downstream disposal to upstream mechanisms and control points (Mohamed Yusoff *et al.*, 2024).

Two research streams point to plausible routes for mitigation. Behavioural studies show that routines and capabilities (particularly planning competence, shopping practices, and leftover reuse) shape waste outcomes and can be influenced through targeted interventions (Romani *et al.*, 2018; Stancu *et al.*, 2016; Vittuari *et al.*, 2023). In parallel, data-driven and digital approaches suggest that technology may help reduce cognitive burden by restoring inventory visibility, structuring decisions, and making outcomes more legible (Agya, 2025; Woolley *et al.*, 2022). However, evidence on consumption-stage interventions remains uneven and highly contingent on design quality and evaluation rigour (Reynolds *et al.*, 2019). Moreover, despite the proliferation of digital platforms aimed at reducing food waste, many existing solutions remain functionally fragmented, supporting isolated tasks rather than end-to-end household food management (HFM) (Cane and Parra, 2020). This fragmentation shifts the burden of integration onto users, increasing friction and limiting sustained adoption.

Against this background, this study examines how digital interventions for HFM can be designed to support end-to-end waste prevention across the household food journey. It addresses this design gap by translating consolidated evidence into a Quality-by-Design (QbD) logic for HFM. The paper makes four contributions: (i) it conducts a literature review to identify candidate design requirements for HFM; (ii) it develops a capability mapping of existing applications across the household food journey to identify coverage patterns and control gaps; (iii) it reports a user requirements survey to prioritise functionalities and adoption-relevant criteria; and (iv) it proposes an integrated Measurement-Control-Feedback (MCF) framework, translated into a system architecture and implemented in a high-fidelity prototype, FoodFlow, providing a basis for subsequent empirical evaluation.

II. Background

Consumption-stage food waste is often framed as a behavioural issue. Yet the literature consistently shows that it arises from a broader configuration of factors in which household practices interact with infrastructures, product characteristics, information conditions, and decision heuristics over time (Roodhuyzen *et al.*, 2017; Schanes *et al.*, 2018). This complexity makes both measurement and comparison difficult. Definitions vary across studies, and many categorisation schemes do not adequately connect waste types to the mechanisms that generate them, thereby limiting the design of targeted prevention strategies.

A process perspective helps explain why HFM is especially vulnerable to waste. HFM is not a single act of disposal, but a chain of interdependent stages (planning, purchasing, storage, preparation and consumption, and leftover management) through which small informational and execution failures may accumulate into waste events. In this

study, that process lens is used as an analytical abstraction rather than as a literal equivalence between household life and industrial production. Domestic food management remains socially situated, contingent, and shaped by routines, relationships, and uneven capabilities; however, a stage-based representation helps identify where waste-generating breakdowns recur and where preventive intervention is most feasible. Figure 1 provides this design-oriented synthesis by linking stage-specific breakdowns to corresponding Measurement-Control-Feedback (MCF) design opportunities, while Table I consolidates the main stage-specific criticalities, dominant behavioural drivers, and supporting references.

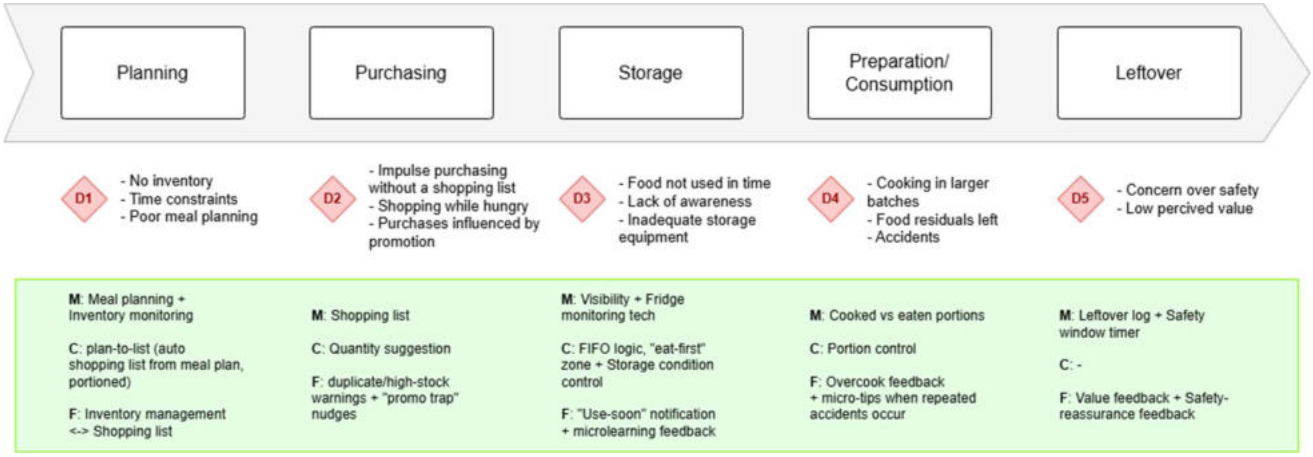


Fig. 1. Stage-based diagnostic map of HFM, analytically representing the main stages of the household journey, the dominant waste-generating breakdowns (D1-D5), and the corresponding Measurement-Control-Feedback (MCF) design opportunities. D = dominant breakdown; M = measurement function; C = control function; F = feedback function.

TABLE I. MAIN STAGE-SPECIFIC CRITICALITIES, DOMINANT BEHAVIOURAL DRIVERS, AND KEY REFERENCES ASSOCIATED WITH HOUSEHOLD FOOD WASTE ACROSS THE HFM JOURNEY

Stage	Main criticality	Dominant behavioural driver	Key ref.
Planning	Planning without checking home stock cause mismatch between needs and supplies.	Weak planning and organisational routines (stock-checking, meal planning, household coordination).	Roodhuyzen et al. (2017)
Purchasing	Over-purchasing (off-list / impulse) amplified by promotions and "shopping while hungry".	Shopping practices shaped by impulse and attraction to special offers (low adherence to list).	Vittuari et al. (2023)
Storage	Food not used in time / "forgotten food"; spoilage due to inadequate storage conditions (incl. cross-moldiness).	Weak storing routines and low stock visibility (keeping food too long; inadequate packaging/temperature; not using freezer).	Woolley et al. (2022)
Preparation / Consumption	Cooking/serving too much; residuals left on plate/packaging; accidents (burning/overcooking).	Portioning & preparation skills/heuristics (difficulty scaling recipes; "better too much than too little").	Reynolds et al. (2019)
Leftover	Leftovers not planned for / not consumed; safety concerns; low perceived value → discard.	Attitudes + perceived food safety + limited skills for preserving/repurposing leftovers.	Stancu et al. (2016)

Across the HFM journey, waste emerges from recurrent breakdowns with different behavioural and informational profiles. In planning, meals are often organised without checking existing household stocks, reflecting weak organisational routines and limited coordination capacity (Manika *et al.*, 2022). Food preparation planning, by contrast, has been identified as a protective competence against waste (Boulet *et al.*, 2021). In purchasing, overbuying and impulse decisions translate upstream planning gaps into downstream surplus (Vittuari *et al.*, 2023). During storage, low inventory visibility and inadequate preservation practices increase the likelihood of forgotten food and spoilage (Mohamed Yusoff *et al.*, 2024). In preparation and consumption, overproduction and portioning heuristics generate avoidable residuals, while only a limited subset of portion-focused interventions has shown quantified effects in the evidence base. Finally, leftovers are frequently devalued or discarded because of safety concerns and weak reuse routines, despite their central role in explaining household waste outcomes.

The intervention and technology literature adds a further constraint: solutions must not only exist, but also be adopted and demonstrate measurable impact. Reynolds et al. (2019) review consumption-stage interventions and highlight a limited evidence base, marked by heterogeneity in design, metrics, and reproducibility, thereby calling for more robust evaluation and more standardised reporting. In parallel, technology-oriented research suggests that data-driven and digital systems may absorb part of the cognitive complexity that households struggle to manage. Woolley et al. (2022) propose a data-driven inventory and recipe-optimisation logic to support more organised and predictable purchasing and consumption when human cognition cannot realistically process competing trade-offs. Mapping studies of digital platforms likewise emphasise the rapid expansion of food-waste-related platforms, while also calling for clearer classifications and more user-centred forms of innovation that connect technology to everyday practices (Cane and Parra, 2020). A global review further shows that the effectiveness of technological solutions remains

contingent on consumer behaviour and that adoption barriers vary across income contexts, requiring context-sensitive combinations of technological design and behavioural nudges (Agya, 2025).

Taken together, this literature suggests that HFW can be approached as a stage-structured design problem within HFM, in which failures of visibility, routine execution, and learning accumulate across the household journey. On this basis, Figure 1 and Table I are used not as literal process models, but as design devices for a QbD interpretation of HFM. In this context, QbD refers to the systematic translation of stage-specific breakdowns into intervention requirements, while the MCF logic specifies the minimum design functions needed to support visibility, decision-making, and behavioural adjustment. This provides the conceptual basis for the framework developed in the following sections.

III. Methodology

The study followed a structured multi-step design to translate prior evidence on HFW into the specification of a coherent digital intervention. The workflow combined four stages: (i) a literature review, (ii) capability mapping of existing applications, (iii) a user requirements survey, and (iv) the translation of convergent findings into an integrated design framework, system architecture, and prototype.

A. Literature review and requirement elicitation

A literature search and screening process guided by the PRISMA method (Page *et al.*, 2021) was conducted to identify established drivers, barriers, and intervention leverage points related to HFW at the consumption stage. The search was conducted in Scopus and covered publications from 2010 to 2025, a period consistent with both the diffusion of digital technologies in everyday life and the growing academic attention to food waste reduction and sustainability. The search string was applied to titles, abstracts, and keywords: "food waste" AND ("household" OR "downstream" OR "customer") AND ("technolog" OR "platform" OR "app" OR "digital") AND NOT (crop OR "anaerobic digestion" OR "agriculture"). Inclusion criteria were limited to English-language publications in Environmental Science, Engineering, Computer Science, Social Sciences, and Economics/Management, and to document types considered methodologically reliable, including journal articles, conference papers, and review studies. Studies were retained only when they explicitly addressed the household. The screening process resulted in a final corpus of 47 studies.

B. Capability mapping of existing applications

To complement the literature-derived requirements with an assessment of currently available solutions, a market analysis was conducted on consumer-facing digital applications and web platforms aimed at reducing HFW. The scan was carried out in October 2025 and combined targeted searches in the Apple App Store, Google Play Store, and public product pages, using descriptors related to food waste, pantry management, expiry tracking, meal planning, and food sharing. Solutions were screened to retain household-relevant tools and exclude applications narrowly tied to a single retailer or to professional food-service settings. A total of 30 applications/platforms were retained and coded against a capability taxonomy aligned with the stages of the household journey and with selected cross-cutting functions. Capability intensity was assessed using an ordinal scale (absent, present, partial, robust) to support structured comparison across solutions. This step was used to identify recurring strengths, areas of technological maturity, and underdeveloped functions relevant to subsequent requirement definition.

C. User requirements survey

The candidate requirement set was then examined through a user requirements survey designed to prioritise functionalities and identify adoption constraints. The questionnaire was administered online via Google Forms and yielded 88 valid responses from participants in Italy involved in HFM. Participation was voluntary and anonymous, and respondents were informed about the purpose of the study before completing the questionnaire. The sample was not intended to be statistically representative; rather, it was used to capture user-perceived priorities relevant to early-stage design definition. It showed a predominance of women (71.4%) over men (28.6%), and a marked prevalence of respondents aged 25-34 (61.4%), followed by the 45-54 group (17.0%), 18-24 (9.1%), 35-44 (8.0%), and a small share (4.5%) aged over 55. The questionnaire consisted of approximately 20 items organised into five sections: socio-demographic profile, food management habits, evaluation of desired technological functions, usage preferences, and open comments. The instrument combined Likert-type items (1-5), multiple-choice questions, and open-ended responses. Quantitative data were analysed using descriptive statistics (frequencies, percentages, and mean values), while open responses were coded qualitatively to extract additional insights on user needs, perceived barriers, and adoption preferences. The survey was used to assess whether the requirements identified through the literature review and market analysis were consistent with user-perceived needs and to identify possible mismatches with implications for service design.

D. From requirements to architecture and implementation

The convergent findings from the literature review, capability mapping, and survey were interpreted and translated into four Critical-to-Quality (CTQ) design characteristics: low cognitive load, state traceability, anticipatory decision support, and interpretable feedback. These CTQs were then mapped onto a MCF logic across the stages of the household journey, yielding the Integrated Design Framework. The same mapping informed a layered system architecture specifying how state capture, decision support, and feedback can be orchestrated as a continuous service loop rather than as fragmented utilities. Finally, a high-fidelity prototype (FoodFlow) was developed to implement the architecture in an inspectable interface and to provide a concrete basis for subsequent empirical evaluation.

The literature review indicates that scholarly attention is concentrated upstream in the HFM journey. Planning, purchasing, and especially storage emerge as the most extensively addressed stages, whereas leftover management remains comparatively marginal despite its recurrent association with avoidable waste (Reynolds *et al.*, 2019; Schanes *et al.*, 2018). Across the reviewed studies, the most frequently recurring strategies include inventory tracking, expiry management, recipe-based use of available ingredients, shopping-list support, and lighter nudging, with a particularly strong emphasis on storage-related visibility and timing problems, where breakdowns appear more readily measurable and actionable (Woolley *et al.*, 2022). By contrast, more peripheral strategies, such as food scanning and gamified engagement mechanisms, appear less frequently represented. Two implications are especially relevant for design. First, several strong alignments emerge between recurring breakdowns and feasible intervention logics: date-label confusion aligns with expiry management and packaging-related information; inefficient purchasing and low stock visibility align with inventory tracking and smart-list functions; and spoilage aligns with stock monitoring and expiry-risk signalling (Stancu *et al.*, 2016). Second, important blind spots remain. Leftovers are often treated as an extension of inventory rather than as a primary design object, even though their disposal is shaped not only by information deficits but also by safety concerns, perceived risk, and low residual value. Likewise, impulse purchasing and promotion-driven overbuying appear only partially addressable through list-based tools, suggesting limits to purely informational interventions when waste is co-produced by retail environments and situational cues (Vittuari *et al.*, 2023).

The screening of currently available digital solutions, as described in Section III.B, suggests an active but uneven technology landscape across the HFM journey. As shown in Figure 2, capability coverage is concentrated around selected functions rather than distributed across an end-to-end process logic. The App×Function mapping indicates that relatively mature features include shopping lists, inventory tracking, and expiry alerts, whereas more integrated capabilities (such as linking household inventory to meal planning, anticipatory "use-it-up" support, or structured leftover management) appear less consistently represented. This pattern is consistent with literature showing that HFW is embedded in everyday routines and practices, and emerges from the interaction between habits, situational constraints, and operational capabilities (Schanes *et al.*, 2018; Stancu *et al.*, 2016; Vittuari *et al.*, 2023).

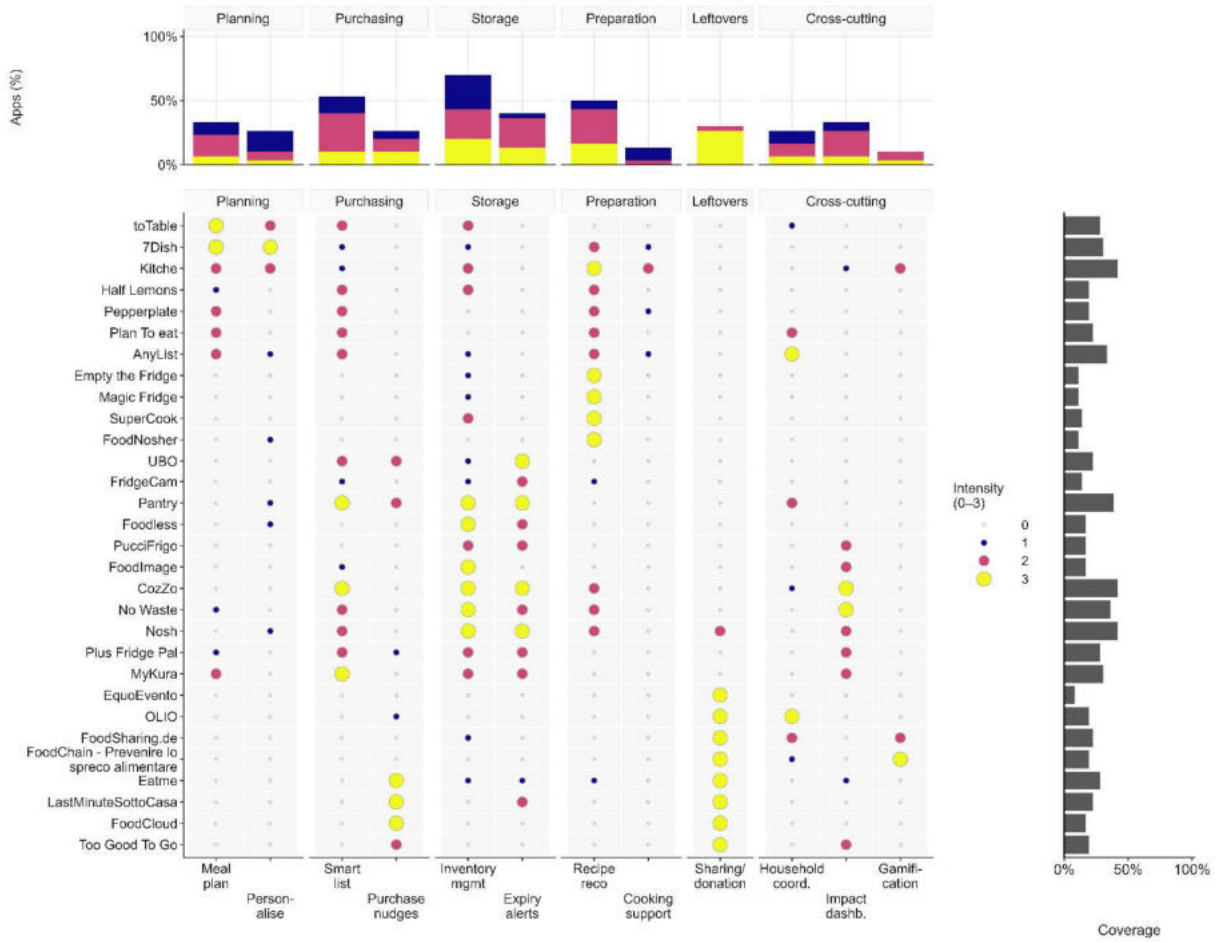


Fig. 2. Capability mapping of currently available consumer-facing applications across the HFM journey. The figure compares application coverage and implementation intensity across stage-specific and cross-cutting functions.

The asymmetry becomes clearer at cluster level (Table II). Inventory and tracking apps tend to concentrate on measurement and basic control functions, but are generally weak in representing leftovers as a distinct object of management. Meal-planning tools are often only loosely connected to actual household inventory and rarely include a strong "use-it-up" logic for food already at home. Recipe engines typically support reactive use rather than continuous monitoring or impact feedback, while redistribution platforms mainly operate further downstream, addressing surplus after it has already materialised. Taken together, these patterns indicate functional fragmentation: achieving broad coverage across the household journey would often require combining multiple tools, with repeated data entry and potential usability friction.

TABLE II. CLUSTER-LEVEL SYNTHESIS OF EXISTING APPLICATIONS, MAIN GAPS, AND QUALITY-CONTROL CAPABILITIES

App Cluster	Representative Apps	Notable Gap	Control capability
Inventory and Tracking	CozZo, Kitche, Nosh, No Waste	Often weakly integrated with upstream planning. Although some include automated data entry (e.g., OCR/vision), leftovers are generally treated as generic inventory rather than as a distinct management object.	M, C, F
Meal Planning & Purchasing	AnyList, Plan To Eat, Pepperplate, 7Dish	Often loosely connected to actual household inventory. They may support list-making and spending control, but rarely incorporate a strong "use-it-up" logic for food already at home and close to expiry.	C
Recipe Engines	SuperCook, Magic Fridge, Empty the Fridge	Primarily support reactive recipe generation and generally lack continuous measurement functions or impact-oriented feedback.	–
Sharing and Redistribution	Too Good To Go, OLIO, FoodCloud	Mainly address downstream surplus redistribution rather than upstream household management processes. Their contribution is therefore more visible at the mitigation stage than at the prevention stage..	M, F

Note: M = measurement capability; C = control capability; F = feedback capability; - = capability not clearly represented at cluster level.

The survey results (collected as per Section III C) are broadly consistent with this interpretation. A substantial share of respondents reported using no monitoring tool (55.7%), while only 3.4% relied on dedicated applications.

Panel A of Figure 3 highlights the most frequently reported difficulties in household food management: food expiring before consumption (35%), lack of inventory visibility (30%), leftover management (26%), and insufficient meal planning (22%). These results do not establish causal mechanisms, but they do suggest that respondents perceive visibility, timing, and routine coordination as recurring points of difficulty within the household food journey. This interpretation is aligned with prior literature identifying planning, shopping, storage, and leftover practices as major contributors to household food waste (Schanes *et al.*, 2018; Stancu *et al.*, 2016).

Panel B of Figure 3 shows a clear prioritisation of operationally useful features. Expiry reminders, recipe suggestions, shopping lists, and meal planning received the highest average ratings, whereas the gamified leaderboard was consistently rated lowest. Rather than rejecting engagement mechanisms altogether, this pattern suggests that respondents place greater value on functions that directly support everyday food management tasks. In design terms, this implies that engagement features may be better positioned as secondary enhancers rather than as primary drivers of adoption, especially when core visibility and planning needs remain insufficiently supported.

Panel C shifts attention from individual features to broader adoption-related criteria. Ease of use was the most frequently selected characteristic (85%), followed by personalisation (58%) and interface clarity (51%), while privacy (27%) and cost (22%) were less frequently indicated. Given the exploratory nature of the sample, these results should be interpreted cautiously; however, they suggest that perceived usefulness alone is unlikely to sustain adoption if digital support is experienced as cognitively demanding or poorly aligned with household routines. From this perspective, usability and personalisation emerge not as peripheral refinements, but as relevant conditions for translating technical functionality into everyday use.

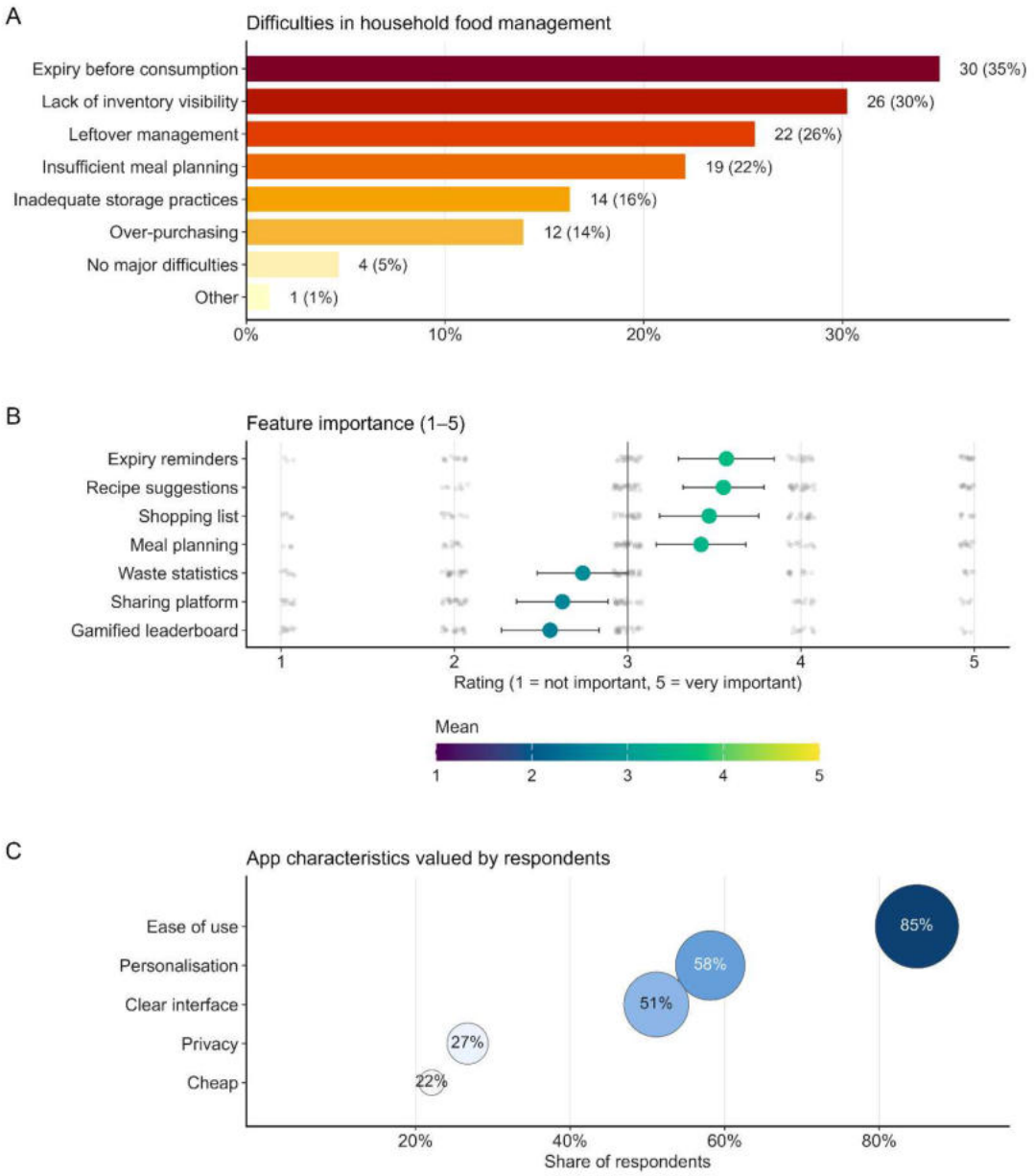


Fig. 3. Survey results on user difficulties, feature priorities, and adoption criteria for digital household food management tools.

V. Results II – Design Framework and Prototype

Taken together, the market scan and survey provide a descriptive basis for a design-oriented interpretation of HFM. Rather than treating HFW solely as a disposal-stage outcome, the findings suggest that relevant difficulties are distributed across planning, purchasing, storage, preparation, and leftover management, and are often associated with limited visibility, weak coordination, and discontinuous decision support. On this basis, HFM can be usefully interpreted through a quality-engineering lens as a stage-structured management problem in which information gaps and routine breakdowns accumulate over time. This interpretation does not imply that domestic life functions like an industrial process in a literal sense; rather, it provides an analytical basis for identifying where digital intervention may be most effective.

Figure 4 presents the Integrated Design Framework, which applies a MCF logic across the household journey. At each stage, the framework specifies how process-relevant state information may be captured (Measurement - M), translated into decision support or corrective action (Control - C), and returned to the user as signals that can inform subsequent choices (Feedback - F). In this sense, the framework offers a design proposition inspired by iterative improvement logic, including the rationale of the Plan-Do-Check-Act (PDCA) cycle (Deming, 2000), while adapting it

to the context of household food management. Its purpose is not to claim full process controllability, but to identify a structured basis for reducing avoidable breakdowns through better visibility, coordination, and feedback.

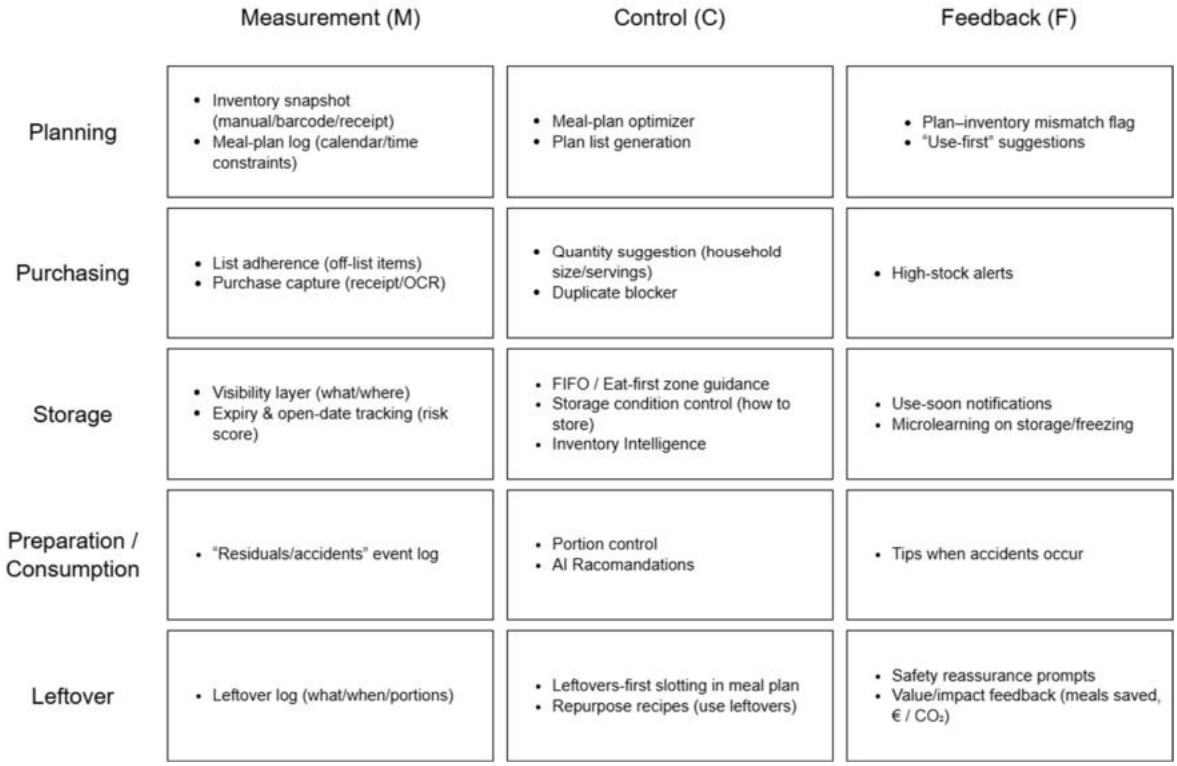


Fig. 4. Integrated Measurement-Control-Feedback (MCF) framework across the household food management journey.

Figure 5 translates this framework into a layered conceptual architecture. The presentation layer is designed around user-facing decisions with immediate operational relevance, such as what is expiring, what can be cooked, and how meal planning should be organised. The logic layer specifies a set of interacting service components, including inventory intelligence, meal-plan optimisation, recipe recommendation, nutritional guidance, gamification/nudging, and impact statistics. The data layer combines internal records and external data sources to support these functions. Privacy and security are positioned as transversal design constraints rather than ex post additions. At this stage, the architecture should be understood as a feasible design configuration intended to operationalise the framework, not as a validated implementation claim.

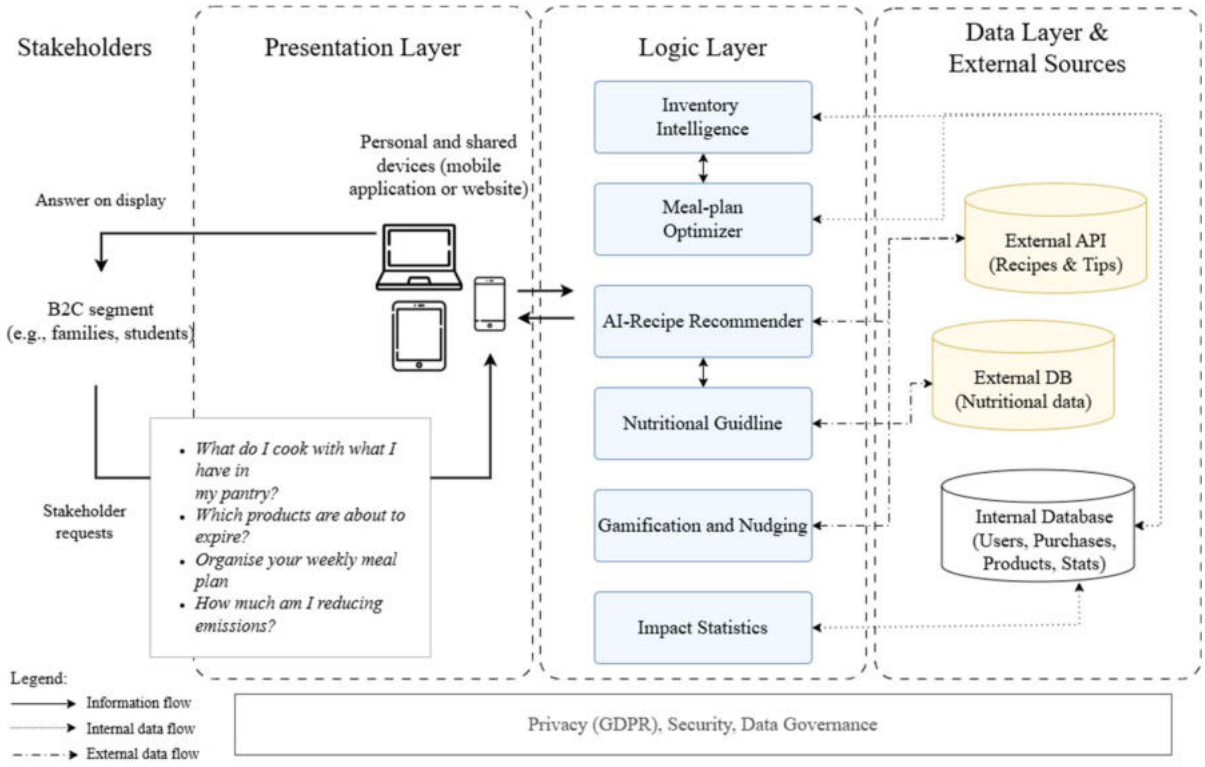


Fig. 5. Layered system architecture translating the Measurement-Control-Feedback (MCF) framework into an executable digital service.

Figure 6 presents FoodFlow as a high-fidelity conceptual prototype that instantiates the proposed logic at interface level. Three modules illustrate the design proposition. The dashboard (Figure 6(a)) aggregates household state indicators, including item counts, expiry alerts, nutritional plan adherence, and estimated CO2-related impact metrics, into a single operational view. The Smart Pantry (Figure 6(b)) illustrates the Measurement dimension by displaying quantity, expiry status, and urgency cues, while linking at-risk inventory to recipe suggestions. The Weekly Meal Plan (Figure 6(c)) illustrates how planning, portion adjustment, and downstream list generation can be connected within a single workflow. Taken together, these modules do not demonstrate effectiveness or adoption in real-world settings; rather, they show that the proposed MCF logic can be translated into an integrated and inspectable interface, thereby providing a testable basis for subsequent empirical validation.

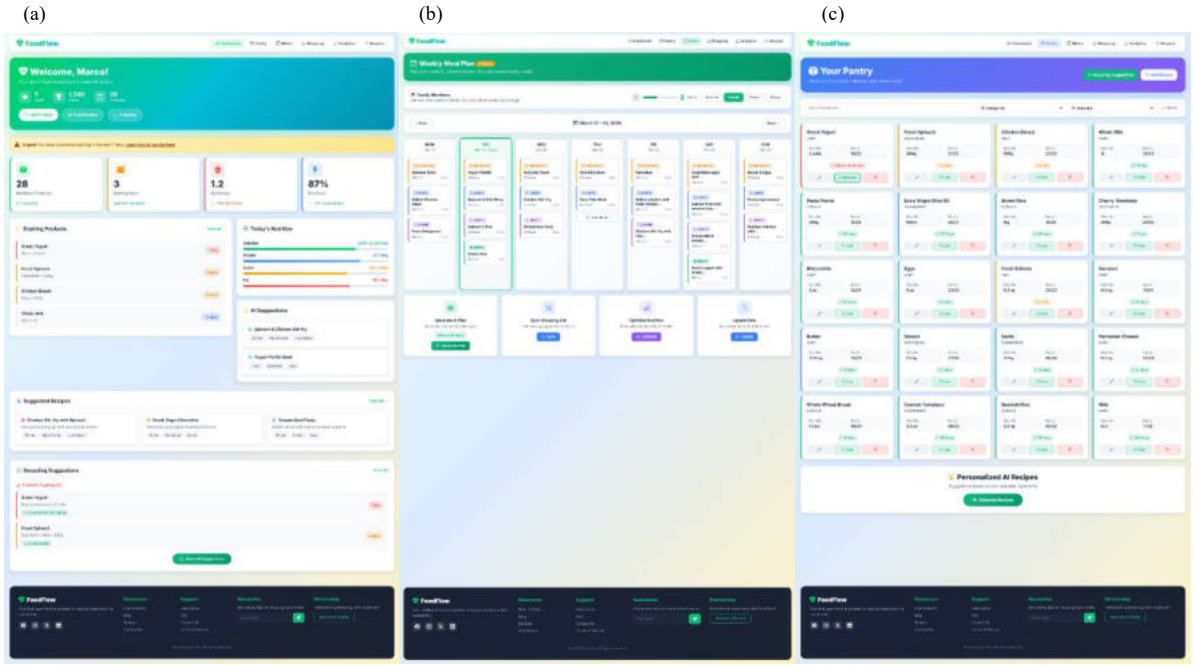


Fig. 6. FoodFlow high-fidelity prototype: (a) dashboard, (b) Smart Pantry, and (c) Weekly Meal Plan.

VI. Discussion and Conclusions

Household food waste should not be treated only as a matter of consumer moral education. The findings of this study suggest that it can also be interpreted as a stage-structured management problem, in which waste emerges from recurring breakdowns in visibility, coordination, and decision support across the household food journey. The principal contribution of this research lies in identifying recurrent forms of functional fragmentation in the current digital landscape and in proposing an Integrated MCF Design Framework to address them. By mapping MCF functions onto each stage of household food management, the framework adapts a quality-oriented design logic (Deming, 2000; ISO 9001:2015) to the domestic context. The conceptual prototype, FoodFlow, shows that the proposed framework can be translated into a coherent and inspectable digital service configuration, thereby providing an empirically grounded and theoretically structured design proposition for subsequent real-world testing, even though its effects on household food waste reduction have not yet been established empirically. Consistent with emerging literature on sustainable digital services (Agya, 2025; Cane and Parra, 2020; UNEP, 2021; Woolley et al., 2022), the system is designed to support prevention upstream rather than merely record waste downstream. By reducing data-entry friction and strengthening planning and visibility functions, it offers a plausible design response to the gap between pro-sustainability intentions and the practical constraints of everyday food management.

The implications of this study extend beyond software development and are relevant to both service design and public policy. If passive information campaigns are insufficient, digital interventions should be designed to reduce cognitive load and support timely action within everyday routines. In this sense, the paper points toward a form of "friction-reducing sustainability", in which recovering at-risk food becomes easier and more actionable within the household workflow. This perspective is compatible with behavioural intervention logic and with the SDG 12.3 objective of reducing per-capita food waste (FAO, 2019). For policymakers and designers, the MCF framework may also serve as an evaluative lens for comparing digital solutions, shifting attention from the mere number of features offered to the extent to which systems support visibility, decision-making, and feedback across the main stages of household food management.

This study also has some limitations to be mentioned. First, the empirical component relies partly on self-reported data and is therefore subject to recall and social desirability bias. Second, the survey sample is heterogeneous but skewed toward younger and more digitally familiar respondents, which limits the transferability of the observed preferences. Third, the proposed architecture remains at prototype stage, and neither the behavioural effects of the interface nor the long-term reliability of AI-assisted recommendations have yet been tested in operational settings. These limitations constrain the causal and generalisable claims that can be made at this stage, while leaving the conceptual contribution of the framework intact.

Future research should therefore move from design plausibility to empirical validation. A first step would be the development of a minimum viable product (MVP) to test user interaction over time and compare alternative nudging and feedback strategies. A second step would be more rigorous outcome evaluation through longitudinal field studies or randomised controlled trials using objective indicators, including physical waste measurements, rather than self-reported estimates alone. A third direction concerns external validity: the MCF framework could be adapted and tested in adjacent contexts, such as food service or small retail, where stage-based process logic may also help reduce avoidable waste.

Acknowledgment

This paper is part of the project NODES which has received funding from the MUR-M4C2 1.5 of PNRR funded by the European Union – NextGenerationEU (Grant agreement no. ECS00000036).

References

Agya, B.A. (2025), "Technological solutions and consumer behaviour in mitigating food waste: A global assessment across income levels", Sustainable Production and Consumption, Elsevier, Vol. 55, pp. 242–256, doi: 10.1016/J.SPC.2025.02.020.

Boulet, M., Hoek, A.C. and Raven, R. (2021), "Towards a multi-level framework of household food waste and consumer behaviour: Untangling spaghetti soup", Appetite, Academic Press, Vol. 156, p. 104856, doi: 10.1016/J.APPET.2020.104856.

Cane, M. and Parra, C. (2020), "Digital platforms: mapping the territory of new technologies to fight food waste", British Food Journal, Emerald Group Holdings Ltd., Vol. 122 No. 5, pp. 1647–1669, doi: 10.1108/BFJ-06-2019-0391/FULL/PDF.

Deming, W.E. (2000), Out of Crisis - Deming's 14 Points for Management, MIT Press.

FAO. (2019), The State of Food and Agriculture 2019, The State of Food and Agriculture 2019, FAO, doi: 10.4060/ca6030en.

"ISO 9001:2015 - Quality management systems — Requirements". (n.d.). , available at: https://www.iso.org/standard/62085.html (accessed 28 October 2025).

Manika, D., Iacovidou, E., Canhoto, A., Pei, E. and Mach, K. (2022), "Capabilities, opportunities and motivations that drive food waste disposal practices: A case study of young adults in England", Journal of Cleaner Production, Elsevier, Vol. 370, p. 133449, doi: 10.1016/J.JCLEPRO.2022.133449.

Mohamed Yusoff, N.I. Bin, Godsell, J. and Woolley, E. (2024), "Towards zero waste: A comprehensive framework for categorizing household food waste", Sustainable Production and Consumption, Elsevier, Vol. 48, pp. 1–13, doi: 10.1016/J.SPC.2024.05.002.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., et al. (2021), "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews", BMJ, BMJ Publishing Group, Vol. 372, doi: 10.1136/BMJ.N71,.

Reynolds, C., Goucher, L., Quested, T., Bromley, S., Gillick, S., Wells, V.K., Evans, D., et al. (2019), "Review: Consumption-stage food waste reduction interventions – What works and how to design better interventions", Food Policy, Pergamon, Vol. 83, pp. 7–27, doi: 10.1016/J.FOODPOL.2019.01.009.

Romani, S., Grappi, S., Bagozzi, R.P. and Barone, A.M. (2018), "Domestic food practices: A study of food management behaviors and the role of food preparation planning in reducing waste", Appetite, Academic Press, Vol. 121, pp. 215–227, doi: 10.1016/j.appet.2017.11.093.

Roodhuyzen, D.M.A., Luning, P.A., Fogliano, V. and Steenbekkers, L.P.A. (2017), "Putting together the puzzle of consumer food waste: Towards an integral perspective", Trends in Food Science & Technology, Elsevier, Vol. 68, pp. 37–50, doi: 10.1016/J.TIFS.2017.07.009.

Schanes, K., Dobernig, K. and Gözet, B. (2018), "Food waste matters - A systematic review of household food waste practices and their policy implications", Journal of Cleaner Production, Elsevier, Vol. 182, pp. 978–991, doi: 10.1016/J.JCLEPRO.2018.02.030.

Stancu, V., Haugaard, P. and Lähteenmäki, L. (2016), "Determinants of consumer food waste behaviour: Two routes to food waste", Appetite, Academic Press, Vol. 96, pp. 7–17, doi: 10.1016/j.appet.2015.08.025.

UNEP - UN Environment Programme. (2021), "Reducing Consumer Food Waste Using Green and Digital Technologies | UNEP - UN Environment Programme", available at: https://www.unep.org/resources/publication/reducing-consumer-food-waste-using-green-and-digital-technologies (accessed 25 June 2025).

Vittuari, M., Garcia Herrero, L., Masotti, M., Iori, E., Caldeira, C., Qian, Z., Bruns, H., et al. (2023), "How to reduce consumer food waste at household level: A literature review on drivers and levers for behavioural change", Sustainable Production and Consumption, Elsevier, Vol. 38, pp. 104–114, doi: 10.1016/J.SPC.2023.03.023.

Woolley, E., Luo, Z., Jellil, A. and Simeone, A. (2022), "A data driven approach to reducing household food waste", Sustainable Production and Consumption, Elsevier, Vol. 29, pp. 600–613, doi: 10.1016/J.SPC.2021.11.004.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Quality Challenges in Multi-Human Multi-Robot Collaborative Assembly Design: A Coherence-Based Methodology to Analyse Process Failures

Matteo Capponi[1], Aurora Sofia di Matteo[2], Luca Mastrogiacomo[3], Fiorenzo Franceschini[4]

1 Politecnico di Torino, Italy - Dipartimento di Ingegneria Gestionale e della Produzione (DIGEP); matteo.capponi@polito.it
2 Politecnico di Torino, Italy - Dipartimento di Ingegneria Gestionale e della Produzione (DIGEP); aurora.dimatteo@polito.it
3 Politecnico di Torino, Italy - Dipartimento di Ingegneria Gestionale e della Produzione (DIGEP); luca.mastrogiacomo@polito.it
4 Politecnico di Torino, Italy - Dipartimento di Ingegneria Gestionale e della Produzione (DIGEP); fiorenzo.franceschini@polito.it

***Abstract*- As collaborative robotics in manufacturing evolves from single-human single-robot (1H-1R) collaboration to multi-human multi-robot (MH-MR) teams, interaction complexity increases significantly. Although the traditional Failure Mode, Effects and Criticality Analysis (FMECA) tool is widely adopted in quality engineering, in collaborative manufacturing processes it is primarily structured around 1H-1R interactions. This makes it less effective at capturing failure modes that emerge from multi-agent dynamics. This paper therefore proposes an exploratory coherence-based methodology for analysing and evaluating failure modes in MH-MR collaborative assembly processes. The methodology is inspired by the "3Cs" model which identifies Collaboration, Coordination and Communication as critical aspects for effective teamwork. A case-study regarding the collaborative assembly of a skateboard involving two human operators and two cobots is then presented. This highlights how timing misalignments, ambiguous feedback, role ambiguity and spatial interference can affect process and product quality when multiple agents operate in parallel within a shared workspace. The proposed framework provides process designers with a structured tool supporting the development of more robust, human-aware collaborative manufacturing processes in MH-MR systems.**

Keywords: Multi-Human Multi-Robot Collaboration, Process Quality, Assembly, FMECA, Manufacturing, Process Failure

I. INTRODUCTION

Manufacturing systems are undergoing a critical change, driven by the shift from Industry 4.0 to Industry 5.0. The latter emphasises flexibility, resilience and human-centred production [1]. This transition is also reflected in European policy perspectives on Industry 5.0, which frame future manufacturing around human-centricity, sustainability and resilience [2]. In this context, collaborative robotics plays a central role, enabling humans and robots to work closely together and perform manufacturing tasks jointly. Initial industrial applications of human-robot collaboration (HRC) have primarily centred on dyadic configurations, typically comprising one human operator and one collaborative robot operating within a shared workspace [3], [4]. From a quality and reliability perspective, one-human-one-robot (1H-1R) systems have been widely investigated. Previous studies have identified common causes of process failures in collaborative assembly [5], [6]. Well-established quality tools, such as Failure Mode, Effects and Criticality Analysis (FMECA), have been successfully adopted to structure and prioritise these risks. Recent technological advances are enabling a shift towards more complex collaborative configurations characterised by the simultaneous presence of multiple human operators and multiple robots working together on shared assembly tasks, also called "MH-MR" teams. The authors adopt the following definition of a MH-MR team: "*a Multi-Human Multi-Robot (MH-MR) team is a structured group composed of at least three agents, including both human and robotic members, working together towards a shared goal through continuous interaction, information exchange, effective coordination, clear task allocation, and dynamic collaboration between agents throughout the production process*" [7]. Multi-human-multi-robot (MH-MR) teams offer significant potential benefits in terms of flexibility, scalability and workload distribution, particularly in complex, low-volume assembly environments [7], [8], [9]. However, the increased number of interacting agents introduces new dynamics relating to coordination, communication, shared situational awareness and decision-making, substantially increasing system complexity. These dynamics concern both physical collaborations, involving shared workspace occupation, object handling and task execution, and cognitive collaboration, involving decision-making, mutual adaptation, shared awareness and interpretation of system states. The interdependencies between multiple humans and robots result in emergent behaviours at the team level, where disturbances affecting one agent can spread throughout the system. From a quality engineering perspective, this raises important questions about the nature of process errors and failure modes in MH-MR assembly. Failures relating to miscoordination, conflicting task priorities, ineffective communication or loss of shared awareness may not immediately present as safety issues but can significantly impact product quality and operational performance through delays, rework or assembly errors. While the literature covers topics related to one human one robot collaboration [3], [5], [10], there is a lack of systematic classifications of process failure modes tailored to MH-MR collaborative assembly. This paper addresses this gap by proposing an exploratory coherence-based methodology to identify failure modes that are specific to collaborative assembly carried out by teams composed of both humans and robots. Based on existing analyses of failures in 1H-1R collaborative assembly [5], the proposed taxonomy broadens the scope of analysis to include new categories of failures relating to coordination, communication and decision-making at the team level. This exploratory and conceptual contribution aims to frame the problem and provide a structured foundation for future empirical validation and methodological developments. The paper is organised as follows. Section 2 reviews the relevant literature on failure analysis in human-robot collaboration and MH-MR manufacturing systems. Section 3 introduces a conceptual

methodology to analyse process failure modes in MH-MR teams while Section 4 presents a case-study on the collaborative assembly of a skateboard. Section 5 and 6 respectively present and discuss the main results while Section 7 summarises the main conclusions, limitations, and future developments.

II. RELATED WORK

Over the past decade, human-robot collaboration (HRC) in manufacturing, particularly for assembly tasks, has been explored as a means of combining human flexibility with robotic precision [1], [11], [12], [13], [14]. Effective task structuring and coordination are critical for supporting reliable collaboration between humans and robots and exploiting their respective strengths [15], [16], [17]. In recent years, researchers developed comprehensive ways to evaluate the performance of HRC systems. Gervasi et al. [11] addressed this issue by defining multiple "HRC dimensions" and proposing an integrated framework with which to compare different collaboration configurations. In practice, the evaluation of an HRC system usually involves considering metrics such as productivity, team efficiency and ergonomics/operator workload. The concept of "fluency" in HRC has also been introduced to capture how smoothly a human-robot team works as a unit [18]. Recent literature surveys have compiled a wide range of performance indicators and benchmarking tasks for collaborative robots in manufacturing [7], [9]. Overall, the state of the art shows that HRC systems can achieve improved safety and efficiency over manual processes when evaluated on both productivity and human factors [13], [19], [20]. The introduction of HRC to the factory floor brings benefits, as well as novel process failure modes and quality challenges. Coronado et al. [21] evaluated quality in human-robot interaction ("HRI") by providing an overview of factors used to evaluate performance and human well-being quality aspects in HRI applications. Similarly, Verna et al. [22] proposed a novel tool to evaluate manufacturing quality in human-robot collaboration combining the analysis of product defects variant with the evaluation of human well-being. Kokotinis et al. [10] also proposed a novel quality index to assess several aspects of human-robot collaboration, including quality, in production contexts. While these studies focus on quality as a broad and multidimensional concept, a more detailed understanding requires considering how quality is affected by errors and failures occurring during collaboration. Humans and robots working together may misinterpret mutual actions or signals, which can lead to misaligned efforts or even collisions. Such miscoordination can result in assembly errors or quality defects that would not occur in a purely manual or automated process [5]. Antonelli and Stadnicka [23] observe that HRC in assembly can lead to different types of process failures, which are not seen in manual assembly or fully robotic cells, due to the novel interactions between humans and robots. Hence, ensuring process quality in HRC requires addressing new error modes, ranging from human errors (e.g. a worker positioning a part incorrectly for the robot) and robot errors (e.g. misdetections or improper force application) to interface issues (e.g. miscommunication of intentions). Haggy et al. [24] investigated a combination of Bow-Tie Analysis and FMEA for assessing HRC safety. By integrating the two techniques, the authors evaluated a collaborative assembly case study. However, most HRC research to date assumes a dyadic interaction. The upcoming Industry 5.0 brings scenarios in which multiple humans and multiple robots (see Figure 1) collaborate in semi-structured environments [1]. Transitioning from dyadic (i.e., one-to-one) collaboration to team-like configurations has great potential, offering higher flexibility, adaptability and productivity by dynamically allocating human and robotic resources where they are needed. For instance, in a complex assembly process, two robots could assist a pair of human workers on different subtasks simultaneously, or hand parts between each other and the humans [7]. Nourmohammadi et al. [25] have already demonstrated the benefits of this approach: they developed a model for balancing an assembly line that allows for multiple operators (both human and cobot) at the same station. They also showed that enabling such parallel human-robot collaboration can significantly improve line productivity. Despite these early successes, scaling up to MH-MR teams introduces new challenges that largely remain unresolved. As Capponi et al. [7] observe, the manufacturing field is focusing attention to more complex configurations of multi-human multi-robot team, and traditional 1H-1R evaluation models may no longer be adequate. In this regard, Ma et al. [8] showed that the traditional metrics of human-robot interaction (HRI) for performance and efficiency must be expanded to incorporate teamwork metrics, including cognitive and environmental aspects of teamwork. Similarly, Riedelbauch et al. [9] developed a holistic set of metrics and evaluation methods to assess human robot teams, including dimensions such as productivity, safety, flexibility, job quality, etc. Another crucial aspect in designing processes involving multiple agents, is to ensure that the right number of robots assist at the right time, humans are neither overloaded nor idle, and that safety is maintained. Scalability is therefore recognised as a major challenge in human-robot teaming research [26]. Furthermore, aspects such as communication protocols and situational awareness are of critical importance in MH-MR settings: robots must not only communicate with humans, but also with each other while keeping humans informed [27]. Additionally, multiple humans may require new interfaces to manage robot teams collectively. In terms of process quality, maintaining high quality and reliability may become more difficult when more agents are involved. For example, new process failure modes could occur when two robots collaborate on a task (e.g. synchronisation failures between robots), or when multiple humans divide their attention among several robots. Capponi et al. [7] emphasised that integrating multiple humans and robots introduces new challenges such as governing interactions within the team, communication means, and human behaviour. This paper addresses this gap by building on three complementary fields of literature: failure analysis in dyadic HRC, teamwork-oriented evaluation of human-robot teams, and quality-oriented assessment of collaborative manufacturing. These contributions are extended toward a coherence-based analysis of process failures in MH-MR collaborative assembly.

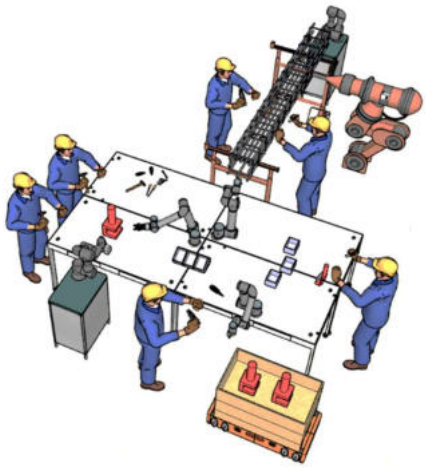


Fig. 1. Multi-human multi-robot team in a general manufacturing scenario

III. METHODOLOGY

From a methodological perspective, this work adopts an exploratory conceptual modelling approach supported by expert-based assessment. To capture emergent process failures in MH-MR teams, the authors adopted a teamwork-oriented perspective based on the "3Cs" model introduced by Ma et al [8], which identifies collaboration, coordination, and communication as the three core dimensions that govern effective teamwork:

- Collaboration: Ensuring that team members share common goals, intentions and plans, and work together towards them. This involves building and maintaining shared mental models among humans and robots. In practice, collaboration means that all parties understand the joint objectives aligning their efforts for mutual success [28].
- Coordination: Ensuring that the interdependent tasks and resources of a team are managed so that the actions of its members are synchronised. In teams comprising both humans and robots, there are inherent informational and physical dependencies between members, thus activities must be carefully sequenced and synchronised to avoid conflict [29]
- Communication: The exchange of information necessary to enable coordination and collaboration over time. As the team works, critical data, status updates and intentions must be transmitted and understood by the agents involved. This encompasses verbal or written messages, signals, cues and interface feedback between humans and robots [30].

In dyadic systems, risks are often confined to the interaction between a single operator and a robot, and failure analysis tends to focus on task- or agent-specific issues. In multi-human multi-robot teams agents involved collaborate and coordinate to successfully perform a task. In this study, the authors hypothesised that the performance of MH-MR teams emerges from maintaining coherence conditions across the process. Coherence refers to the consistent alignment of system components, agent behaviours and shared information necessary for the successful execution of collaborative tasks. Accordingly, a process failure mode can be defined as a violation of a coherence condition necessary for accomplishing a collective task. Based on this assumption, a structured methodology is proposed to guide designers in identifying process failure modes for MH-MR-oriented FMECA. The methodology is composed of three main phases: (i) MH-MR characterisation; (ii) Coherence requirement analysis and (iii) Failure mode derivation.

A. Phase 1: MH-MR characterisation

A MH-MR team can be characterised across six macro-dimensions:

- **Team composition** defined as the structural configuration of agents (i.e., number of humans and robots, types, etc.)
- **Task allocation** refers to the distribution of operational activities among agents
- **Agent responsibilities** regard the clarity of the authority and of the decision-making mechanisms
- **Coordination** defined as the synchronization of interdependent tasks and resources
- **Collaboration** defined as the sharing of common goals, intentions and plans
- **Communication** understood as the exchange of information enabling coordination and collaboration over time

These macro-dimensions define the areas in which coherence must be preserved and in which misalignments may structurally emerge.

B. Phase 2: Coherence requirement analysis

For each macro-dimension, the designer identifies the conditions required for successful task performance. The analysis focuses on multiple coherence features, including:

- **Goal coherence**: alignment between the objectives of individual agents and the overall system goals.
- **Information coherence**: consistency of shared data and representation of the system state.
- **Temporal coherence**: synchronisation of actions and decision timing.
- **Spatial coherence**: compatibility of physical workspace occupation and movement.
- **Cognitive coherence**: shared situational awareness among agents.

Each macro-dimension requires a specific combination of these coherence features. This phase makes explicit the alignment conditions that must be satisfied during operation.

C. Phase 3: Failure mode derivation

Failure modes are systematically derived as violations of the coherence requirements identified in Phase 2. For each macro-dimension and associated coherence features, the designer evaluates: how partial or complete misalignment may occur and how they can propagate across agents and the whole system. The derived failure modes are then organized within a FMECA framework. Each process failure mode is documented using the standard FMECA schema [31]:

- **Process Failure Mode**: A deviation from expected performance within a task or interaction involving one or more agents.
- **Causes**: The specific factors that lead to the occurrence of the failure mode. Causes may be technical (e.g., sensing error), procedural (e.g., misaligned task sequence), or cognitive (e.g., incorrect assumption), and can originate from either human or robot behaviour or their interaction.
- **Effects**: The observable consequences of the failure on task execution, process fluency, product integrity, or team behaviour.
- **Responsible Agents**: Human, Robot, or Joint thus indicating the agent primarily associated with the origin of the failure.

In addition to the standard qualitative descriptions, each failure mode is assigned a Severity (S), Occurrence (O), and Detection (D) rating according to a 1-10 ordinal scale. Severity reflects the impact of the process failure on safety, product integrity, or process continuity (1 = negligible effect; 10 = safety-critical or system-disrupting effect). Occurrence estimates the likelihood of the failure under nominal operating conditions (1 = rare; 10 = frequent). Detection represents the probability that the failure will be identified before leading to significant consequences (1 = easily detectable; 10 = difficult to detect). The industrial case study presented in the following section illustrates the application of the proposed methodology to a specific assembly scenario.

IV. CASE-STUDY: MH-MR COLLABORATIVE ASSEMBLY OF A SKATEBOARD

To support the conceptual development of the proposed methodology, this paper presents a case study of the collaborative assembly of a skateboard (see Figure 2) performed by a multi-human, multi-robot (MH-MR) team.

Fig. 2. (a) Parts of a skateboard and (b) the assembled product

This scenario was designed and executed in a laboratory environment (Mind4Lab, Politecnico di Torino, Italy). Assembly took place on a dedicated workstation equipped with two collaborative robots (Universal Robots UR3e), as shown in Figure 3. The robotic agents were integrated into the workspace to operate in physical proximity with the human operators. The operational team was composed of two human operators and two robotic agents. Each human-robot pair worked on one side of the table, assembling their respective portion of the skateboard in parallel. The two sub-teams performed equivalent tasks simultaneously, requiring coordination during the process for the use of shared

resources and process alignment. In this configuration, the robotic agents mainly performed predefined physical support actions, such as part positioning and tool presentation, while human operators retained most cognitive and adaptive functions, including decision-making, verification, timing adaptation, mutual coordination, alignment and joining. This parallel execution modality was selected to reproduce realistic industrial dynamics and maximise inter-agent interaction phenomena.

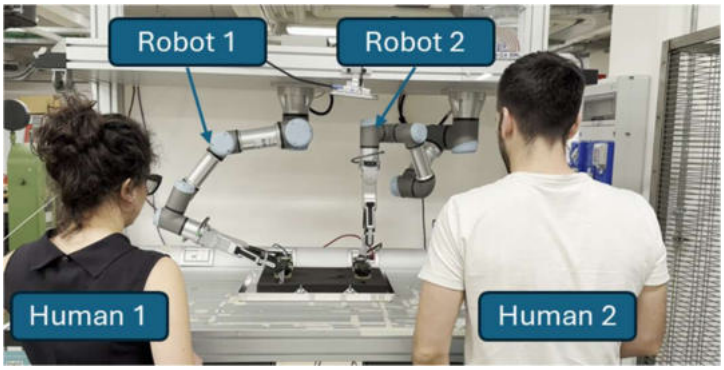


Fig. 3. The assembly work area with two collaborative robots UR3 (Robot 1 and Robot 2) and two human operators (Human 1 and Human 2)

The product to assemble was a standard skateboard comprising three main subsystems: trucks, the board, and the wheels. The process was divided into three macro-phases (see Figure 4): (i) Truck assembly; (ii) Trucks on the board assembly and (iii) Wheels and trucks assembly. The detail steps of the assembly process are provided in Table 1.

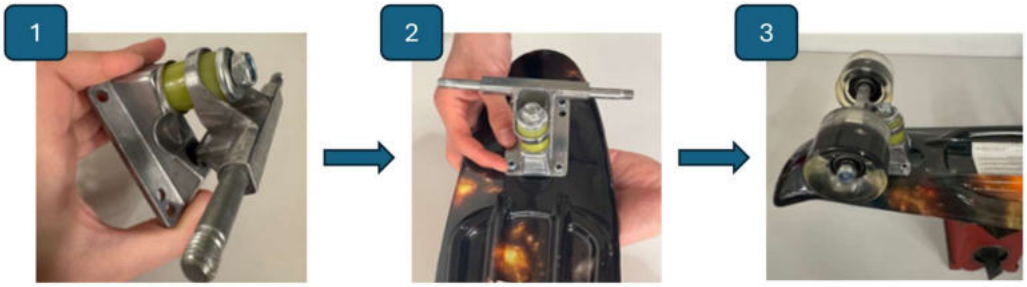


Fig. 4. The three phases of the assembly of the skateboard: (1) truck assembly; (2) trucks on the board assembly (3) wheels and trucks assembly.

Task allocation followed a symmetric model: each human-robot pair was responsible for half portion of the product (front and rear). Table 1 summarises the elementary steps within each macro-phase. While the robotic agents' assignments were predefined, the human operators had some flexibility in their decision-making, particularly with regard to mutual adaptation and action timing.

TABLE I. LIST OF THE ELEMENTARY STEPS NECESSARY TO ASSEMBLE THE SKATEBOARD AND THE RELATED ALLOCATION TO THE AGENT INVOLVED (H1:HUMAN 1, H2: HUMAN 2; R1:ROBOT 1 AND R2:ROBOT 2)

Assembly phase	Detailed steps	Agent involved
A) Assembly of the truck (x2)	A.1 Communicate and align on action	H1/H2
	A.2 Send input to cobot	H1/H2
	A.3 Pick and place the baseplate (BP)	R1/R2
	A.4 Insert the bolt (BL1) into the baseplate (BP), ensuring that the threaded part of the bolt is facing upwards.	H1/H2
	A.5 Insert the larger metal protector (LMP) and then the larger yellow grommet (LYG) onto the bolt (BL1).	H1/H2
	A.6 Insert the black pin cup (PC) on the hanger (HNG) and on the baseplate (BP).	H1/H2
	A.7 Insert the smallest yellow grommet (SYG) and then the smallest metal protector (SMP) onto the bolt (BL1).	H1/H2
	A.8 Screw the nut (N1) onto the bolt, pushing the head of the bolt while tightening and the truck is assembled.	H1/H2
	B.1 Send input to cobot	H1/H2

B) Assembly of the wheels on the truck (x2)	B.2 Pick and place wheel (WL)	R1/R2
	B.3 Mount the wheel on the hanger	H1/H2
	B.4 Insert the washer (W) and tighten with a nut (N3).	H1/H2
	B.5 Pick and place the wheel	R1/R2
	B.6 Mount the wheel on the hanger.	H1/H2
	B.7 Insert the washer (W) and tighten with a nut (N3).	H1/H2
	B.8 Communicate and align on action	H1/H2
	B.9 Send input to cobot	H1/H2
	B.10 Verify input received from both humans	R1/R2
C) Assembly of the truck on the board	C.1 Take the board (B) and put its underside upwards on the specific supports so that the tips of the board (B) protrude from the supports.	R1/R2
	C.2 Put the truck on the board, making sure that the holes in the baseplate (BP) match those in the board (B). (x2)	H1/H2
	C.3 Insert the bolts (BL2) into the four holes, tighten them with the nuts (N2) to join the truck and the board. (x2)	H1/H2

In this configuration, the two operators work in parallel within a shared workspace, performing either symmetrical or complementary actions on mirrored components simultaneously. Explicit interaction between humans occurs at specific stages, particularly when initiating robot actions (e.g. steps A.1 and B.8). At these times, synchronisation and confirmation of readiness are necessary to trigger robot operation. Beyond these explicit interaction points, however, the process involves continuous implicit mutual adjustments between humans. The operators regulate their pace, spatial positioning and manipulation strategies based on their peripheral awareness of each other's actions. These micro-adjustments are not formally described in the task sequence and cannot be fully predicted at the planning stage since they emerge from shared situational awareness and real-time adaptation within the common workspace.

V. Results

Process failure analysis conducted through the extended FMECA framework highlights conventional and emergent failure modes in collaborative assembly involving multi-human multi-robot teams. A series of structured focus groups involving a panel of domain experts was used to identify MH-MR-specific failure modes. The panel was composed of engineers and academic researchers expert in process quality management, human-robot collaboration and collaborative robotics system design. All participants had prior experience of industrial assembly systems. The focus group sessions were conducted using the methodology proposed in this study. The discussion was specifically structured according to the MH-MR macro-domains and the corresponding coherence features identified in section 3. Failure modes were derived by systematically examining potential violations of goal, temporal, spatial, informational and cognitive coherence within the defined MH-MR configuration.

This analysis led to the development of two distinct FMECA tables. The first one (Table 2) addresses the dyadic (1H-1R) configuration and represents an extension of the framework proposed by Maisano et al. [32]. In this table, each process failure mode is associated with the corresponding assembly step, potential causes, effects, responsible agent, and the assigned Severity (S), Occurrence (O), and Detection (D) ratings. The second table (Table 3) concerns the MH-MR configuration. In addition to the conventional FMECA elements (assembly step, failure mode, causes, effects, responsible agent, and S-O-D scores), this extended table includes two additional columns: the MH-MR macro-dimension involved, and the specific coherence feature violated. The S-O-D values were determined through expert judgement during the focus group sessions, based on the participants' experience. It is important to note that failure modes identified in the dyadic configuration remain conceptually valid within MH-MR systems. Execution-level errors such as incorrect assembly sequences, insufficient torque, or component misalignment may still occur locally within a multi-agent team. However, in MH-MR contexts these process failures may propagate differently, interacting with additional agents and coherence constraints. The MH-MR FMECA therefore extends the dyadic analysis. Compared with the dyadic configuration, the MH-MR analysis does not replace execution-level failure modes, but adds failure modes emerging from team-level interaction dynamics, such as temporal desynchronisation, spatial interference, ambiguous feedback, loss of shared situational awareness and unclear task ownership.

TABLE II. FMECA TABLE FOR THE 1H-1R COLLABORATIVE ASSEMBLY OF THE SKATEBOARD

Assembly step	Process Failure Mode	Cause(s)	Effect(s)	Responsible	S	O	D
A.4	Bolt inserted upside down	Human error; lack of visual guide	Assembly defect; disassembly required	Human	6	6	4
A.5	Missing grommet in sequence	Distraction or part skipped	Poor performance; rework	Human	6	6	6
A.5	Grommets and protectors assembled in wrong order	Human forgets step order; unclear visual hierarchy	Assembly defect; disassembly required	Human	6	4	6

A.8/B.4/B.7/ C.4	Nut falls before threading begins	Unstable grip; inattentive handling	Nut lost; rework or component search	Human	4	6	4
A.8/B.4/B.7/ C.4	Insufficient torque applied	Under-tightening due to fatigue or distraction	Loosening under load; safety risk	Human	10	4	8
A.3/B.2/B.5/ C.1	Misaligned part presentation	Poor calibration, difficulties in picking phase	Human struggles to assembly; process delay; risk of dropping	Robot	6	4	6
A.3/B.2/B.5/ C.1	Gripper slippage during hand-over	Inadequate grip force estimation or sensor error	Part falls or bounces; risk of damage or safety incident	Robot	8	4	6
A.4/A.5/A.6/ A.7/A.8	Baseplate slips from gripper due to excessive force by human	Human applies force before verifying robot release or stable grasp	Part drop, rework needed, delay	Joint	8	4	6

The 1H-1R FMECA highlights process failure modes primarily associated with execution-level errors occurring within single human-robot interactions. Most process failures originate from human procedural mistakes or from robot-related handling inaccuracies. In this configuration, failures are typically localised and directly attributable to a specific agent. The Severity (S), Occurrence (O), and Detection (D) ratings reflect this execution-focused nature. For example, insufficient torque application exhibits the highest Severity (S = 10) due to potential safety implications under load, combined with relatively high Detection difficulty (D = 8), since under-tightening may not be immediately visible. In contrast, failures such as a nut falling before threading begins show lower Severity (S = 4) and high detectability (D = 4), as the issue is immediately observable and correctable.

On the other hand, the MH-MR FMECA reveals that many critical process failures in MH-MR systems are not caused by individual agents' actions, but by breakdowns in coordination, communication and collaboration within the team. Coordination failures in MH-MR systems are primarily caused by violations of temporal and spatial coherence. Since multi-agent execution requires procedural coupling, even minor desynchronisation can propagate throughout the system. For instance, if one human agent fails to confirm readiness or input signals are not provided in a timely manner, temporal coherence is disrupted, resulting in execution delays and process uncertainty. Similarly, uncoordinated spatial actions, such as lifting or placing components without confirming their position, represent violations of spatial coherence and can potentially escalate into safety-relevant events (e.g. S = 8). While the agents may function correctly in isolation, the absence of shared timing and spatial awareness mechanisms can negatively affect the process.

Communication-related process failures correspond to violations of information coherence. Ambiguous robot confirmation signals or missed human inputs are not technological malfunctions per se, but rather breakdowns in communicative alignment. When feedback is weak or non-salient, humans may misinterpret the system state, and this misalignment may remain undetected until a subsequent action fails (often reflected in higher detection scores). Regarding collaboration, situations in which one human moves a component without verifying their teammates' readiness indicate a loss of shared situational awareness and reflect a coherence violation rather than an isolated error. Hence, failures associated with collaboration extend beyond the exchange of information and encompass goals, cognition and authority. For example, role ambiguity can lead to situations in which neither person takes action because ownership of the task is unclear. In this case, the issue relates to cognitive coherence, since the agents do not share the same understanding of task ownership, and although the severity may be moderate, the absence of action can be difficult to detect promptly.

Overall, the MH-MR FMECA shows that multi-agent process failure modes are best understood as structured coherence violations across coordination, communication, collaboration, and related structural domains.

TABLE III. FMECA TABLE FOR MH-MR COLLABORATIVE ASSEMBLY OF THE SKATEBOARD

Assembly step	Failure Mode	Cause(s)	Effect(s)	Responsible	MH-MR macro-dimension	Violated coherence dimension	S	O	D
A.2/B.1/B.9	One human forgets to press input	Distraction or misunderstanding of readiness	Task halted; unnecessary delay	Human	Coordination	Temporal coherence	4	6	4
A.4/A.5/A.6/ A.7/A.8	Inputs not given in timely succession	Desynchronized press due to lack of prompt	Execution delay; human uncertainty	Joint	Coordination	Temporal coherence	4	6	6
C.1	Uncoordinated lifting or placement	No shared cue or tactile handoff	Board misalignment or drop	Human	Coordination	Goal coherence	8	4	8
B.9	Robot provides weak or ambiguous confirmation	Non-salient feedback channel	Human confused about process state	Robot	Communication	Information coherence	4	6	8
C.2	Humans interfere during precision tasks	Physical proximity; overlapping intent	Micro-delays or safety stops	Human	Team structure	Spatial coherence	6	6	6

A.2/B.1/B.9	Robots do not receive human input	Latency or wrong channel used	No action triggered	Human	Communication	Information coherence	4	4	8
A.3/B.2/B.5/ C.1	Robot enters safety mode due to collision	Humans enter robot space unexpectedly	Process interruption; requires restart or reset	Human	Team structure	Spatial coherence	8	4	6
C.2	Human moves board without checking teammate's status	Lack of mutual monitoring; absence of shared visual cues	Collision or misalignment	Human	Collaboration	Cognitive coherence	8	4	8
C.2	Neither human acts due to unclear responsibility	Role ambiguity; absence of explicit leadership in task	Task skipped or delayed; passive blocking	Human	Agent responsibilities	Cognitive coherence	6	4	10

VI. Discussion

The proposed coherence-based extension of FMECA complements previous work on failure analysis in dyadic human-robot assembly by shifting the focus from isolated task execution to interaction-level misalignments within MH-MR teams. While prior studies on HRC failures mainly address human errors, robot handling inaccuracies or interface-related issues within one-to-one collaboration [5], [23], the present analysis shows that additional failure modes emerge when several agents operate in parallel and share physical, informational and cognitive dependencies. This finding is consistent with teamwork-oriented perspectives on human-robot teams, which argue that performance evaluation should include coordination, communication and shared awareness dimensions beyond conventional productivity or safety indicators [8], [9]. From a process design perspective, the results suggest that MH-MR failure analysis should not only verify whether each agent can perform its assigned task, but also whether the team can preserve temporal, spatial, information, goal and cognitive coherence throughout the process. Accordingly, the proposed approach should be interpreted as a conceptual design-support tool rather than as a fully validated predictive model. Its main value lies in making multi-agent dependencies explicit and in helping designers anticipate failure modes that may remain hidden in conventional dyadic FMECA analyses.

VII. Conclusions

This study proposes an exploratory coherence-based methodology for analysing and evaluating process failure modes in multi-human, multi-robot collaborative assembly systems, by extending the traditional Failure Mode Effects and Criticality Analysis (FMECA). Applying the approach to a collaborative assembly scenario involving two human operators and two cobots shows its practical relevance. The analysis highlights that process failures in MH-MR systems are often not attributable to isolated agent limitations, but to breakdowns in shared coherence conditions within parallel and shared activities. By explicitly linking failure modes to MH-MR analytical macro-dimensions, the proposed methodology helps process designers to systematically anticipate multi-agent risks during the design of collaborative manufacturing systems. The main outputs produced by the methodology are the MH-MR characterisation structure, the taxonomy of coherence features and the extended FMECA schema linking failure modes to team-level dimensions and coherence violations.

However, the present work remains exploratory in nature. The identification, classification, and evaluation of failure modes are derived from a specific assembly scenario. Consequently, the resulting taxonomy should be adapted for different industrial contexts. The assigned severity, occurrence and detection ratings are qualitative and based on expert judgement rather than large-scale empirical data. Furthermore, the case study represents a controlled collaborative setting with limited variability in task structure, environmental constraints, and operator heterogeneity. Real industrial environments may introduce additional sources of uncertainty, such as varying levels of operator expertise, or unstructured workspace conditions. These factors could generate further classes of process failure mode that are not captured in the current analysis.

Therefore, future research should aim to validate and refine the proposed methodology by applying it to multiple industrial case studies and thus extending the taxonomy beyond the specific assembly process examined here. Particular attention should be given to collecting empirical performance data to support a more quantitative evaluation of S-O-D ratings. Overall, this work should be viewed as an initial step towards establishing a structured process failure analysis framework for MH-MR systems, providing a conceptual and methodological basis for further empirical development.

References

[1] P. K. R. Maddikunta *et al.*, "Industry 5.0: A survey on enabling technologies and potential applications," *J. Ind. Inf. Integr.*, vol. 26, p. 100257, Mar. 2022, doi: 10.1016/j.jii.2021.100257.

[2] European Commission. Directorate General for Research and Innovation., *Industry 5.0: towards a sustainable, human centric and resilient European industry*. LU: Publications Office, 2021. Accessed: Apr. 23, 2025. [Online]. Available: https://data.europa.eu/doi/10.2777/308407

[3] R. Gervasi, M. Capponi, L. Mastrogiacomo, and F. Franceschini, "Manual assembly and Human–Robot Collaboration in repetitive assembly processes: a structured comparison based on human-centered performances," *Int. J. Adv. Manuf. Technol.*, vol. 126, no. 3, pp. 1213–1231, May 2023, doi: 10.1007/s00170-023-11197-4.
[4] M. A. Goodrich and A. C. Schultz, "Human–Robot Interaction: A Survey," *Found. Trends® Human–Computer Interact.*, vol. 1, no. 3, pp. 203–275, Jan. 2008, doi: 10.1561/1100000005.
[5] D. A. Maisano, D. Antonelli, and F. Franceschini, "Assessment of Failures in Collaborative Human-Robot Assembly Workcells," *IFIP Adv. Inf. Commun. Technol.*, vol. 568, pp. 562–571, 2019, doi: 10.1007/978-3-030-28464-0_49.
[6] R. Gervasi, L. Mastrogiacomo, and F. Franceschini, "Human-robot collaboration in a repetitive assembly process: a preliminary investigation on operator's experience and product quality outputs," International Conference on Quality Engineering and Management, 2022, pp. 72–92. ISBN: 978-989-54911-1-7.
[7] M. Capponi, A. S. Di Matteo, and L. Mastrogiacomo, "Exploring multi-human multi-robot collaboration in assembly processes: Challenges and opportunities in quality and manufacturing," in *Materials Research Proceedings*, pp. 410–417, 2025, doi: 10.21741/9781644903735-48.
[8] L. M. Ma, M. Ijtsma, K. M. Feigh, and A. R. Pritchett, "Metrics for Human-Robot Team Design: A Teamwork Perspective on Evaluation of Human-Robot Teams," *ACM Trans. Hum.-Robot Interact.*, vol. 11, no. 3, p. 30:1-30:36, Settembre 2022, doi: 10.1145/3522581.
[9] D. Riedelbauch, N. Höllerich, and D. Henrich, "Benchmarking Teamwork of Humans and Cobots—An Overview of Metrics, Strategies, and Tasks," *IEEE Access*, vol. 11, pp. 43648–43674, 2023, doi: 10.1109/ACCESS.2023.3271602.
[10] G. Kokotinis, G. Michalos, Z. Arkouli, and S. Makris, "On the quantification of human-robot collaboration quality," *Int. J. Comput. Integr. Manuf.*, vol. 36, no. 10, pp. 1431–1448, Oct. 2023, doi: 10.1080/0951192X.2023.2189304.
[11] R. Gervasi, L. Mastrogiacomo, and F. Franceschini, "A conceptual framework to evaluate human-robot collaboration," *Int. J. Adv. Manuf. Technol.*, vol. 108, no. 3, pp. 841–865, May 2020, doi: 10.1007/s00170-020-05363-1.
[12] R. Jahanmahin, S. Masoud, J. Rickli, and A. Djuric, "Human-robot interactions in manufacturing: A survey of human behavior modeling," *Robot. Comput.-Integr. Manuf.*, vol. 78, p. 102404, Dec. 2022, doi: 10.1016/j.rcim.2022.102404.
[13] M. Dhanda, B. A. Rogers, S. Hall, E. Dekoninck, and V. Dhokia, "Reviewing human-robot collaboration in manufacturing: Opportunities and challenges in the context of industry 5.0," *Robot. Comput.-Integr. Manuf.*, vol. 93, p. 102937, Jun. 2025, doi: 10.1016/j.rcim.2024.102937.
[14] A. K. Inkulu, M. V. A. R. Bahubalendruni, A. Dara, and S. K., "Challenges and opportunities in human robot collaboration context of Industry 4.0 - a state of the art review," *Ind. Robot Int. J. Robot. Res. Appl.*, vol. 49, no. 2, pp. 226–239, Jan. 2021, doi: 10.1108/IR-04-2021-0077.
[15] C. Petzoldt, D. Niermann, E. Maack, M. Sontopski, B. Vur, and M. Freitag, "Implementation and Evaluation of Dynamic Task Allocation for Human–Robot Collaboration in Assembly," *Appl. Sci.*, vol. 12, no. 24, Art. no. 24, Jan. 2022, doi: 10.3390/app122412645.
[16] M. Calzavara, M. Faccio, I. Granata, and A. Trevisani, "Achieving productivity and operator well-being: a dynamic task allocation strategy for collaborative assembly systems in Industry 5.0," *Int. J. Adv. Manuf. Technol.*, Aug. 2024, doi: 10.1007/s00170-024-14302-3.
[17] T. Kiyokawa *et al.*, "Difficulty and complexity definitions for assembly task allocation and assignment in human–robot collaborations: A review," *Robot. Comput.-Integr. Manuf.*, vol. 84, p. 102598, Dec. 2023, doi: 10.1016/j.rcim.2023.102598.
[18] G. Hoffman, "Evaluating Fluency in Human–Robot Collaboration," *IEEE Trans. Hum.-Mach. Syst.*, vol. 49, no. 3, pp. 209–218, Jun. 2019, doi: 10.1109/THMS.2019.2904558.
[19] R. Gervasi, M. Capponi, L. Mastrogiacomo, and F. Franceschini, "Eye-tracking support for analyzing human factors in human-robot collaboration during repetitive long-duration assembly processes," *Prod. Eng.*, vol. 19, pp. 47–64, 2025, doi: 10.1007/s11740-024-01294-y.
[20] M. Capponi, R. Gervasi, L. Mastrogiacomo, and F. Franceschini, "Assembly complexity and physiological response in human-robot collaboration: Insights from a preliminary experimental analysis," *Robot. Comput.-Integr. Manuf.*, vol. 89, p. 102789, Oct. 2024, doi: 10.1016/j.rcim.2024.102789.
[21] E. Coronado, T. Kiyokawa, G. A. G. Ricardez, I. G. Ramirez-Alpizar, G. Venture, and N. Yamanobe, "Evaluating quality in human-robot interaction: A systematic search and classification of performance and human-centered factors, measures and metrics towards an industry 5.0," *J. Manuf. Syst.*, vol. 63, pp. 392–410, Apr. 2022, doi: 10.1016/j.jmsy.2022.04.007.
[22] E. Verna, S. Puttero, G. Genta, and M. Galetto, "A Novel Diagnostic Tool for Human-Centric Quality Monitoring in Human–Robot Collaboration Manufacturing," *J. Manuf. Sci. Eng.*, vol. 145, no. 121009, Sep. 2023, doi: 10.1115/1.4063284.
[23] D. Antonelli and D. Stadnicka, "Predicting and preventing mistakes in human-robot collaborative assembly," *IFAC-Pap.*, vol. 52, no. 13, pp. 743–748, Jan. 2019, doi: 10.1016/j.ifacol.2019.11.204.
[24] A. Haggy *et al.*, "Safety Assessment of Human-Robot Collaborations Using Failure Mode and Effects Analysis and Bow-Tie Analysis," presented at the 21st International Conference on Informatics in Control, Automation and Robotics, Feb. 2026, pp. 432–439. Accessed: Feb. 06, 2026. [Online]. Available: https://www.scitepress.org/Link.aspx?doi=10.5220/0013017500003822
[25] A. Nourmohammadi, M. Fathi, and A. H. C. Ng, "Balancing and scheduling assembly lines with human-robot collaboration tasks," *Comput. Oper. Res.*, vol. 140, p. 105674, Apr. 2022, doi: 10.1016/j.cor.2021.105674.

[26] M. Natarajan *et al.*, "Human-Robot Teaming: Grand Challenges," *Curr. Robot. Rep.*, vol. 4, no. 3, pp. 81–100, Sep. 2023, doi: 10.1007/s43154-023-00103-1.
[27] A. Dahiya, A. M. Aroyo, K. Dautenhahn, and S. L. Smith, "A Survey of Multi-Agent Human-Robot Interaction Systems," Dec. 10, 2022, *arXiv*: arXiv:2212.05286. doi: 10.48550/arXiv.2212.05286.
[28] G. Hoffman and C. Breazeal, "Collaboration in Human-Robot Teams," in *AIAA 1st Intelligent Systems Technical Conference*, Chicago, Illinois: American Institute of Aeronautics and Astronautics, Sep. 2004. doi: 10.2514/6.2004-6434.
[29] L. Mingyue Ma, T. Fong, M. J. Micire, Y. K. Kim, and K. Feigh, "Human-Robot Teaming: Concepts and Components for Design," in *Field and Service Robotics*, M. Hutter and R. Siegwart, Eds., Cham: Springer International Publishing, 2018, pp. 649–663. doi: 10.1007/978-3-319-67361-5_42.
[30] J. MacMillan, E. E. Entin, and D. Serfaty, "Communication overhead: The hidden cost of team cognition," in *Team cognition: Understanding the factors that drive process and performance*, Washington, DC, US: American Psychological Association, 2004, pp. 61–82. doi: 10.1037/10690-004.
[31] D. H. Stamatis, *Failure Mode and Effect Analysis: Fmea from Theory to Execution*. ASQ Quality Press, 2003. ISBN: 978-0-87389-598-9
[32] D. A. Maisano, D. Antonelli, and F. Franceschini, "Assessment of Failures in Collaborative Human-Robot Assembly Workcells," presented at the 20th Working Conference on Virtual Enterprises (PRO-VE), Springer International Publishing, Sep. 2019, p. 562. doi: 10.1007/978-3-030-28464-0_49.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Understanding Assembly Process Quality through Product Complexity and Cognitive Demand: Preliminary Experimental Evidence

Riccardo Gervasi [1], Laura Angioletti [2], Michela Balconi [3], Luca Mastrogiacomo [4], and Fiorenzo Franceschini [5]

1 Department of Management and Production Engineering (DIGEP), Politecnico di Torino; riccardo.gervasi@polito.it
2 International research center for Cognitive Applied Neuroscience (IrcCAN), Università Cattolica del Sacro Cuore di Milano; laura.angioletti1@unicatt.it
3 International research center for Cognitive Applied Neuroscience (IrcCAN), Università Cattolica del Sacro Cuore di Milano; michela.balconi@unicatt.it
4 Department of Management and Production Engineering (DIGEP), Politecnico di Torino; luca.mastrogiacomo@polito.it
5 Department of Management and Production Engineering (DIGEP), Politecnico di Torino; fiorenzo.franceschini@polito.it

***Abstract*— This paper investigates how psychophysiological and attentional demands vary across the different phases of manual assembly processes with different levels of product complexity. Within the Industry 5.0 perspective, understanding when operators experience higher mental demand is essential for improving assembly process quality in terms of design, instructions, training, and human-centered support systems to avoid errors and defects. An experimental study was conducted on two assembly processes: a mechanical equipment and a tile cutter, representing lower and higher complexity products, respectively. Participants performed the assemblies while multimodal data were collected through eye-tracking, electrodermal activity (EDA) biosensor, and functional near-infrared spectroscopy (fNIRS). Each assembly process was decomposed into macro-phases, and psychophysiological metrics were analyzed separately for each one. Results showed that the simpler product exhibited limited differences among macro-phases. In contrast, the more complex product revealed clearer heterogeneity, with some phases emerging as attentional and cognitive bottlenecks, particularly through eye-tracking measures. These findings confirmed that assembly complexity is not uniformly distributed across process steps and demonstrate the value of a multimodal, phase-based analysis for identifying critical moments where targeted support interventions, such as improved instructions, cognitive assistance, or collaborative technologies, may be most beneficial for process quality improvement.**

Keywords – Industry 5.0, Cognitive ergonomics, Assembly process quality, Quality engineering.

I. Introduction

In assembly processes, operators must optimally allocate their attention, cognitive resources, and motor control to achieve a specific objective while adhering to constraints related to sequence, precision, time, and quality (Calzavara *et al.*, 2023; Gervasi *et al.*, 2023; Vitti *et al.*, 2025). Within the context of Industry 5.0, analyzing the involvement of mental resources becomes of fundamental importance for improving process quality (Bartolomei *et al.*, 2026). Understanding which steps in the process are most costly from a psychophysiological perspective is crucial for reducing errors, improving product and instruction design, optimizing training, and, more generally, increasing safety and efficiency (Gervasi *et al.*, 2024; Gualtieri *et al.*, 2022; Hollnagel, 1997). However, what makes an operation challenging is not always visible from the outside: two phases may appear similar in terms of manual actions, but differ markedly in terms of mental load, arousal, attentional demands, or decision uncertainty (Calzavara *et al.*, 2024; Wang and Jiao, 2025). For this reason, in recent years there has been growing interest in multimodal approaches that combine psychophysiological and behavioral signals, capable of capturing subtle and dynamic changes during task performance (Capponi *et al.*, 2024a; Villani *et al.*, 2020).

The objective of this paper is to explore how psychophysiological and attentional demands vary within assembly processes of products with different product complexity, in order to better understand where and when support interventions may be needed to improve process quality from a cognitive perspective. In the proposed study, assembly processes of two products of different complexity were observed through three families of psychophysiological measures: (i) Electrodermal Activity (EDA), a sensitive indicator of sympathetic nervous system activation and therefore of arousal and activation; (ii) functional Near-Infrared Spectroscopy (fNIRS), with estimation of aggregate oxyhemoglobin (Oxy-Hb) and deoxyhemoglobin (Deoxy-Hb) in frontal, fronto-central, and temporo-central brain regions, to capture variations in cortical activation associated with executive control, planning, and monitoring; (iii) Eye-tracking, using oculomotor and pupillary metrics (i.e., fixation duration, number of fixations per second, pupillary dilation, peak velocity, and saccade amplitude), to describe the dynamics of visual attention and, indirectly, aspects related to visual search, cognitive load, and exploration strategies.

The most interesting methodological aspect is that assembly is not treated as a single event, but as a sequence of macro-phases. In particular, the simplest product (i.e., a mechanical equipment) was segmented into three macro-phases, while the most complex product (i.e., a tile cutter) was segmented into six macro-phases. Psychophysiological metrics were calculated separately for each macro-phase, enabling a fine-grained analysis aimed at identifying when activation peaks, changes in attention, and signs of increased mental demand emerged.

This approach aims to highlight how operational complexity can be not uniform: in an assembly process, there are typically preparation and orientation steps, repetitive assembly phases, and critical moments that require alignment, verification, decisions, and quality control. Analyzing the psychophysiological signals within the assembly macro-phases therefore allows to

This study was carried out within the «Empathically enhanced robot for the collaboration with HUmans in MANufacturing (E.HU.MAN)» project – funded by the Ministero dell'Università e della Ricerca – within the PRIN 2022 program (D.D.104–02/02/2022). This manuscript reflects only the authors' views and opinions, and the Ministry cannot be considered responsible for them.

go beyond an overall average difference and produce more useful results for targeted interventions (e.g., improving instructions, redesigning components, inserting checks or supports in the most critical steps) to improve process quality.

The paper is structured as follows. In Section II, the methodology implemented in the conducted for an exploratory case study is described. Section III presents the results obtained. Finally, discussion, conclusions, and directions for future research are provided in Section IV.

II. METHODOLOGY

The present study consisted in the assembly of two products of different product complexity: a mechanical equipment (Fig. 1) and a tile cutter (Fig. 2). Each assembly process was divided into macro-phases according to the following criterions: (i) each macro-phase had to represent the complete joining of two types of components (i.e., involving elementary tasks of identification, handling, alignment, and joining), thus engaging the same executive and cognitive functions (i.e., planning, working memory, monitoring, inhibitory control, response regulation, sustained attention, visuo-motor coordination) (Balconi *et al.*, 2023; Diamond, 2013), albeit with varying degrees of intensity; (ii) each macro-phase had to last at least 15 seconds to ensure a sufficient amount of data from slower psychophysiological signals, such as fNIRS and EDA. The mechanical equipment assembly process was decomposed into the following macro-phases:

- MMP1: positioning of Base, and assembly of the first oval flange (EF1) using two screws B1 and two nuts N1.
- MMP2: assembly of the square flange (SF) using two screws B1 and two nuts N1.
- MMP3: assembly of the oval flange (EF2) using the two screws B2 and two nuts N1, and final tightening check of all bolts.

The tile cutter was decomposed into the following macro-phases:

- TMP1: positioning of Base, and assembly of the two side supports (C1a and C1b) using a screw B1 and a nut N1 for each.
- TMP2: assembly of the first component of the cutting mechanism (C2) with the component C3 using the screw B2 and the nut N2.
- TMP3: assembly of the round blade (L1) with the component C3 using a screw B3 and a nut N3.
- TMP4: assembly of the component C4 with component C3 using a screw B3 and a nut N3.
- TMP5: inserting the two rail rods (P1a and P1b) into the holes of component C2 and assembling them to the side supports C1a and C1b.
- TMP6: assembly of the handle with component C3.

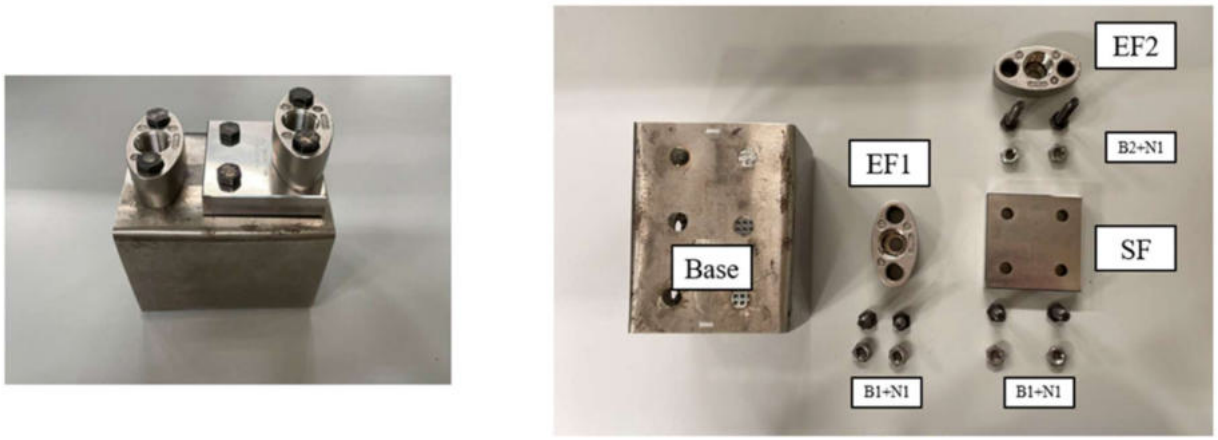


Fig. 1. The mechanical equipment and its components.

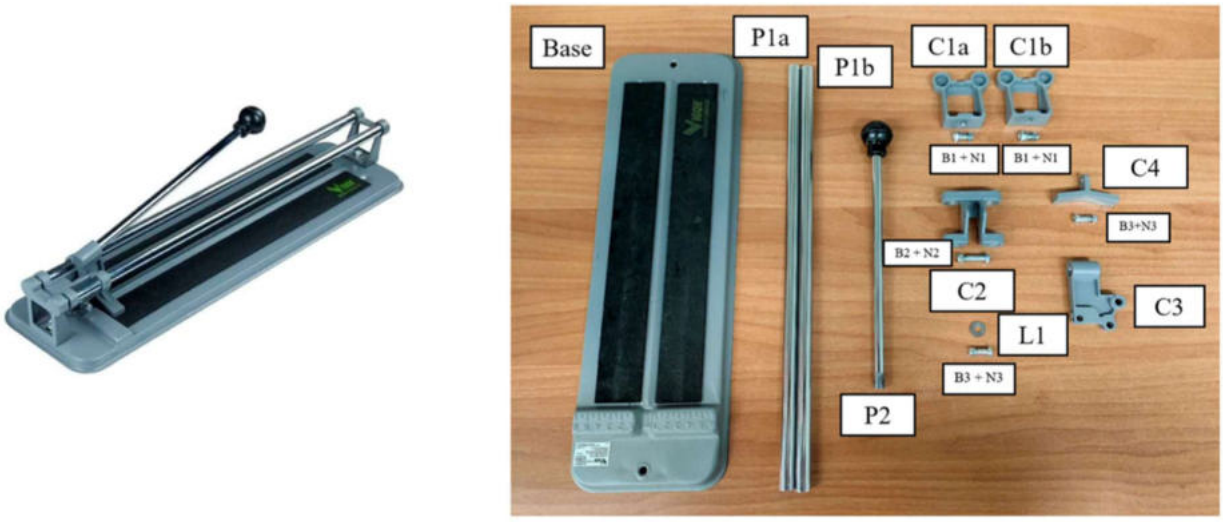


Fig. 2. The tile cutter and its components.

A. Physiological signals and behavioral data

The Shimmer3R GSR+ biosensor (Fig. 3a) was used to record EDA by placing the two electrodes on the first phalanges of the non-dominant hand according to standard procedures to minimize artifacts. Tobii Pro Glasses 3 device (Fig. 3c) was used to collect eye-tracking data (i.e., fixations, saccades, and pupil dilation) during the assembly process. Cortical hemodynamic activity was recorded using the fNIRS NIRScout system (Fig. 3b) with a 24-channel array of optodes (8 light emitters and 16 detectors) (Fig. 4) with a sampling rate of 6.25 Hz (Balconi and Vanutelli, 2016). The emitters were placed on positions AFp1, AFp2, CFC3, CFC4, CCP5, CCP6, PCP7, and PCP8. The detectors were placed on positions FAF5, FAF6, FAF1, FAF2, FFC3, FFC4, CFC5, CFC6, CCP3, CCP4, CCP7, CCP8, PCP5, PCP6, PCP9, and PCP10. The emitter-detector distance was 30 mm for contiguous optodes, and near-infrared light of two wavelengths (i.e., 760 and 850 nm) was used. Positioning was performed with respect to cranial references of the 10-5 system (Oostenveld and Praamstra, 2001) to cover the frontal (F) (channels 1, 2, 13, and 14), fronto-central (FC) (channels 3, 4, 5, 6, 7, 15, 16, 17, 18, and 19), and temporo-central (TC) (channels 8, 9, 10, 11, 12, 13, 20, 21, 22, 23, and 24) regions (Fig. 4). The relative concentrations of Oxy-Hb and Deoxy-Hb were derived.

Fig. 3. The implemented biosensor: (a) Shimmer3R GSR+ (Shimmer, 2026), (b) NIRScout system (NIRx, 2026), (c) Tobii Pro Glasses 3 (Tobii, 2026).

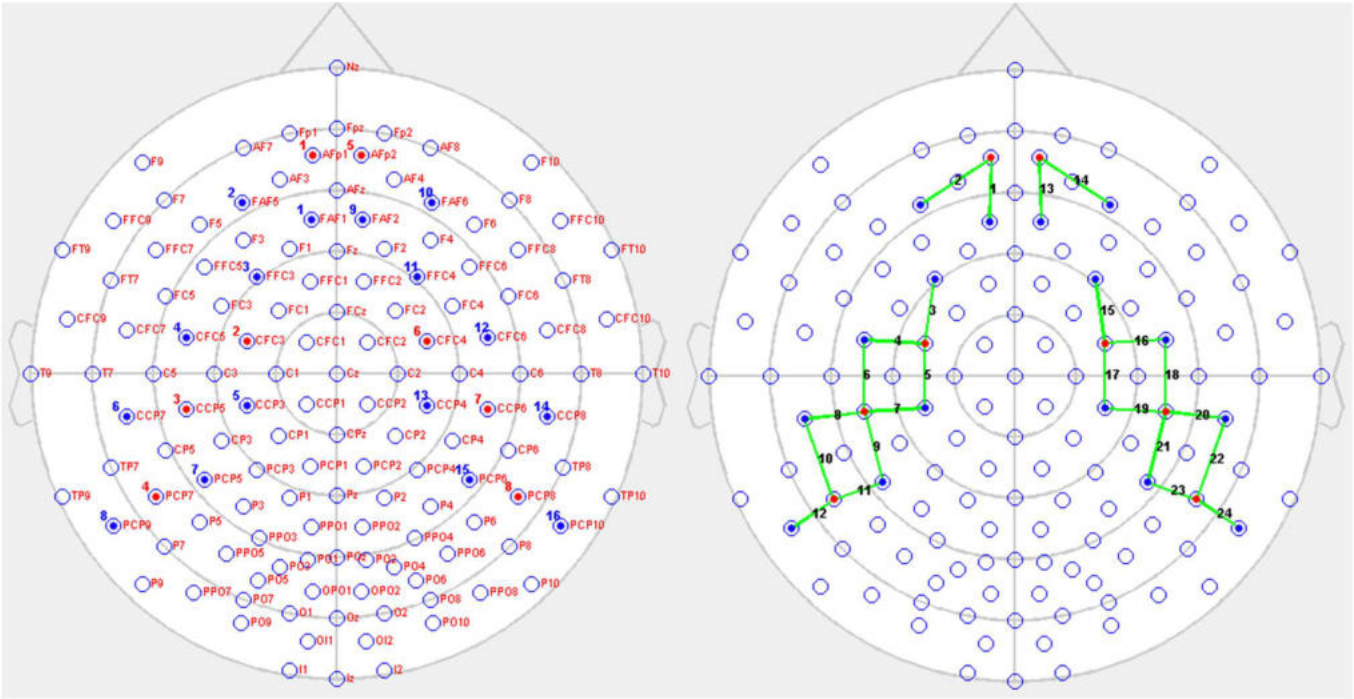


Fig. 4. On the left, position of fNIRS emitters (in red) and detectors (in blue) according to the international 10-5 system; on the right, position of the 24 channels.

B. Experimental procedure

The study was carried out in the in the "Mind4Lab" (Manufacturing Industry 4.0 Laboratory) at the "Politecnico di Torino" involving 20 participants (10 males and 10 females aged between 20 and 28 years old) without any previous experience in assembling the products.

Initially, the objectives of the study and its procedure were explained to the participant and their demographics were collected. The researcher subsequently seated the participant in the experiment location and explained the various steps of the assembly process. After any questions were discussed, the biosensors were set up on the participant. Next, the participant was invited to relax and remain still to record 2 minutes of physiological signals at rest (i.e., the baseline of the physiological signals). After this phase was completed, the participant started the assembly phase. For each product, the participant performed a training trial to learn the assembly process and then, if everything was clear, performed the assembly without any external support. Each participant performed the assembly of each product in random order. At the end of each trial, feedback from the participant was collected. The entire experimental session lasted 60 minutes on average.

C. Data processing

The software "Tobii Pro Lab" was used to elaborate eye-tracking data and extract eye-tracking metrics. Five of the most common metrics were analyzed: average duration of fixations, number of fixations per second, average pupil diameter, average peak velocity of saccades, and average amplitude of saccades. The "nirsLAB" software was used to analyzed fNIRS data and derive average changes in Oxy-Hb and Deoxy-Hb for each region. EDA data were processed and cleaned from artifacts (e.g., unusual peaks, extreme values) and average Skin Conductance Level (SCL) was computed. A summary of the response variables is reported in Table I.

TABLE I. SUMMARY OF THE INDEPENDENT AND DEPENDENT VARIABLES.

Type of variable	Variable	Description
Independent variables	Product	Product to be assembled (Mechanical equipment, Tile-cutter)
	Macro-phase	Macro-phase of the product (Mechanical equipment: MMP1, MMP2, and MMP3; Tile cutter: TMP1, TMP2, TMP3, TMP4, TMP5, and TMP6)
Dependent variables	Avg. duration of fixations	The average duration of fixations [ms]
	Nr. of fixations per second	The number of fixations per second [1/s]
	Avg. pupil diameter	The average pupil dilation [ms]
	Avg. peak velocity of saccades	The average peak velocity of saccades [°/s]
	Avg. amplitude of saccades	The average amplitude of saccades [°]
	Oxy-Hb	The average changes in oxyhemoglobin [mmol∗mm]
	Deoxy-Hb	The average changes in deoxyhemoglobin [mmol∗mm]
	SCL	Average skin conductance level [μS]

III. RESULTS

In this section, the comparison between the macro-phases for each product is analyzed based on the response variables collected. Additionally, significant differences in the response variables between macro-phases are highlighted through the Wilcoxon signed-rank test (Wilcoxon, 1945), in order to correctly manage paired data. To control family-wise error rate due to multiple pairwise comparisons, the Holm method (Holm, 1979) was implemented to obtain the adjusted *p*-values.

A. Mechanical equipment

Table II contains the resulting *p*-values of the pairwise Wilcoxon signed-rand tests among the three macro-phases for each response variable. Fig. 5 shows the comparison of the eye-tracking metrics and SCL between the three macro-phases of the mechanical equipment. Looking at the eye-tracking metrics, there were no significant differences between the macro-phases except for the amplitude of the saccades. The amplitude of the saccades significantly distinguished MMP1 from the subsequent macro-phases (i.e., MMP2 and MMP3). MMP1 included an initial setup phase, in which operators performed more visual searches by looking at and identifying among multiple components, even similar ones (i.e., EF1 and EF2), which gradually decreased in the subsequent macro-phases. This tended to produce larger saccades than in the subsequent macro-phases, where attention can remain more focused. The SCL showed significant differences between all macro-phases, highlighting a decreasing trend. This result highlights a gradual decrease in mental effort on the part of the operator due to both a decreasing variety of components to choose from and a repetitiveness of operations as the macro-phases as the macro-phases follow one another.

TABLE II. PAIRWISE WILCOXON SIGNED-RANK TEST *P*-VALUES ADJUSTED WITH HOLM CORRECTION FOR MECHANICAL EQUIPMENT MACRO-PHASES.

Pairwise comparison	*p*-value											
	Avg. duration of fixations	Nr. of fixations per second	Avg. pupil diameter	Avg. peak velocity of saccades	Avg. amplitude of saccades	SCL	Oxy-Hb (F)	Oxy-Hb (FC)	Oxy-Hb (TC)	Deoxy-Hb (F)	Deoxy-Hb (FC)	Deoxy-Hb (TC)
MMP1-MMP2	0.792	0.846	0.696	0.958	0.020*	<0.001***	0.579	0.504	1.000	1.000	1.000	1.000
MMP1-MMP3	0.854	0.202	0.585	0.327	0.001**	<0.001***	0.579	0.046*	1.000	1.000	1.000	1.000
MMP2-MMP3	0.854	0.165	0.696	0.592	0.208	<0.001***	0.579	0.669	1.000	1.000	1.000	1.000

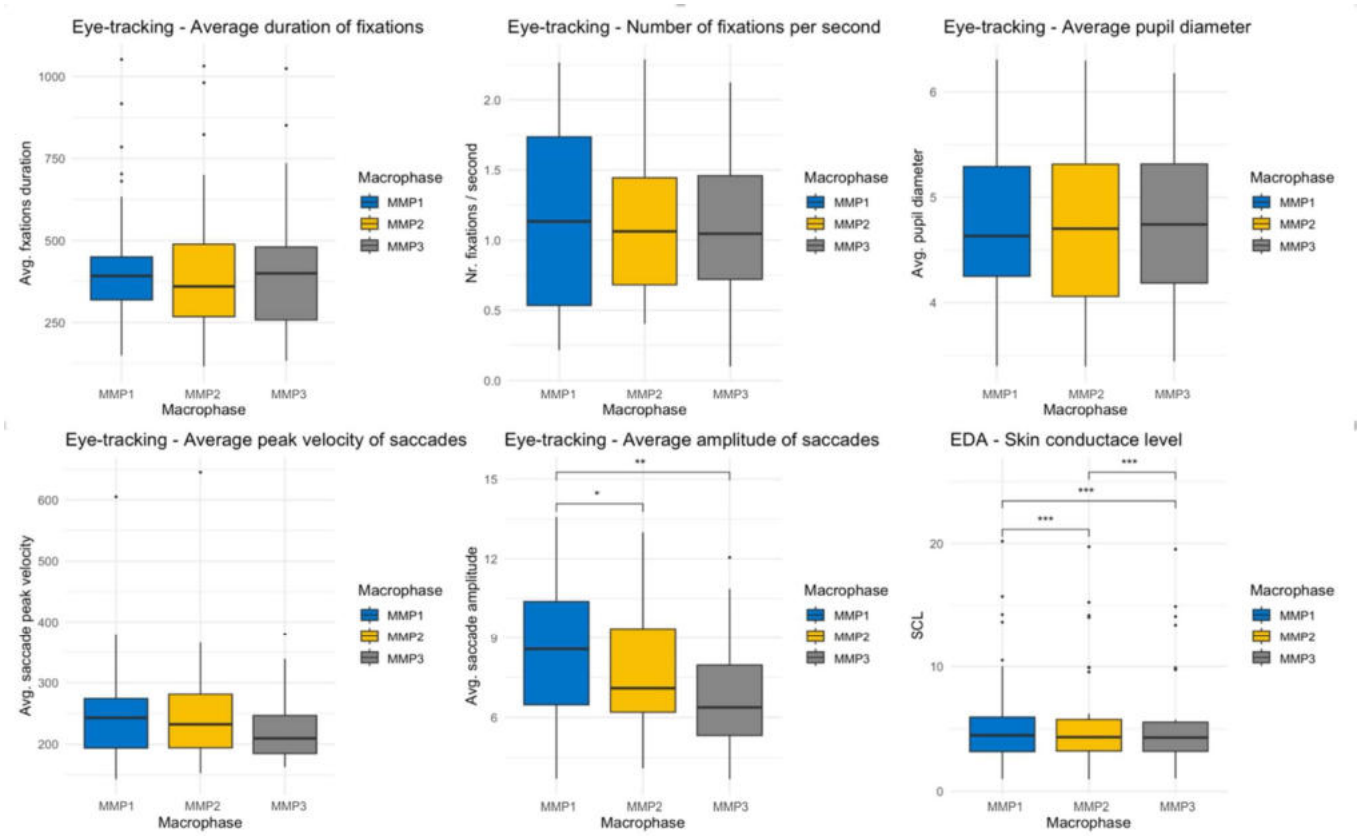


Fig. 5. Boxplot comparison of the eye-tracking metrics and EDA between the macro-phases of mechanical equipment ('*' $p<0.05$; '**' $p<0.01$; '***' $p<0.001$).

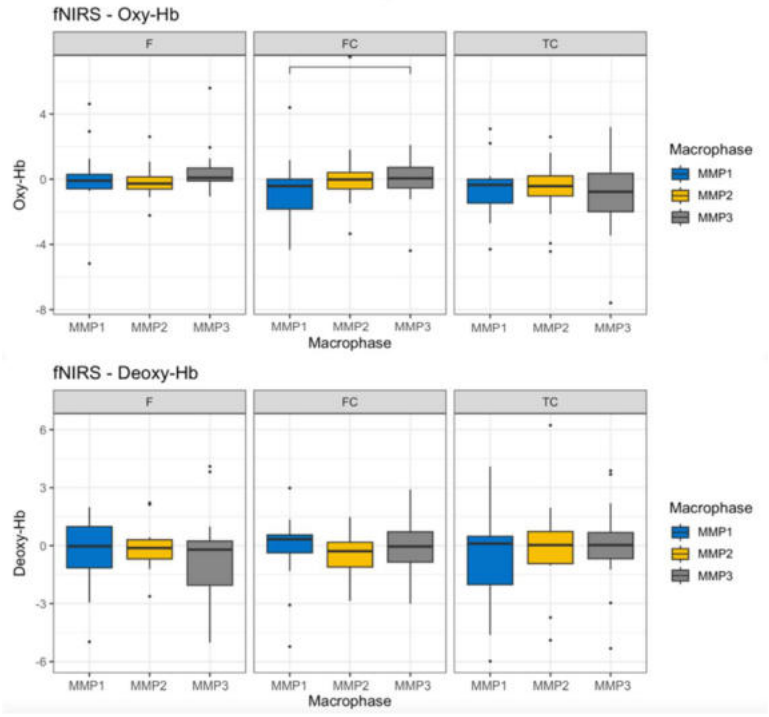


Fig. 6. Boxplot comparison of the average changes in Oxy-Hb and Deoxy-Hb between the macro-phases of mechanical equipment ('*' $p<0.05$; '**' $p<0.01$; '***' $p<0.001$).

Fig. 6 shows the comparison of the fNIRS metrics (i.e., Oxy-Hb and Deoxy-Hb) between the three macro-phases of the mechanical equipment. A slightly significant difference (p=0.046) emerged in the FC Oxy-Hb between the initial phase (MMP1) and the final phase (MMP3). The increase in Oxy-Hb in the fronto-central region was consistent with a greater demand for executive control in the final phase for the final tightening of all screws. However, it should be noted that there was no extensive evidence of this pattern in other ROIs, and FC Deoxy-Hb did not show a significant decrease between MMP1 and MMP3.

B. Tile cutter

Table III contains the resulting p-values of the pairwise Wilcoxon signed-rand tests among the six macro-phases for each response variable. Fig. 7 shows the comparison of the eye-tracking metrics and SCL between the six macro-phases of the tile cutter. Looking at the metrics related to saccade behavior, TMP2 emerged as significantly different from all other macro-phases. Specifically, the average amplitude and average peak velocity of saccades were lower in TMP2, highlighting greater cognitive effort compared to the other macro-phases, due to greater visual attention focused on a rather restricted area during the assembly of component C2 with C3.

Observing pupil diameter, which can be interpreted as an indicator of cognitive load and arousal, higher values are observed in TMP2 and TMP5. In particular, significant differences emerge between TMP2 and TMP1 (p=0.015), TMP5 and TMP1 (p=0.002), and TMP5 and TMP6 (p=0.045).

Focusing on the number of fixations per second, TMP5 (i.e., insertion of the two rail rods in C2 and fixing them to the supports) and TMP6 (i.e., assembly of the handle with C3) higher values than the previous macro-phases were observed, highlighting a more dynamic visual attention behavior. TMP5 differed significantly from TMP3 (p=0.016), suggesting that inserting the rods was not only more cognitively demanding (i.e., greater pupil dilation) but also more visually fragmented (i.e., more fixations per second). TMP6 differed significantly from TMP1 (p=0.017), TMP3 (p=0.016), and TMP4 (p=0.045), highlighting the presence of rapid shifts in gaze between the fastening point and the overall configuration for a final check.

No significant differences emerged between the macro-phases in terms of SCL, Oxy-Hb and Deoxy-Hb (Fig. 8), suggesting a fairly stable cortical activation throughout the process.

TABLE III. PAIRWISE WILCOXON SIGNED-RANK TEST *P*-VALUES ADJUSTED WITH HOLM CORRECTION FOR TILE CUTTER MACRO-PHASES.

Pairwise comparison	*p*-value											
	Avg. duration of fixations	Nr. of fixations per second	Avg. pupil diameter	Avg. peak velocity of saccades	Avg. amplitude of saccades	SCL	Oxy-Hb (F)	Oxy-Hb (FC)	Oxy-Hb (TC)	Deoxy-Hb (F)	Deoxy-Hb (FC)	Deoxy-Hb (TC)
TMP1-TMP2	1.000	0.942	0.015*	<0.001***	<0.001***	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP1-TMP3	1.000	1.000	1.000	1.000	0.253	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP1-TMP4	0.833	1.000	0.150	0.404	0.074	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP1-TMP5	0.197	0.127	0.002**	0.264	0.052	1.000	1.000	1.000	1.000	1.000	0.866	1.000
TMP1-TMP6	0.056	0.017*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP2-TMP3	1.000	0.161	0.238	0.038*	0.014*	1.000	0.710	1.000	1.000	1.000	1.000	1.000
TMP2-TMP4	1.000	0.92	1.000	0.035*	0.016*	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP2-TMP5	0.648	0.942	1.000	0.029*	0.004**	1.000	1.000	0.866	1.000	1.000	0.916	1.000
TMP2-TMP6	0.062	0.920	1.000	0.010**	0.002**	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP3-TMP4	0.196	1.000	0.150	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP3-TMP5	0.249	0.016*	0.073	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP3-TMP6	0.120	0.016*	1.000	1.000	0.258	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP4-TMP5	1.000	0.097	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP4-TMP6	0.754	0.045*	1.000	1.000	0.253	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TMP5-TMP6	1.000	1.000	0.045*	0.268	0.088	1.000	1.000	1.000	0.664	1.000	1.000	1.000

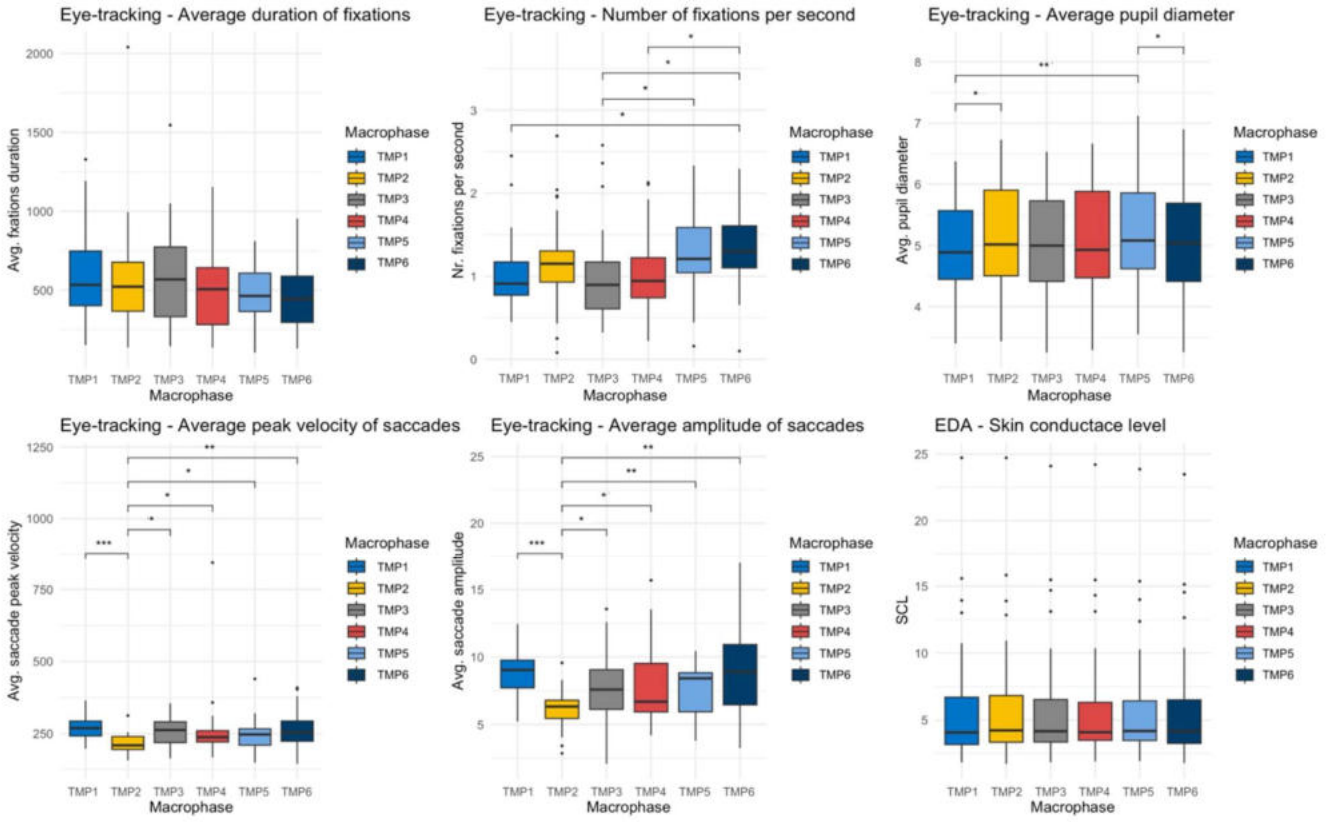


Fig. 7. Boxplot comparison of the eye-tracking metrics and EDA between the macro-phases of tile cutter ('*' p<0.05; '**' p<0.01; '***' p<0.001).

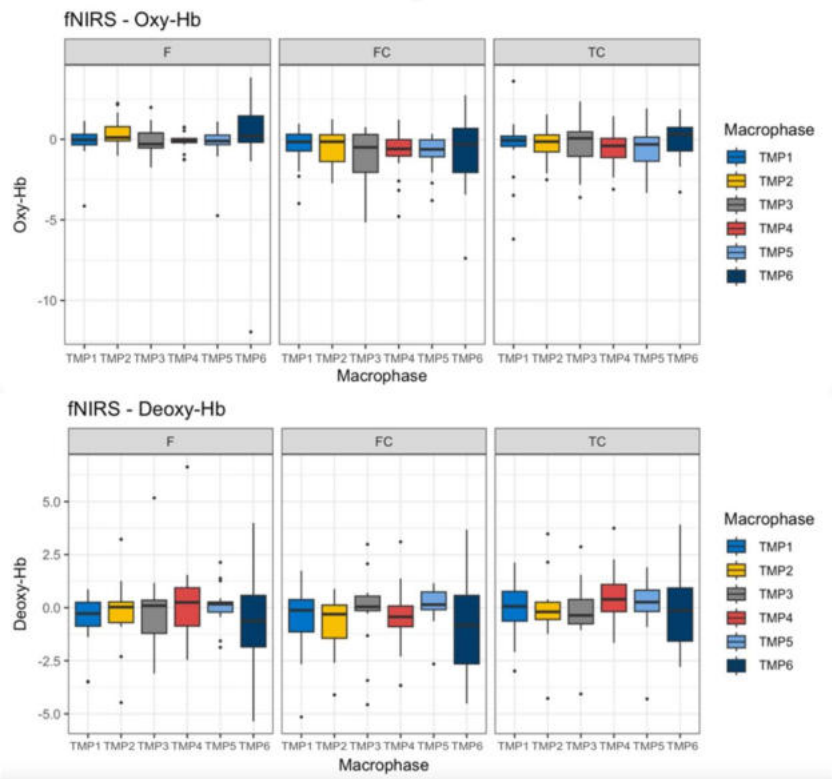


Fig. 8. Boxplot comparison of the average changes in Oxy-Hb and Deoxy-Hb for each ROI between the macro-phases of tile cutter ('*' p<0.05; '**' p<0.01; '***' p<0.001).

IV. Discussion and conclusions

The aim of this study was to analyze process quality from a cognitive perspective, by characterizing how psychophysiological and attentional demands vary within assembly processes, using a multimodal approach (i.e., eye tracking, fNIRS, and EDA) and comparing macro-phases within each product. The results proved to be different depending on the complexity of the product: in the simplest task (i.e., mechanical equipment), the differences between macro-phases are limited and circumscribed, while in the most complex task (i.e., tile cutter), clearly distinctive macro-phases emerge, suggesting that the complexity of the product translates into greater heterogeneity of demands throughout the process.

In the mechanical equipment assembly process, the only eye-tracking metric sensitive to differences between macro-phases was the amplitude of saccades, with MMP1 distinct from MMP2 and MMP3. From an operational point of view, MMP1 included positioning the base and starting assembly with the first oval flange, a phase in which it is plausible that operators must quickly construct a spatial representation of the work area and identify the correct components and orientations. This type of "attentive setup" is compatible with greater visual exploration and therefore with larger saccadic patterns compared to later macro-phases, which are more repetitive and focused on tightening already known elements. This effect was also observed through SCL, with gradual relaxation due to the receptiveness of the operations. At the cortical level, the difference between MMP1 and MMP3 observed on fronto-central Oxy-Hb suggests that the beginning and final phases were not equivalent in terms of executive control processes. MMP3 involved assembling the second oval flange with B2 screws (different from B1) and required greater control of the final assembly conditions (i.e., alignment and correct overall tightening). It is relevant that the fNIRS effect was selective to fronto-central region, while other indices did not differ. This may indicate that in an assembly process of a simple product the demands change more in terms of executive control in some specific steps than in terms of arousal or global visual strategies.

In the tile cutter assembly process, eye-tracking metrics revealed the heterogeneity of the macro-phases due to an uneven cognitive resource use across them. A key finding was that TMP2 (i.e., the assembly of C2 with C3) represented an attentional and cognitive bottleneck. TMP2 systematically differed from all other macro-phases in terms of saccadic metrics (i.e., amplitude and peak velocity). The pattern found was consistent with a phase of fine manipulation and concentration in a restricted area, indicating the need for greater cognitive effort to complete the macro-phase. At the same time, TMP2 showed also greater pupil diameter, suggesting that it required greater sustained attention. Participants' feedback confirmed that the alignment between the component C2 and C3 was the most difficult phase to carry out correctly, as it was not intuitive, thus requiring mental effort. TMP5 (i.e., the insertion and fixing of the rail rods) emerged as another macro-phase requiring high attention, presenting greater pupil dilation and number of fixations per second. This macro-phase required several visual checks for the correct insertion of the rods into the C2 supports and for the alignment of the rods in the C1a and C1b side supports, which contributed to increasing the cognitive effort. Surprisingly, EDA and fNIRS metrics did not reveal any statistically significant differences among the macro-phases. This may be due to the variety of component shapes and the low repetitiveness of assembly operations between macro-phases, which required a constant level of executive control involvement.

The overall pattern that emerged is consistent with the idea that in a complex task certain macro-phases represent attentional and cognitive bottlenecks (Capponi *et al.*, 2024b). From an application perspective, the paper showed how to objectively identify these situations, suggesting where support interventions (e.g., cobots, cognitive assistance systems, targeted training) should be concentrated to improve process quality and thus prevent errors and defects. For instance, concerning the tile cutter assembly process, additional cognitive support should be implemented in TMP2 and TMP5 to reduce mental effort and reduce error risk.

Future work will focus on validating the preliminary results by enlarging the experiment to include a higher number of participants and other products to be assembled, as well as characterizing the effect of different kind of cognitive support interventions in an assembly process.

References

Balconi, M., Acconito, C., Allegretta, R.A. and Crivelli, D. (2023), "What Is the Relationship between Metacognition and Mental Effort in Executive Functions? The Contribution of Neurophysiology", *Behavioral Sciences*, Multidisciplinary Digital Publishing Institute, Vol. 13 No. 11, p. 918, doi: 10.3390/bs13110918.

Balconi, M. and Vanutelli, M.E. (2016), "Hemodynamic (fNIRS) and EEG (N200) correlates of emotional inter-species interactions modulated by visual and auditory stimulation", *Scientific Reports*, Nature Publishing Group, Vol. 6 No. 1, p. 23083, doi: 10.1038/srep23083.

Bartolomei, M., Gervasi, R., Acconito, C., Angioletti, L., Cannizzaro, D., Balconi, M., Mastrogiacomo, L., *et al.* (2026), "Evaluating mental workload: A taxonomic approach to evaluation tools based on ISO 10075", *Applied Ergonomics*, Vol. 130, p. 104659, doi: 10.1016/j.apergo.2025.104659.

Calzavara, M., Faccio, M. and Granata, I. (2023), "Multi-objective task allocation for collaborative robot systems with an Industry 5.0 human-centered perspective", *The International Journal of Advanced Manufacturing Technology*, Vol. 128 No. 1, pp. 297–314, doi: 10.1007/s00170-023-11673-x.

Calzavara, M., Faccio, M., Granata, I. and Trevisani, A. (2024), "Achieving productivity and operator well-being: a dynamic task allocation strategy for collaborative assembly systems in Industry 5.0", *The International Journal of Advanced Manufacturing Technology*, Vol. 134 No. 7, pp. 3201–3216, doi: 10.1007/s00170-024-14302-3.

Capponi, M., Gervasi, R., Mastrogiacomo, L. and Franceschini, F. (2024a), "Assembly complexity and physiological response in human-robot collaboration: Insights from a preliminary experimental analysis", *Robotics and Computer-Integrated Manufacturing*, Vol. 89, p. 102789, doi: 10.1016/j.rcim.2024.102789.

Capponi, M., Gervasi, R., Mastrogiacomo, L. and Franceschini, F. (2024b), "Assessing perceived assembly complexity in human-robot collaboration processes: a proposal based on Thurstone's law of comparative judgement", *International Journal of Production Research*, Taylor & Francis, Vol. 62 No. 14, pp. 5315–5335, doi: 10.1080/00207543.2023.2291519.

Diamond, A. (2013), "Executive Functions", *Annual Review of Psychology*, Annual Reviews, Vol. 64 No. Volume 64, 2013, pp. 135–168, doi: 10.1146/annurev-psych-113011-143750.

Gervasi, R., Capponi, M., Mastrogiacomo, L. and Franceschini, F. (2023), "Manual assembly and Human–Robot Collaboration in repetitive assembly processes: a structured comparison based on human-centered performances", *The International Journal of Advanced Manufacturing Technology*, Vol. 126 No. 3, pp. 1213–1231, doi: 10.1007/s00170-023-11197-4.

Gervasi, R., Capponi, M., Mastrogiacomo, L. and Franceschini, F. (2024), "Eye-tracking support for analyzing human factors in human-robot collaboration during repetitive long-duration assembly processes", *Production Engineering*, doi: 10.1007/s11740-024-01294-y.

Gualtieri, L., Fraboni, F., De Marchi, M. and Rauch, E. (2022), "Development and evaluation of design guidelines for cognitive ergonomics in human-robot collaborative assembly systems", *Applied Ergonomics*, Vol. 104, p. 103807, doi: 10.1016/j.apergo.2022.103807.

Hollnagel, E. (1997), "Cognitive ergonomics: it's all in the mind", *Ergonomics*, Taylor & Francis, Vol. 40 No. 10, pp. 1170–1182, doi: 10.1080/001401397187685.

Holm, S. (1979), "A Simple Sequentially Rejective Multiple Test Procedure", *Scandinavian Journal of Statistics*, [Board of the Foundation of the Scandinavian Journal of Statistics, Wiley], Vol. 6 No. 2, pp. 65–70.

NIRx. (2026), "NIRScout", available at: https://nirx.net/nirscout (accessed 20 January 2026).

Oostenveld, R. and Praamstra, P. (2001), "The five percent electrode system for high-resolution EEG and ERP measurements", *Clinical Neurophysiology*, Vol. 112 No. 4, pp. 713–719, doi: 10.1016/S1388-2457(00)00527-7.

Shimmer. (2026), "Shimmer3R GSR+ Unit", available at: https://www.shimmersensing.com/product/shimmer3-gsr-unit/ (accessed 20 January 2026).

Tobii. (2026), "Tobii Pro Glasses 3", available at: https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3 (accessed 20 January 2026).

Villani, V., Righi, M., Sabattini, L. and Secchi, C. (2020), "Wearable devices for the assessment of cognitive effort for human-robot interaction", *IEEE Sensors Journal*, presented at the IEEE Sensors Journal, pp. 1–1, doi: 10.1109/JSEN.2020.3001635.

Vitti, M., Cotruvo, A. and Facchini, F. (2025), "Toward human-centric Industry 5.0: a framework for cognitive workload assessment in maintenance and assembly", *Journal of Intelligent Manufacturing*, doi: 10.1007/s10845-025-02716-z.

Wang, S. and Jiao, R.J. (2025), "Cognitive intelligent task allocation for human-automation symbiosis in Industry 5.0 manufacturing systems via non-cooperative game theory: a bi-level optimization approach", *The International Journal of Advanced Manufacturing Technology*, Vol. 136 No. 3, pp. 1717–1739, doi: 10.1007/s00170-024-14890-0.

Wilcoxon, F. (1945), "Individual Comparisons by Ranking Methods", *Biometrics Bulletin*, [International Biometric Society, Wiley], Vol. 1 No. 6, pp. 80–83, doi: 10.2307/3001968.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Sustainability Assessment of Photovoltaic Solar Parks: A Four-Dimensional Delphi-Based Approach

Maria de Fátima Cabral [1], Adelina Baptista [2]

1 Master in Total Quality Management, Águeda School of Technology and Management, University of Aveiro, Portugal; corresponding author: mariafcabral@ua.pt

2 Águeda School of Technology and Management, University of Aveiro, Portugal

***Abstract*— This paper analyses the sustainability of photovoltaic solar parks using a qualitative Delphi approach structured around four dimensions: environmental, economic, social and technical. While renewable energy projects are commonly assessed through the traditional sustainability triad, technical constraints are often treated separately despite their direct influence on long-term performance.**

The study involved a pilot phase with three specialists followed by two Delphi rounds with a stable panel of twelve experts from Portugal, Spain and Chile. The pilot phase led to the inductive inclusion of the technical dimension, which became a central element of the proposed framework. The first round explored expert perceptions through semi-structured interviews, and the second round clarified priorities and relationships between dimensions.

Results show broad agreement on the environmental importance of solar parks in the energy transition, but also reveal significant challenges related to land use, biodiversity, waste management, uneven socio-economic benefits, community acceptance, storage constraints and long-term technological robustness. The findings demonstrate that sustainability cannot be inferred from low operational emissions alone and depends on how projects are located, designed, governed and integrated into local territories.

The paper proposes a four-dimensional sustainability assessment framework and illustrates how a qualitative Delphi process can support decision making in complex renewable energy projects. The results highlight the interdependence between environmental, socio-economic and technical factors, reinforcing the need for integrated approaches in quality engineering, project management and energy planning.

Keywords – photovoltaic solar parks, sustainability, renewable energy, Delphi method, energy transition

I. Introduction

Climate change, geopolitical instability and the persistent growth of energy demand have intensified the search for energy systems that are not only low-carbon but also socially and economically viable. Within this context, photovoltaic energy has become one of the most visible symbols of the energy transition. International climate commitments, especially the Paris Agreement and the 2030 Agenda, have reinforced the centrality of renewable energy in decarbonisation strategies and in the pursuit of Sustainable Development Goal 7, which promotes access to affordable, reliable and sustainable energy (United Nations Framework Convention on Climate Change, n.d.; United Nations, 2015).

However, the expansion of renewable energy does not automatically guarantee sustainability. The literature consistently shows that the relationship between energy supply, economic growth and environmental protection is complex and sometimes contradictory. Renewable technologies may reduce operational greenhouse gas emissions while simultaneously creating land-use pressure, new material dependencies, territorial conflicts and end-of-life waste streams (Hinrichs-Rahlwes, 2012; Ito, 2017; Dong et al., 2018). This ambiguity is particularly visible in the case of large-scale photovoltaic parks, whose rapid deployment is often accompanied by environmental, social and technical trade-offs.

In Portugal, this discussion is especially relevant. National renewable electricity production increased sharply over the last decade, and photovoltaic generation rose from 799 GWh in 2015 to 7,098 GWh in 2024, confirming the accelerating role of solar energy in the national electricity mix (Direção-Geral de Energia e Geologia, 2024). Yet the deployment of solar parks has also intensified debates on land occupation, impacts on local ecosystems, uneven economic benefits, limited public participation and the technical resilience of systems across their life cycle. The issue is not whether solar energy is preferable to fossil fuels in climate terms; it is whether the current mode of implementation is robust enough to support a genuinely sustainable transition.

This paper addresses this question by drawing on the author's research on the sustainability of photovoltaic solar parks. It adopts a four-dimensional framework (environmental, economic, social and technical) and uses a qualitative Delphi design to structure expert perceptions. The technical dimension is particularly relevant in this study because it emerged during the pilot phase as an essential complement to the three classical sustainability dimensions. The paper therefore asks: how sustainable are photovoltaic solar parks when assessed through an integrated, expert informed and life-cycle-oriented, throught a qualitative Delphi approach?

In addition to analysing expert perceptions, the study also contributes to the development of an integrated sustainability assessment framework intended to support decision-making in renewable energy planning, sustainability evaluation and quality-oriented project management.

To frame this contribution, the literature on sustainability was examined in the specific context of photovoltaic solar parks, considering the environmental, economic, social and technical dimensions adopted in this study.

The concept of sustainability has evolved from a resource-preservation concern into an integrated framework that combines environmental integrity, social equity and economic viability. In energy studies, this triadic understanding is now inseparable from debates on transition pathways, because decarbonisation requires structural changes in technologies, supply chains, territorial planning and governance (United Nations, 2015). For renewable-energy infrastructures, sustainability therefore cannot be reduced to the absence of direct emissions during operation; it must consider upstream and downstream effects, distributional outcomes and long-term technical performance.

Photovoltaic solar parks are frequently described as environmentally advantageous because electricity generation does not involve direct combustion and can displace fossil-based generation. Yet several studies warn that the sustainability of these systems depends on where and how they are implemented. Environmental concerns include biodiversity loss, landscape alteration, soil sealing or compaction, microclimatic effects, embodied impacts from manufacturing and transport, and the management of end-of-life waste (Hosenuzzaman et al., 2015; Pan and Hashemzadeh, 2023; Marcuzzo et al., 2022). Life-cycle thinking is therefore essential, because low operational emissions may coexist with relevant impacts during pre-installation, construction, operation, maintenance and decommissioning (Hauschild et al., 2017; Adiansyah et al., 2025).

In the Delphi process, social and economic issues emerged as strongly interconnected, leading to their joint interpretation as a socio-economic dimension in the second round. Large-scale solar parks can generate employment, stimulate local services and diversify regional economies, especially during construction. At the same time, benefits may remain short lived or unevenly distributed, while communities may perceive landscape change, agricultural land conversion and weak participation in decision-making as forms of territorial injustice (Bursztyn, 2018; Ramadhan and Naseeb, 2011; Alves et al., 2023). In practice, social acceptance is often shaped less by the environmental narrative of clean energy than by whether communities feel informed, heard and fairly treated. This places promoters, regulators and local authorities at the centre of sustainability governance.

The technical dimension is often underplayed in broader sustainability debates, even though it has direct consequences for all other dimensions. The efficiency decay of modules, storage constraints, compatibility between technologies, maintenance requirements and end-of-life recoverability influence environmental performance, cost structures and even public trust in the long term viability of projects (Bogdanov et al., 2019; Praene et al., 2021; Wanderley and Oliveira, 2024). Hussain et al. (2024) highlight the value of integrated sustainability approaches in renewable-energy projects, but the study on which this paper is based suggests that, in the specific case of solar parks, technical issues need to be treated not as a residual engineering detail but as a full sustainability dimension. This gap provides the main rationale for the present study.

II. METHODOLOGY

The study used a qualitative and exploratory Delphi design because the sustainability of photovoltaic solar parks is a complex and contested issue, combining technical, environmental and socio-territorial considerations that cannot be adequately captured through a single metric. Although Delphi is frequently associated with quantitative consensus building, it can also be used qualitatively when the aim is to structure, compare and refine expert perceptions in areas where multiple interpretations coexist (Hsu and Sandford, 2007; Skulmoski et al., 2007). In this research, Delphi was adapted to privilege thematic interpretation rather than statistical inference. Therefore, the purpose of the study was not to measure statistical consensus, but to identify, compare and interpret expert-based patterns of convergence, divergence and prioritisation. The analysis was based on qualitative thematic coding, supported by descriptive counts of coded mentions only as an auxiliary tool for organising the findings.

The methodological process followed the three stages proposed by Day and Bobeva (2005): exploration, distillation and use, as illustrated in Figure 1.

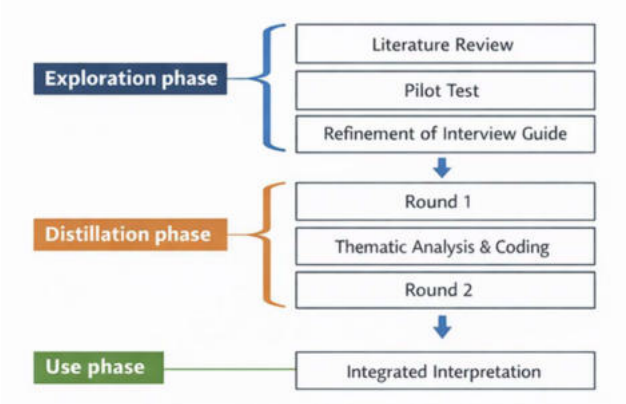


Fig. 1. Stages of the Delphi methodological process adopted in this study. *Source: Own elaboration, Adapted from Day and Bobeva (2005)*

The exploration phase encompassed the conceptual framing of the study, the development and refinement of the data collection instrument, and the initial configuration of the expert panel. This stage began with a literature review on climate change, sustainability, renewable energy and photovoltaic solar parks, with particular attention to life cycle impacts and sustainability trade-offs. The initial version of the interview guide was informed by existing frameworks on renewable energy project sustainability, particularly the work of Hussain et al. (2024), which provided a relevant conceptual starting point. However, given the exploratory nature of this study and its focus on capturing context specific expert perceptions, the instrument was subsequently reformulated into a more open and flexible structure.

A preliminary test was conducted with three specialists, playing a critical role in methodological refinement. The initial guide included only environmental, economic and social dimensions. During the pilot phase, experts consistently raised issues related to energy storage, degradation of panel performance over time, technological compatibility, maintenance and long-term system viability. These contributions revealed limitations in the initial conceptual structure and led to the inductive inclusion of a fourth dimension, technical sustainability. This adjustment represents one of the most distinctive aspects of the study, as it emerged directly from empirical input rather than being predefined.

Following this refinement, the interview guide was structured into twelve semi-structured questions, organised across the four dimensions. The evolution from the initial version to the final Round 1 interview guide is presented in Appendix A, Table IVThe environmental dimension addressed land use, ecosystem impacts and waste management; the economic dimension focused on local benefits, financial barriers and operational costs; the social dimension explored community effects and social acceptance; and the technical dimension examined system performance, durability and innovation. This stage ensured that the research instrument was both theoretically grounded and empirically validated.

In parallel, the expert panel was defined through purposive sampling, aiming to ensure diversity of experience, sectoral representation and technical expertise. A total of twenty-four invitations were sent, resulting in seventeen responses and a final panel of twelve specialists who participated in both Delphi rounds. The panel included experts from Portugal, Spain and Chile, with varied professional backgrounds spanning construction, energy, public institutions and non-governmental organisations. This diversity supported the objective of capturing multiple perspectives on the sustainability of photovoltaic solar parks. The main characteristics of the expert panel are summarised in Table I.

TABLE I.

EXPERT PANEL PROFILE

Characteristic	Summary
Invited / final panel	24 invited; 12 experts participated in both rounds
Geographic scope	Portugal (10), Spain (1), Chile (1)
Experience	5 with >15 years; 3 with 5–10 years; 2 with 11–15 years; 2 with <5 years
Organisation size	11 from large organisations (250+ employees); 1 from a medium-sized organisation
Sectoral profile	Construction/civil implementation predominated, with additional experts from energy, public entities and non-governmental organisations.

The distillation phase corresponded to the core Delphi process and involved two iterative rounds of data collection and analysis. The first round consisted of individual semi-structured interviews based on the validated guide, aiming to elicit in depth perceptions, identify key themes and explore areas of convergence and divergence. The data were analysed using qualitative thematic coding in WebQDA, structured around the four sustainability dimensions and their subthemes.

Although the analysis remained interpretative, a systematic approach was adopted to enhance transparency and analytical consistency. Recurring patterns mentioned by six or more experts were considered indicative of convergence, while less frequent patterns were analysed as partial or divergent insights. This approach did not aim at statistical generalisation but rather at structuring and interpreting expert perceptions in a consistent and traceable manner.

The second round translated the findings from the first round into a more focused questionnaire consisting of six questions, organised into environmental, socio-economic and technical dimensions, as presented in Appendix B, Table V This stage aimed to clarify priorities and deepen the understanding of previously identified issues, including environmental impacts, mitigation factors, recycling challenges, socio-economic benefits, social acceptance and technological constraints. Given that some questions allowed multiple responses, the analysis combined qualitative interpretation with descriptive counts of coded mentions. These counts supported the identification and comparison of patterns but were not treated as inferential statistics.

The main stages, instruments and outputs of the Delphi process are summarised in Table II.

TABLE II.

SUMMARY OF THE DELPHI RESEARCH DESIGN

Stage	Participants	Instrument	Main output
Pilot phase	3 experts	Pre-test of the interview guide 6 questions across 3 dimensions	Refinement of wording and inclusion of the technical dimension
Round 1	12 experts	Semi-structured interviews; 12 questions across 4 dimensions	Identification of themes, convergences, divergences and cross-dimensional links
Round 2	12 experts	Focused questionnaire; 6 questions derived from Round 1	Clarification of priorities, descriptive comparison of mentions and interpretive consolidation

Finally, the use phase involved the integrated interpretation of results across both rounds. This stage focused on synthesising the findings and examining the interrelationships between the environmental, economic, social and technical dimensions. Rather than analysing each dimension in isolation, the study adopted a systemic perspective, highlighting how decisions and impacts in one dimension influence the others. This integrative approach constitutes a central contribution of the research.

The objective of this design was not to produce statistical generalisations, but to develop a structured interpretive framework capable of supporting integrated sustainability assessment and decision making in complex engineering projects.

III. Findings and Discussion

The results obtained from the two Delphi rounds reveal that the sustainability of photovoltaic solar parks is perceived by specialists as both necessary and inherently complex. While solar energy is widely recognised as a key solution for reducing greenhouse gas emissions and supporting the energy transition (Pan and Hashemzadeh, 2023), the findings show that its sustainability cannot be assumed in absolute terms. Instead, it is conditional upon how projects are designed, implemented and integrated into their territorial, social and technical contexts. This interpretation is consistent with the challenges identified in the literature, which highlight that renewable energy systems, despite their environmental advantages, may generate significant trade-offs across multiple dimensions.

To provide an overview of the main findings, Table III summarises the key convergences and tensions identified across the two Delphi rounds. This synthesis reflects the evolution from initial perceptions in Round 1 to more structured and prioritised insights in Round 2.

~~TABLE V.~~ TABLE III.

SUMMARY OF THE MAIN FINDINGS ACROSS THE TWO DELPHI ROUNDS

Dimension	Main convergence	Main tension or caveat
Environmental	Climate benefits were consistently recognized across the panel; land use and biodiversity were prioritised as key impacts, strongly influenced by technical and territorial conditions	Environmental performance depends heavily on location, project scale, technological choices and end-of-life waste management
Economic	Short-term local economic benefits during the construction phase were widely acknowledged, often linked to social outcomes at community level	Benefits do not consistently translate into long-term territorial development or sustained local value
Social	Community acceptance was closely linked to transparency, participation and perceived local benefits, reflecting strong interdependence with economic outcomes	Communication alone was considered insufficient without effective value sharing, stakeholder engagement and trust-building mechanisms
Technical	Energy storage emerged as the most critical technological challenge, influencing both environmental performance and economic viability	Optimism about innovation coexists with uncertainty regarding long-term robustness, system integration and lifecycle sustainability

From an environmental perspective, the first Delphi round revealed strong convergence regarding key impacts, particularly land use change, vegetation removal, biodiversity loss and waste generation. These concerns align with existing literature, which emphasises that large-scale solar developments may significantly alter ecosystems and territorial dynamics (Pan and Hashemzadeh, 2023; Hernandez et al., 2015).

In the second round, these concerns became more clearly prioritised. Biodiversity loss and land use change were identified as the most critical impacts, while geographic location and environmental sensitivity emerged as the main factors for mitigation. This progression from identification to prioritisation reflects the iterative nature of the Delphi method and indicates a maturation of expert perceptions.

The findings reinforce the idea that low carbon emissions alone do not guarantee environmental sustainability. As highlighted by Rielli and Campos (2025), rigorous environmental impact assessments and effective mitigation and compensation mechanisms are essential to ensure that renewable energy projects do not generate disproportionate territorial impacts.

Waste management emerged as another critical issue across both rounds. Specialists consistently identified the end of life management of photovoltaic components as a major challenge, particularly due to technical complexity, high costs and limited recycling infrastructure (Marcuzzo et al., 2022; Domínguez et al., 2019). The degradation of modules over time, leading to increased waste generation, further reinforces these concerns (Oñate et al., 2024).

Importantly, the second round clarified that recycling challenges should not be interpreted as purely technical, but rather as a combination of technical, logistical and economic constraints. This aligns with literature highlighting the need for integrated lifecycle approaches and improved recycling technologies (Brito et al., 2023; Silveira and Urbanetz Junior, 2022; Konzen and Pereira, 2020).

The first Delphi round revealed a strong interconnection between economic and social aspects, leading to their integration into a single socio-economic dimension in the second round. This methodological adjustment reflects the empirical reality that economic benefits, social acceptance and stakeholder engagement are deeply intertwined.

From an economic perspective, specialists recognised that solar parks generate short-term local benefits, particularly during the construction phase, including employment, local services and economic activity. However, these benefits were consistently described as temporary and unevenly distributed, raising concerns about their long term contribution to territorial develop-ment. This finding is consistent with Bursztyn (2018), who highlights structural limitations in the economic sustainability of renewable energy systems, as well as with Ramadhan and Naseeb (2011), who point out the omission of key variables in eco-nomic viability assessments.

From a social perspective, the results highlight persistent challenges related to community acceptance, land-use conflicts and lack of stakeholder involvement (Granovsky-Larsen and Larreátegui Benavides, 2023; Alves et al., 2023; Mota et al., 2021). Across both rounds, specialists emphasised that resistance to solar projects is not necessarily directed at the technology itself, but rather at the way projects are implemented within local contexts.

The second round reinforced this interpretation by identifying transparent communication, stakeholder engagement, benefit-sharing mechanisms and participatory processes as key factors for improving acceptance and generating durable socio-economic value. These findings suggest that social acceptance is strongly linked to perceptions of fairness, inclusion and territorial benefit, rather than to environmental arguments alone.

In this dimension, one of the most significant contributions of this study concerns the role of project promoters. The findings indicate that promoters are not merely technical or regulatory actors, but central agents in shaping the relationship between solar projects and the territories in which they are implemented.

Beyond compliance with legal and technical requirements, promoters influence how impacts are communicated, how mitigation measures are designed and how value sharing mechanisms are implemented. In practice, the

sustainability of solar parks appears to depend less on the abstract environmental benefits of renewable energy and more on the concrete experiences of local communities, including transparency, participation and perceived fairness.

This reinforces the idea that socio-economic sustainability is strongly mediated by governance practices, stakeholder engagement and the ability of promoters to integrate territorial concerns into project design and implementation.

Another relevant insight emerging from the study relates to environmental impact assessment. Although not dominant across all responses, one critical perspective suggested that, in some contexts, environmental studies may be treated primarily as compliance requirements rather than as genuine decision support tools. This perception is analytically significant because it highlights a potential gap between formal procedures and effective environmental governance. If environmental assessments are perceived as procedural rather than substantive, their ability to ensure meaningful mitigation and to build trust may be compromised.

This concern is supported by the literature, which emphasises the importance of rigorous assessment processes and credible mitigation strategies in ensuring the sustainability of renewable energy projects (Campos et al. 2025). It also helps explain broader patterns of social resistance, reinforcing that trust is built through demonstrable responsiveness to local concerns rather than through communication alone.

The technical dimension, introduced following the pilot phase, proved to be a critical addition to the analysis. In the first round, specialists identified several technological challenges, including energy storage, equipment durability, technological compatibility and long-term maintenance.

In the second round, energy storage clearly emerged as the most critical issue, consistent with the literature highlighting its importance for system stability and renewable energy integration (Bogdanov et al., 2019). Other factors, such as durability and compatibility, were recognised but considered less critical in relative terms.

Despite these challenges, specialists expressed a generally optimistic view regarding future technological developments. How-ever, this optimism was often accompanied by caution, reflecting concerns about long-term performance, lifecycle impacts and technological maturity. These findings align with studies emphasising the need for integrated planning, lifecycle assessment and alignment with policy frameworks to ensure sustainable technological development (Prasad et al., 2017; Praene et al., 2021; Wanderley and Oliveira, 2024).

Across both Delphi rounds, one of the most important findings was the strong interdependence between environmental, socio-economic and technical dimensions. Environmental impacts were closely linked to technical decisions and territorial planning; economic benefits influenced social acceptance; and technological constraints affected both environmental and economic performance.

These interconnections support a systemic understanding of sustainability, in which environmental, socio-economic and technical dimensions cannot be analysed in isolation. Instead, they form a dynamic and interdependent system, where decisions in one dimension directly influence and condition outcomes in the others. This integrated perspective is synthesised in Figure 2, which illustrates the bidirectional relationships identified throughout the Delphi process.

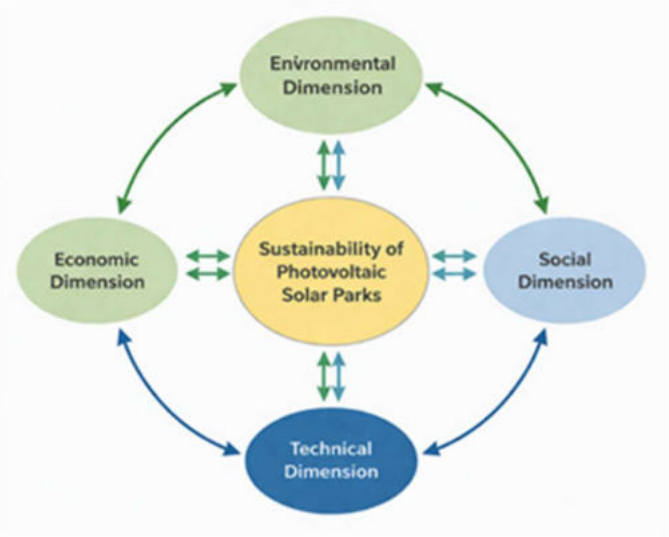


Fig.2. Interdependence of the sustainability dimensions of photovoltaic solar parks. Source: Own elaboration.

The diagram does not represent a hierarchy, but rather a conceptual framework in which sustainability emerges from the interaction between dimensions, reinforcing the need for holistic and context-sensitive approaches in the planning and implementation of photovoltaic solar parks.

V. CONCLUSIONS

This paper has shown that photovoltaic solar parks should be understood as a strategic but contested component of the energy transition. Based on a two-round qualitative Delphi study with a stable expert panel, the research confirms the environmental relevance of solar parks while demonstrating that their sustainability depends on much more than their contribution to decarbonisation. Land use, biodiversity, waste management, uneven socio economic benefits, limited community involvement, storage constraints and long-term technological robustness all influence how sustainable these projects actually are in practice.

The study contributes in two main ways. Substantively, it proposes a four-dimensional framework that explicitly includes technical sustainability alongside the environmental, economic and social dimensions. Methodologically, it shows how a qualitative Delphi design can be used to structure expert knowledge and support the analysis of complex sustainability problems without relying on inferential statistical approaches. The pilot phase, the inclusion of the technical dimension, the stable expert panel and the interpretation of both convergence and divergence are key elements of this contribution.

Beyond the specific case of photovoltaic solar parks, the proposed four-dimensional framework may be useful as a decision-support tool in renewable energy planning, sustainability assessment and quality-oriented project management, particularly in contexts where technical, environmental and socio-territorial factors interact.

The framework was not entirely predefined, as the technical dimension emerged inductively from the pilot phase.

From a practical standpoint, the results suggest that a more sustainable deployment of photovoltaic parks requires rigorous territorial planning, credible environmental assessment, stronger life-cycle thinking, clearer mechanisms of local value creation and structured stakeholder engagement by promoters and public authorities. The transition to renewable energy is therefore not only a matter of changing the energy source, but also, of how projects are designed, governed and embedded in local territories. Future research may extend this work by incorporating community perspectives directly and by testing the proposed four-dimensional framework in other national or project contexts.

REFERENCES

Adiansyah, J.S., Agusdinata, D.B. and Putra, A.P. (2025), "Environmental impacts of solar PV energy systems for small-island communities in Indonesia: A life cycle assessment approach", Energy for Sustainable Development, Vol. 85, 101651.

Alves, A., Marques da Costa, E., Gomes, E. and Niza, S. (2023), "Avaliação espacial da sustentabilidade para a implantação de parques solares fotovoltaicos em Portugal continental", Finisterra, Vol. 58 No. 124, pp. 63-84.

Bogdanov, D., Farfan, J., Sadovskaia, K., Aghahosseini, A., Child, M., Gulagi, A. and Breyer, C. (2019), "Radical transformation pathway towards sustainable electricity via evolutionary steps", Nature Communications, Vol. 10 No. 1, pp. 1-14.

Bursztyn, M. (2018), "Energia solar e desenvolvimento sustentável no Semiárido: o desafio da integração de políticas públicas", Estudos Avançados.

Campos, I., Maraschin, S. and Schwemmlein, K.M. (2025**)**, "Green, innovative, and unfair: The case of unjust energy transitions and civic resistance and opposition in Portugal", *Energy Research & Social Science*, Vol. 124, 104068.

Day, J. and Bobeva, M. (2005), "A generic toolkit for the successful management of Delphi studies", Electronic Journal of Business Research Methods, Vol. 3 No. 2, pp. 103-116.

Direção-Geral de Energia e Geologia (2024), Estatísticas rápidas – nº 241 – dezembro de 2024, DGEG.

Dong, K., Hochman, G., Zhang, Y., Sun, R., Li, H. and Liao, H. (2018), "CO2 emissions, economic and population growth, and renewable energy: Empirical evidence across regions", Energy Economics, Vol. 75, pp. 180-192.

Hauschild, M.Z., Rosenbaum, R.K. and Olsen, S.I. (2017), Life Cycle Assessment: Theory and Practice, Springer, Berlin/Heidelberg.

Hinrichs-Rahlwes, R. (2012), "Renewable energy: Paving the way towards sustainable energy security", Renewable Energy, Vol. 49, pp. 10-14.

Hosenuzzaman, M., Rahim, N.A., Selvaraj, J., Hasanuzzaman, M., Malek, A.B.M.A. and Nahar, A. (2015), "Global prospects, progress, policies, and environmental impact of solar photovoltaic power generation", Renewable and Sustainable Energy Reviews, Vol. 41, pp. 284-297.

Hsu, C.-C. and Sandford, B.A. (2007), "The Delphi technique: Making sense of consensus", Practical Assessment, Research & Evaluation, Vol. 12 No. 10, pp. 1-8.

Hussain, T., Wang, D. and Benqian, L. (2024), "Examining the role of responsible management, CSR, and TQM in enhancing renewable energy projects: An empirical analysis", Ecological Frontiers, Vol. 44, pp. 478-488.

Ito, K. (2017), "CO2 emissions, renewable and non-renewable energy consumption, and economic growth: Evidence from panel data for developing countries", International Economics, Vol. 151, pp. 1-6.

Latunussa, C.E.L., Ardente, F., Blengini, G.A. and Mancini, L. (2016), "Life cycle assessment of an innovative recycling process for crystalline silicon photovoltaic panels", Solar Energy Materials and Solar Cells, Vol. 156, pp. 101-111.

Marcuzzo, R., de Araujo, W.C., Maldonado, M.U. and Vaz, C.R. (2022), "A multi-country simulation-based study for end-of-life solar PV panel destination estimations", Sustainable Production and Consumption, Vol. 33, pp. 531-542.

Pan, Y. and Hashemzadeh, A. (2023), "Circular economy-based assessment framework for enhancing sustainability in renewable energy development with life cycle considerations", Environmental Impact Assessment Review, Vol. 103, 107289.

Praene, J.-P., Fakra, A.H., Benard, F., Ayagapin, L. and Rachadi, M.N.M. (2021), "Comoros's energy review for promoting renewable energy sources", Renewable Energy, Vol. 169, pp. 885-893.

Ramadhan, M., Naseeb, A. (2011), "The cost benefit analysis of implementing photovoltaic solar system in the state of Kuwait", Renewable Energy, Vol. 36 No. 4, pp. 1272-1277.

Rielli, L.E., Campos, I. (2025), "The process matters: Exploring public participation in solar energy projects in Brazil and Portugal", *Energy Research & Social Science*, Vol. 125, 104078.

Skulmoski, G.J., Hartman, F.T. and Krahn, J. (2007), "The Delphi method for graduate research", Journal of Information Technology Education, Vol. 6, pp. 1-21.

United Nations (2015), Transforming our world: The 2030 Agenda for Sustainable Development, United Nations, New York, NY, available at: https://sdgs.un.org/2030agenda.

United Nations Framework Convention on Climate Change (n.d.), The Paris Agreement, UNFCCC, available at: https://unfccc.int/process-and-meetings/the-paris-agreement.

Wanderley, M.V.L. and Oliveira, P.P. (2024), "Barreiras técnicas e econômicas na adoção da energia solar: um estudo bibliográfico", Revista Multidisciplinar do Nordeste Mineiro, Vol. 12, pp. 1-16.

Appendix A

TABLE IV.

EVOLUTION OF THE INTERVIEW GUIDE AFTER THE PILOT PHASE

Dimension	Initial Version — Pilot Phase	Expert Feedback and Suggestions	Final Interview Guide — Round 1
Environmental	1. What are the environmental impacts of solar parks? 2. Is there any visible impact on the territory? 3. Are there any concerns regarding waste?	Experts suggested a deeper analysis, including land use, impacts on ecosystems and waste management throughout the project life cycle.	1. What are the main environmental impacts of solar parks that should be considered when assessing their sustainability? 2. In your opinion, how do solar parks affect land use and local ecosystems? 3. What are the most critical challenges regarding waste management and recycling of solar park components?
Economic	1. Do solar parks generate employment? 2. Is there public support? 3. Are the costs competitive?	Experts suggested clarifying the type of benefits, further exploring financial barriers and comparing costs during the operation phase.	1. What economic benefits, local or regional, do solar parks generate? 2. What are the main financial barriers that limit the development of solar parks? 3. How do maintenance and operation costs of solar parks compare with those of conventional energy sources?
Social	1. Do local communities' benefit? 2. Is there social opposition? 3. Is there any impact on rural areas?	Experts recommended deepening the analysis of social acceptance and long-term effects in rural contexts.	1. How does the implementation of solar parks affect local communities? 2. What challenges are faced regarding the social acceptance of large-scale solar installations? 3. What positive or negative social impacts can solar parks generate in rural areas?
Technical	Not initially included.	Experts identified the absence of topics such as energy efficiency, panel degradation, technological compatibility and maintenance.	1. What technical factors do you consider most critical to ensure the sustainability of solar parks? 2. How does panel efficiency evolve over time and what implications does this have? 3. What challenges exist in the operation, maintenance and technological updating of installed systems?

Appendix B

TABLE V.

SECOND-ROUND FOCUSED QUESTIONNAIRE

Dimension	Second Round Questions
Environmental	**Q1.** Although there is consensus regarding the existence of environmental impacts caused by solar parks, in your opinion, which of these impacts are most relevant or severe? Examples may include soil alteration, biodiversity loss, local temperature increase, or changes in water or air flow. Which factors do you consider decisive for mitigating these impacts? **Q2.** Regarding recycling, all specialists recognised its complexity. What barriers do you identify, including technical, logistical or economic barriers?
Socio-Economic	**Q3.** From your perspective, what mechanisms or practices can ensure that solar parks generate effective and lasting economic benefits for local communities? Examples may include value-sharing mechanisms, compensation measures, local contracts or policies. **Q4.** Do you consider that promoters and responsible entities have been effective in involving and informing local communities? What strategies do you believe would be effective in promoting social acceptance, such as awareness-raising campaigns, direct compensation, inclusion in planning processes or other measures?
Technical	**Q5.** In your opinion, which of the following challenges is currently the most critical, and why: energy storage, equipment durability or compatibility between technologies? **Q6.** Do you think solar parks are technically prepared to respond to future sustainability and innovation requirements?

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Preparing the Workforce for Digital Quality Management: A Comparative Analysis of SAP Implementation Strategies in Two Luxury Goods Manufacturers

Micael Gonçalves[1], Marta Ribeiro[1], Joana Oliveira[1], Ana Lobarinhas[1], Sérgio Sousa[1], Senhorinha Teixeira[1] and Pedro Domingues[1]

[1] ALGORITMI Centre, Department of Production and Systems, School of Engineering, University of Minho, Guimarães, Portugal; micael.goncalves@dps.uminho.pt

***Abstract* — Digital transformation in manufacturing requires not only technological adoption but also effective workforce preparation, particularly in sectors reliant on artisanal skills. This study examines the implementation of SAP across two luxury leather goods component ateliers of ATEPELI, a leading company in Portugal, comparing workforce preparation strategies and their impact on adoption and operational outcomes. ATEPELI Penafiel, the first atelier to implement SAP, adopted a traditional approach focused on structured training and procedural documentation, providing clear guidance for quality processes. ATEPELI Ponte de Lima, the last to implement SAP, combined formal training with participatory initiatives, including gamified exercises, workflow simulations, and collaborative workshops, aimed at engaging employees and reducing resistance to change. Both implementations followed the group's standardized Core Model and SAP deployment methodology, ensuring technical consistency. Findings indicate that while traditional training ensures operational stability and process standardisation, incorporating interactive and experiential approaches enhances employee engagement, adaptation speed, and confidence in using the system. The sequential rollout allowed lessons learned from earlier cases to inform later implementations, highlighting the value of human-centred strategies in digital transformations within artisanal manufacturing contexts. The results suggest that integrating structured training with participatory activities maximises both technical adoption and workforce motivation, providing actionable insights for luxury manufacturers undertaking ERP implementations.**

Keywords – SAP Implementation, Workforce Preparation, Luxury Manufacturing, Digital Transformation

I. Introduction

In the era of Industry 4.0, the digital transformation of manufacturing industries has emerged as a critical imperative for organizations seeking to maintain competitiveness in an increasingly dynamic global market, offering benefits such as increased productivity, reduced operating costs, improved product quality, and product innovation (Ghobakhloo and Iranmanesh, 2021). Among the various technological innovations reshaping industrial operations, Enterprise Resource Planning (ERP) systems stand out as solutions that integrate business areas through a shared information environment across the organization, enhancing data visibility and supporting informed decision-making (Idilbi and Abu-Shanab, 2022).

A notable example of such is SAP, a leading ERP solution that has been widely adopted across sectors to streamline operations and improve organizational efficiency. However, evidence from multiple case studies suggests that successful SAP implementation depends critically on organizational readiness, cultural alignment, and the adoption of structured project management methodologies (Madathala *et al.*, 2025).

The luxury goods industry presents a particularly complex context for digital transformation. Unlike mass-market manufacturing, luxury production is characterised by a profound reliance on manual craftsmanship, artisanal expertise, and heritage techniques that have been refined over generations. This is precisely why luxury consumers are willing to pay premium prices: they associate these products with superior quality, exclusivity, and reputation (Sanz-Lopez *et al.*, 2024). Consequently, any transformation that risks disrupting established quality standards or the craftsmanship culture poses significant strategic risks (Überbacher, Brozzi and Matt, 2020; Holmqvist, Wirtz and Fritze, 2021; Gonçalves *et al.*, 2025, 2026). This is reinforced by the literature, which frequently characterises craft-based industries as lagging behind other sectors in terms of digital transformation maturity (Kocak and Pawlowski, 2023). Navigating this tension requires an approach that values employee engagement and empowerment alongside technological advancement (Cirillo *et al.*, 2021). In this context, the concept of digital literacy becomes crucial. Cetindamar Kozanoglu and Abedin (2021) define it as a multidimensional organisational capability that encompasses the skills, knowledge and abilities employees need to effectively use digital technologies in their work contexts.

The literature on change management has proposed various approaches to preparing workforces for digital transitions. Traditional approaches have emphasised structured training programmes designed to develop specific technical competencies (Bokek-Cohen, 2018). These programmes typically follow a formal curriculum, delivered through classroom instruction or learning modules, with clearly defined learning objectives and assessment criteria. Such approaches offer consistency and scalability, ensuring that all employees receive standardised information about new systems and processes.

An alternative paradigm has emerged that emphasises experiential engagement and participative approaches to change. Gamification has been widely adopted by organizations to support training and education, employing game design elements such as points, leaderboards, stories, rewards, challenges, and competitions in non-game contexts (Wang, Hsu and Fang, 2022). It has been shown to reduce resistance to change by encouraging voluntary learning among employees, improving motivation, engagement, and the development of knowledge, skills and abilities relevant to organizational tasks (Bitrián *et al.*, 2023).

Despite the growing body of research on ERP implementation and human involvement strategies in digital transformation, studies examining different approaches to workforce preparation in luxury manufacturing contexts remain scarce. The luxury sector's unique characteristics, which relies heavily on manual labour and high product quality, suggest that insights from other industries may not directly apply.

The present study addresses this gap by conducting a comparative analysis of two distinct approaches to workforce preparation for SAP implementation within the same luxury goods group in Portugal. The first approach, implemented at ATEPELI Penafiel, emphasized structured training and procedural documentation. The second approach, implemented at ATEPELI Ponte de Lima, was developed after the experience at ATEPELI Penafiel and therefore benefited from the feedback and lessons learned gathered during the experience at ATEPELI Penafiel. Building on this prior knowledge, it incorporated gamification, participatory activities, and experiential engagement alongside formal training.

The main objective is to evaluate the effectiveness of these approaches across key implementation outcomes, including employee adaptation, system adoption, operational continuity, and quality performance. By comparing these two strategies, the study contributes to the existing literature on ERP implementation, digital transformation, and change management by offering insights into how human-centered and preparation strategies can support technology adoption and organizational resilience in artisanal industrial contexts.

The remainder of this paper is structured as follows: first, the "Methodology" section outlines the research approach employed throughout this project. Section III describes the two SAP implementation case studies and the corresponding workforce preparation strategies. The following section, "Results", presents the findings of the comparative analysis between the two workforce preparation approaches. Next, the "Discussion" section addresses its contributions and identifies remaining gaps. Finally, the "Final Remarks" section summarizes the main findings and implications of this work.

II. Methodology

This study adopts a comparative case study methodology, suitable for investigating contemporary phenomena within real-life organisational contexts, particularly where the boundaries between phenomenon and context are not clearly evident (Yin, 2018). The research focuses on two industrial units belonging to the same luxury goods group in Portugal—ATEPELI Penafiel and ATEPELI Ponte de Lima—where distinct approaches to workforce preparation for SAP implementation were deployed (Oliveira, 2023; Lobarinhas, 2024). This design enables a controlled comparison, as both units share similar organisational structures, production processes, and strategic objectives, while differing primarily in the change management approach adopted.

A mixed-methods approach was employed to capture both quantitative performance outcomes and qualitative insights into employee experiences. Quantitative data were collected from internal company records covering key performance indicators (KPIs) before, during, and after SAP implementation. For the Ponte de Lima unit, these KPIs included system adoption rates, production efficiency, error rates, and quality performance metrics. Such indicators are commonly used in ERP evaluation studies to assess implementation success and operational impact (DeLone and McLean., 2003; Nicolaou, 2004).

Qualitative data were gathered through a survey and non-participant observation in the Penafiel unit. The survey was conducted with employees to capture diverse perspectives on training effectiveness, engagement, and perceived challenges. The survey protocol was designed following established guidelines for case study research (Voss, 2010), ensuring consistency while allowing flexibility to explore emergent themes. It consisted of 10 questions, including two demographic items (age and gender), followed by items addressing employee feedback on the implemented practices. The final question allowed respondents to provide open-ended suggestions for improvement and future development. Additionally, observational data were collected during training sessions and early system usage phases to assess behavioural engagement and interaction with the implemented approaches.

To analyse the qualitative data, a thematic analysis was conducted, following the procedure outlined by Braun and Clarke (2006). The use of thematic analysis is appropriate for exploring perceptions and experiences in organisational change contexts, particularly in studies involving digital transformation (Klein and Myers, 1999).

Quantitative data were analysed using descriptive and comparative statistical techniques, enabling the identification of trends and differences between the two units. Pre- and post-implementation performance metrics were compared to assess the relative effectiveness of the structured training approach versus the gamified and participatory approach. Where appropriate, percentage variations and performance ratios were calculated to standardise comparisons across units.

To enhance the validity and reliability of the findings, data triangulation was employed by integrating multiple data sources (a survey, observations, and performance records), as recommended in case study research (Yin, 2018). Furthermore, the use of a single organisational group reduces external variability, strengthening internal comparability. Ethical considerations were also addressed, ensuring confidentiality and informed consent for all participants.

Overall, this methodological approach enables a robust and context-sensitive evaluation of workforce preparation strategies for ERP implementation in the luxury manufacturing sector, contributing to both theory and practice in digital transformation and change management.

III. Case Studies

ATEPELI is a leading producer of luxury leather goods and footwear components, founded in 2011 and part of one of the largest multinational groups in the sector, with a presence in over 80 countries. The company operates three manufacturing units in northern Portugal—Ponte de Lima, Penafiel, and Santa Maria da Feira—distinguished by a focus on craftsmanship, quality, and continuous innovation.

Within the group, ATEPELI manages three specialised component ateliers: ATEPELI Penafiel, ATEPELI Ponte de Lima, and an additional unit in Romania. Unlike manufacturers of finished luxury goods, these ateliers focus exclusively on producing components—such as handles, straps, small leather elements, and internal parts for Small Leather Goods (SLG) and external elements for Leather Goods (LG)—that are later integrated into final luxury products (Fig. 1). The high volume and artisanal nature of production demand a flexible and agile organisational structure to ensure performance standards are met, making this a particularly relevant context for analysing digital transformation and workforce preparation.

Fig. 1. Examples of components produced at ATEPELI.

In response to increasing market demands for both quality and responsiveness, ATEPELI's group launched a global transformation initiative across its ateliers, currently known as One Manuf. This project aims to redesign manufacturing processes by pursuing three objectives: increasing speed through lead time reduction, simplifying operations to enhance efficiency, and improving coordination across production stages (Lobarinhas, Gonçalves and Teixeira, 2024). At the core of this initiative is a standardized "Core Model", supported by SAP, which integrates processes, improves inventory control, and enhances collaboration between ateliers and suppliers.

The SAP rollout was carried out progressively across the group's component ateliers. ATEPELI Penafiel was the first to implement the system, followed by the Romanian atelier, and finally Ponte de Lima—the 19th SAP deployment within the group. This sequential approach created opportunities for learning and adaptation between cases, while also highlighting differences in workforce preparation strategies, which form the focus of this study.

The following section describes the two case studies—ATEPELI Penafiel and ATEPELI Ponte de Lima.

A. ATEPELI Penafiel

The SAP implementation at ATEPELI Penafiel followed a structured and technically oriented approach, focusing on process standardisation and formal training. As the first specialised component atelier within the group to implement SAP, Penafiel played a key role in the initial deployment of the system, with limited prior internal references to support the process. The implementation was conducted through close collaboration between a central project team and a local team composed of representatives from each department, ensuring alignment between global requirements and local operational needs.

The implementation process followed a phased structure aligned with established ERP methodologies, including initial preparation, system design and configuration, and final transition. In parallel, the project was guided by the group's internal framework, structured around three main stages: Core Model, Convergence, and Bascule. The Core Model phase focused on defining processes and system specifications with input from different ateliers, while the Convergence phase involved process documentation, system alignment, and the development of standard operating procedures. Finally, the Bascule phase included business simulations, user training, system transition (Go-Live), and the monitoring of key performance indicators and operational routines.

Within the quality department, the implementation required a detailed analysis of existing processes, including material quality, internal quality, new product quality, and external quality. These processes were mapped and adapted to fit the SAP Quality modules, ensuring that all relevant activities were integrated into the new system. This effort enabled greater visibility over quality operations and supported more consistent and structured data management across the department.

Workforce preparation in this case relied primarily on formal and structured training mechanisms. A comprehensive SAP procedure manual was developed to support users in executing their daily tasks, providing detailed guidance on system transactions across different areas of the quality department. This document served as a key reference tool during and after the transition, contributing to process standardisation and reducing operational variability. The transactions required to perform different tasks within the system are synthetised in TABLE I.

TABLE I. LIST OF THE MAIN TRANSACTIONS USED IN THE SAP SYSTEM.

Transaction	Definition	Function
NCOOIS	Production Order Information System	General transaction: search for production orders, materials, flows, production lines, orders in quarantine, or blocked orders
VA01	Create Customer Order	Creation of Return Sales Order
VA03	Display Customer Order	Display Return Sales Order
VL02N	Change Outbound Delivery	Receipt of Components for Repair
VL74	Output from Handling Units	Label Printing
LT10	Create Transfer Order from Stock Summary	Labelling error (label data inconsistency)
ZVA06	Sales Order Cockpit	Display Return Deliveries
VL06O	Outbound Delivery Monitor	Display Deliveries
LX03	Bin Status Report	Display Stock per Bin
MB56	Batch Where-Used List	Display Production Line for Returns

In addition to the written procedures, training sessions were conducted for employees in the quality department, focusing on the practical use of the SAP system. These sessions aimed to ensure that users were able to perform their tasks within the new digital environment and to clarify any doubts related to system functionalities. Support was also provided during the transition phase, allowing users to progressively adapt to the new system while maintaining operational continuity.

B. ATEPELI Ponte de Lima

The SAP implementation at ATEPELI Ponte de Lima was characterised by a strong emphasis on human management and employee engagement, recognising the critical role of workforce preparation in ensuring a successful digital transition. As the oldest manufacturing site within ATEPELI, the atelier employs artisans with long tenure, in some cases exceeding a decade of experience, making changes to established routines particularly challenging.

To address this, ATEPELI invested in a set of initiatives aimed at familiarising employees with the upcoming changes and promoting their active involvement in the transformation process. One of the most distinctive activities was the creation of a local project mascot. As part of the One Manuf initiative, each atelier develops its own identity, and in Ponte de Lima this process was conducted through a collaborative workshop involving employees and a local social association (Fig. 2 (a)). The activity was extended across production lines and departments, with each team creating its own representation (Fig. 2 (b)). Employees then voted for their preferred option, resulting in the selection of the mascot "Minhota" (Fig. 2 (c)), fostering a sense of ownership and emotional connection with the project.

(a) (b) (c)

Fig. 2. Mascot development process: (a) Activity: Creation of the mascot; (b) Example of the dolls built during the activity; and (c) Final mascot of the project: "Minhota".

In addition to symbolic engagement, several interactive dynamics were implemented to support understanding of operational changes. A layout change activity was conducted using challenges and games to help employees understand the benefits of the new production layout associated with the Logistics and Cutting (LC) building (Fig. 3).

Fig. 3. Activity: Layout Changes.

The inauguration of this new facility was also marked by a dedicated event involving all transferred employees, reinforcing the importance of the transformation and promoting a positive perception of change (Fig. 4).

Fig. 4. LC building inauguration.

Furthermore, a "Marketplace" dynamic was organised to simulate the future production flow using physical elements such as LEGO, allowing employees to visualise how SAP would impact their daily activities and better understand system functionalities and constraints (Fig. 5).

Fig. 5. Activity: Marketplace.

To prepare for the SAP Go-Live at Ponte de Lima, a dedicated laboratory was set up in the LC building, with computers installed at each step of the production process and a simulator warehouse to test the full manufacturing flow. The system was simulated for several weeks, allowing potential errors to be identified and corrected before the official launch (Fig. 6).

Fig. 6. SAP Laboratory inauguration.

Each department received detailed manuals developed by the project team, providing step-by-step guidance for system use, and all affected employees underwent targeted training sessions to familiarize themselves with the new processes. A complete inventory of the atelier was conducted, preceded by two trial inventories, ensuring that final stock data was

accurate before the Go-Live. Production was temporarily paused for two days while the project team prepared the system and verified readiness. To keep the team motivated and focused despite the natural stress, various activities were organized, including games, sports, martial arts classes, and a barbecue lunch (Fig. 7).

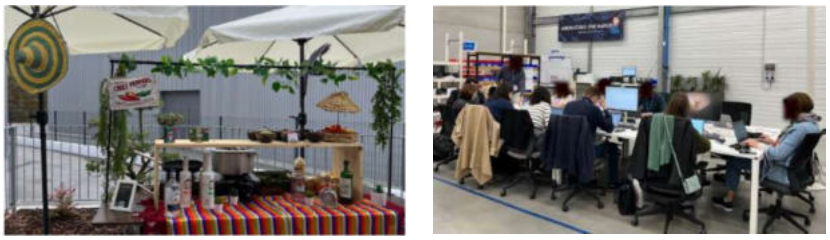


Fig. 7. Work and activities during the final days before the Go-Live.

The first production order was successfully tested through the entire SAP workflow, with no errors, and initial shipments were dispatched alongside the dolls created during the mascot activity (Fig. 8).

Fig. 8. First order dispatched using SAP.

On the Go-Live day, the atelier was decorated, employees received project-related souvenirs, and small engagement activities were organized to celebrate the milestone and maintain motivation (Fig. 9).

Fig. 9. SAP Go-Live celebration.

IV. RESULTS

The results obtained from both case studies highlight the impact of different workforce preparation strategies on SAP adoption and operational performance within luxury leather goods manufacturing.

A. ATEPELI Penafiel

The implementation of SAP at ATEPELI Penafiel had a direct impact on the Quality Department, making employee involvement a critical success factor. A questionnaire was conducted among all 11 employees, achieving a 100% response rate, to assess perceptions regarding the new system.

The findings indicate that the main benefits perceived by employees included improved traceability, more efficient responses to quality blockages, enhanced cross-departmental information flow, and better stock management. These improvements are directly associated with the implementation of specific SAP Quality modules, particularly those related to stock visibility and quality control processes.

However, some limitations were also identified. The most significant drawback reported was the high level of process repetitiveness, mainly due to the constant registration of item locations. Additionally, increased bureaucratisation was observed in certain processes, particularly in external quality management, where resistance from clients led to delays and temporary workflow disruptions.

Regarding training effectiveness, the majority of employees (82%) considered the training adequate for their roles. Nevertheless, the absence of timely training for the external quality module negatively affected adaptation in that area. Despite this limitation, overall system adaptation was perceived as relatively fast and intuitive by 91% of respondents, supported by structured procedures, formal training sessions, and continuous support during the Go-Live phase.

Finally, employees were also given the opportunity to provide suggestions for operational improvements to the system, focused on improving data accessibility, reducing manual intervention in quality workflows, and increasing the reliability of operational information. Overall, these suggestions reflect users' direct interaction with the SAP system in daily operations and provide valuable insights for continuous improvement of the implemented solution.

B. ATEPELI Ponte de Lima

In contrast, the SAP implementation at ATEPELI Ponte de Lima adopted a more participatory and experiential approach to workforce preparation. Despite the unconventional investment in gamified and interactive activities, these initiatives yielded highly positive outcomes.

Employee engagement levels were significantly high, with even long-tenured artisans demonstrating openness and enthusiasm towards the new system. This positive perception contributed to a smooth transition, with no interruptions in productivity during the week following Go-Live.

Operational performance indicators further reinforce the success of this approach. The final inventory presented only a 0.01% deviation, and manufacturing efficiency exceeded 100%. Additionally, the implementation was recognised internally as a benchmark, being praised by its international counterparts.

Overall, the combination of experiential learning, active participation, and continuous communication proved effective in reducing resistance to change, increasing employee confidence, and accelerating system adoption. This approach contrasts with more traditional training models by actively involving workers in the transformation process, thereby supporting smoother system adoption and minimising operational disruptions during SAP implementation.

V. DISCUSSION

The comparative analysis of the SAP implementations at ATEPELI Penafiel and Ponte de Lima highlights the impact of different workforce preparation strategies on the success of digital transformation in luxury manufacturing (Table II). Both ateliers implemented the same system and followed the core stages of ERP deployment—Core Model, Convergence, and Bascule—ensuring consistency in technical configuration and process alignment. However, the approaches to human management and training differed significantly, revealing important insights for managing digital transitions in contexts with highly skilled and traditionally oriented workforces.

TABLE II. SUMMARY OF KEY FINDINGS FROM THE COMPARATIVE ANALYSIS.

Dimension	ATEPELI Penafiel (Structured Training Approach)	ATEPELI Ponte de Lima (Participatory & Experiential Approach)	Comparative Insight
Training Strategy	Formal, structured training; detailed SAP manuals; classroom-style instruction	Gamified activities, simulations (Marketplace), mascot creation, layout games, hands-on SAP lab	Both approaches ensured system readiness, but experiential methods increased engagement
Employee Engagement	Moderate; some resistance among long-tenured artisans; adaptation supported by documentation	Very high; strong enthusiasm and emotional connection to the project; reduced resistance	Participatory activities significantly boosted motivation and openness to change
Adaptation Speed	Generally fast, though slowed in areas lacking timely training (e.g., external quality module)	Very fast; no productivity drop during Go-Live week; smooth transition	Experiential preparation accelerated adaptation compared to traditional training
Operational Continuity	Stable transition; some workflow disruptions due to incomplete module readiness	Fully stable; no interruptions; manufacturing efficiency >100%	Participatory approach better preserved operational continuity
Quality & Process Outcomes	Improved traceability, stock visibility, and information flow; some bureaucratisation	High inventory accuracy (0.01% deviation); strong performance in early SAP usage	Both improved quality processes, but Ponte de Lima achieved superior early performance
Resistance to Change	Present, especially among long-tenured employees; mitigated by structured support	Minimal; engagement activities reduced anxiety and built ownership	Human-centred strategies more effective in reducing resistance
User Confidence & System Adoption	Good overall; dependent on manuals and post-Go-Live support	Very high; hands-on simulations increased confidence before Go-Live	Experiential learning improved confidence and autonomy
Recognition & Benchmarking	Internal success but limited external visibility	Recognised as a benchmark by international counterparts	Participatory approach produced standout organisational results
Overall Effectiveness	Ensured standardisation and operational stability	Maximised engagement, motivation, and performance	Combining both approaches may yield optimal results

At Penafiel, the first atelier to implement SAP, the focus was primarily on formal and structured training. Detailed SAP procedure manuals were developed, providing comprehensive guidance on transactions across the quality department. Training sessions were conducted to familiarise users with system functionalities, and support was offered during the Bascule phase to facilitate adaptation. This approach standardised knowledge, reduced operational errors, and enabled a relatively smooth transition, as most employees reported that the system was intuitive and their adaptation was rapid. Nevertheless, certain limitations were observed, particularly regarding the timing of training and incomplete module development, which affected user preparedness, especially for external quality processes. Resistance to change among long-tenured employees was also noted, reflecting the challenges of implementing digital tools in environments with entrenched routines.

In contrast, the Ponte de Lima implementation emphasized experiential learning and employee engagement alongside structured training. Initiatives such as the creation of the atelier mascot, layout simulations, and the Marketplace exercise actively involved employees in understanding the operational and technological changes (Fig. 2–9). Despite the unusual investment in recreational activities for a project of this nature, very positive outcomes were achieved by focusing on keeping every employee as involved as possible. The excitement and positive spirit generated by these initiatives created a sense that this significant change was for the better, even among artisans who had been with the company for over 13 years. The biggest challenge in projects like this is keeping people motivated, as everyone's cooperation is crucial for success. By incorporating insights from the Penafiel unit and combining gamified exercises, collaborative workshops, and practical simulations, Ponte de Lima succeeded in reducing resistance to change, while maintaining high levels of operational continuity.

Comparing the two cases reveals clear distinctions in outcomes related to workforce preparation. While Penafiel achieved operational stability through formal documentation and traditional training, Ponte de Lima achieved both operational efficiency and higher employee engagement by incorporating participatory methods. This suggests that structured training alone may suffice for functional adoption, but combining it with interactive, experiential approaches can significantly enhance motivation, engagement, and resilience during the transition. The sequential rollout within the group—Penafiel first, Romania second, and Ponte de Lima last—also provided an opportunity to incorporate lessons learned, which likely contributed to the smoother experience at Ponte de Lima.

Overall, these findings highlight the critical role of human-centred strategies in digital transformations within luxury manufacturing contexts. Traditional, artisanal work environments are often characterised by strong organizational identity, where employees identify closely with craft-based practices, which can intensify resistance to change during the introduction of standardized digital systems such as SAP. Such industries benefit from formal procedures to ensure consistency and error reduction, but engagement-focused interventions appear to accelerate adaptation, reduce resistance, and strengthen employees' confidence in using new digital tools. For organizations planning similar SAP deployments, integrating structured training with experiential and participatory activities can therefore maximize both technical and human outcomes.

VI. FINAL REMARKS

This study examined two distinct approaches to workforce preparation for SAP implementation within the ATEPELI luxury leather goods group. ATEPELI Penafiel, as the first atelier to adopt the system, relied primarily on formal training and detailed procedural documentation to ensure process standardization and minimize operational errors. In contrast, ATEPELI Ponte de Lima combined structured training with experiential and participatory activities, such as mascot creation, layout simulations, and a Marketplace exercise, aiming to engage employees actively and reduce resistance to change.

The comparative analysis revealed that while formal, documentation-based training can facilitate technical adoption and ensure continuity, integrating interactive and motivational initiatives significantly enhances employee engagement, adaptation speed, and confidence in using new systems. The sequential implementation across the group allowed lessons learned from earlier deployments to inform subsequent ateliers, contributing to smoother transitions and higher overall efficiency, particularly at Ponte de Lima.

Key outcomes from this research underscore the critical importance of human-centred strategies in digital transformations, particularly in artisanal and luxury manufacturing contexts, where deeply ingrained traditional practices coexist with modern technological systems. The findings suggest that organizations should not only focus on technical implementation but also invest in structured engagement and experiential learning, aligning employees' motivations with operational objectives to maximize adoption success.

Despite its contributions, this study presents some limitations. Different performance and perception indicators were prioritised in each context, which may limit the direct comparability of results. Furthermore, the sequential nature of the implementations means that the second case was informed by the experience of the first, which influenced both the design of the intervention and the outcomes observed, introducing a degree of learning bias. Nevertheless, this also reflects a realistic organisational setting, where prior implementations naturally inform subsequent ones. For future applications, the company can benefit from systematically consolidating lessons learned from both cases, identifying what worked well and what did not, and standardising key indicators to improve the evaluation framework in subsequent implementations.

Notwithstanding these limitations, this study contributes practical insights for luxury manufacturers and similar industries aiming to implement ERP systems. By combining formal training with participatory and gamified initiatives, companies can achieve both operational stability and a motivated workforce, ensuring that digital transformation delivers its intended benefits in quality, efficiency, and innovation.

ACKNOWLEDGMENT

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Unit Project Scope UID/00319/2025 - Centro ALGORITMI (ALGORITMI/UM) https://doi.org/10.54499/UID/00319/2025.

REFERENCES

Bitrián, P. *et al.* (2023) "The use of gamification strategies to enhance employees' attitudes towards e-training systems," *The International Journal of Management Education*, 21(3), p. 100892. Available at: https://doi.org/10.1016/j.ijme.2023.100892.

Bokek-Cohen, Y. (2018) "Conceptualizing employees' digital skills as signals delivered to employers," *International Journal of Organization Theory & Behavior*, 21(1), pp. 17–27. Available at: https://doi.org/10.1108/IJOTB-03-2018-003.

Braun, V. and Clarke, V. (2006) "Using thematic analysis in psychology," *Qualitative Research in Psychology*, 3(2), pp. 77–101. Available at: https://doi.org/10.1191/1478088706qp063oa.

Cetindamar Kozanoglu, D. and Abedin, B. (2021) "Understanding the role of employees in digital transformation: conceptualization of digital literacy of employees as a multi-dimensional organizational affordance," *Journal of Enterprise Information Management*, 34(6), pp. 1649–1672. Available at: https://doi.org/10.1108/JEIM-01-2020-0010.

Cirillo, V. *et al.* (2021) "Technology vs. workers: the case of Italy's Industry 4.0 factories," *Structural Change and Economic Dynamics*, 56, pp. 166–183. Available at: https://doi.org/10.1016/j.strueco.2020.09.007.

Delone, W.H. and McLean., E.R. (2003) "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," *Journal of Management Information Systems*, 19(4), pp. 9–30. Available at: https://doi.org/10.1080/07421222.2003.11045748.

Ghobakhloo, M. and Iranmanesh, M. (2021) "Digital transformation success under Industry 4.0: a strategic guideline for manufacturing SMEs," *Journal of Manufacturing Technology Management*, 32(8), pp. 1533–1556. Available at: https://doi.org/10.1108/JMTM-11-2020-0455.

Gonçalves, M. *et al.* (2026) "Quality Management Processes for Sap Implementation in a Luxury Leather Goods Industry," *Journal of Engineering, Management and Information Technology*, 4(1), pp. 37–50. Available at: https://doi.org/10.61552/JEMIT.2026.01.004.

Gonçalves, M.T. *et al.* (2025) "Analysis of Quality Management Procedures in a Luxury Leather Goods Industry." Available at: https://doi.org/10.14488/IJCIEOM2025_FULL_0087_38238.

Holmqvist, J., Wirtz, J. and Fritze, M.P. (2021) "Digital Luxury Services: Tradition versus Innovation in Luxury Fashion," *Services Marketing*. WORLD SCIENTIFIC (US), pp. 550–552. Available at: https://doi.org/10.1142/9781944659806_0018.

Idilbi, M. and Abu-Shanab, E.A. (2022) "Critical Success Factors for ERP Implementation," *International Journal of Technology and Human Interaction*, 18(1), pp. 1–22. Available at: https://doi.org/10.4018/IJTHI.297613.

Klein, H.K. and Myers, M.D. (1999) "A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems1," *MIS Quarterly*, 23(1), pp. 67–93. Available at: https://doi.org/10.2307/249410.

Kocak, S. and Pawlowski, J. (2023) "Digital Transformation in the Craft Sector: Identifying Digital Competencies and Attitudes for Managers and Employees," *Proceedings of the 2023 Computers and People Research Conference on ZZZ*. New York, NY, USA: ACM, pp. 1–6. Available at: https://doi.org/10.1145/3579168.3632720.

Lobarinhas, A.L. (2024) *Evaluation of quality management procedures for the implementation of SAP in a luxury leather goods industry*. Master's thesis. University of Minho.

Lobarinhas, A.L., Gonçalves, M.T. and Teixeira, S.F.C.F. (2024) "The Role of Employees in Implementing an SAP System - A Practical Case in the Luxury Leather Goods Industry," *11º ENEGI - Encontro Nacional de Engenharia e Gestão Industrial*. Coimbra, Portugal.

Madathala, H. *et al.* (2025) "Navigating SAP ERP Implementation: Identifying Success Drivers and Pitfalls," *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, pp. 75–83. Available at: https://doi.org/10.1109/IDCIOT64235.2025.10914890.

Nicolaou, A.I. (2004) "Firm Performance Effects in Relation to the Implementation and Use of Enterprise Resource Planning Systems," *Journal of Information Systems*, 18(2), pp. 79–105. Available at: https://doi.org/10.2308/jis.2004.18.2.79.

Oliveira, J. (2023) *Contributos para a implementação de módulos SAP QUALIDADE numa empresa de produção de componentes de marroquinaria de luxo*. Master's thesis. University of Minho.

Sanz-Lopez, F. *et al.* (2024) "Is the digitalisation the future of the luxury industry?," *Heliyon*, 10(21), p. e40029. Available at: https://doi.org/10.1016/j.heliyon.2024.e40029.

Überbacher, R., Brozzi, R. and Matt, D.T. (2020) "Innovation in craftsmanship family SMEs in times of digitalization," *Picola Empressa Small Business*. Available at: https://doi.org/10.14596/pisb.345.

Voss, C. (2010) "Case Research in Operations Management," *Researching Operations Management*. Routledge, pp. 176–209.

Wang, Y.-F., Hsu, Y.-F. and Fang, K. (2022) "The key elements of gamification in corporate training – The Delphi method," *Entertainment Computing*, 40, p. 100463. Available at: https://doi.org/10.1016/j.entcom.2021.100463.

Yin, R.K. (2018) *Case Study Research and Applications: Design and Methods*. Sixth Edition. Thousand Oaks, California: SAGE Publications, Inc.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

The future of quality in manufacturing industry: From Quality 4.0 to Quality 5.0

Inês Rocha Melo [1], Pedro Marques [1,2] , Graça Silva [2], and Pedro Saraiva [3]

[1] Faculty of Engineering - Industrial Engineering and Management, Universidade Lusófona, Campo Grande 376, 1749-024 Lisbon, Portugal; ines.rocha.melo@ulusofona.pt (Inês Rocha Melo); p5037@ulusofona.pt (Pedro Marques)

[2] ISEG - Lisbon School of Economics and Management, Universidade de Lisboa, Rua do Quelhas 6, 1200-781 Lisbon, Portugal; p5037@ulusofona.pt (Pedro Marques); gracamsilva@iseg.ulisboa.pt (Graça Silva)

[3] University of Coimbra, Chemical Process Engineering and Forest Products Research Centre, Coimbra, Portugal; pas@eq.uc.pt

***Abstract* - This paper examines the transition from Quality 4.0 to Quality 5.0 in the manufacturing industry, with the aim of supporting organisations in navigating this paradigm shift. Through a literature review of academic publications retrieved from the Scopus database, complemented by grey literature from leading consulting firms and industry practitioners, the study maps the evolution of quality management from Quality 1.0 to Quality 5.0, identifies the skill sets required of future quality professionals, and outlines key organisational requirements for implementing Quality 5.0. The main contribution is a bottom-up Quality 5.0 implementation framework comprising four stages: Current State Assessment, Roadmap Definition, Implementation, and Monitoring and Improvement. The findings suggest that successfully adopting Quality 5.0 goes beyond technological readiness, as it requires a human-centric organisational culture, strong leadership commitment, and a proactive approach to risk management. The proposed framework offers practical guidance for quality managers and decision-makers seeking to align their organisations with the principles of Industry 5.0, while also contributing to the theoretical consolidation of this emerging field.**

Keywords – Quality 4.0, Quality 5.0, Industry 4.0, Industry 5.0

I. Introduction

The concept of quality in the manufacturing industry has undergone a significant transformation in the recent decades due to technological advancements and shifts in market dynamics (Gunasekaran et al., 2019). For most authors, this evolution is associated with successive industrial revolutions, which have continuously reshaped manufacturing systems and organisational practices, and reflects a broader understanding of quality, extending from defect management to encompassing entire production and service delivery systems (Möller et al., 2022). In the context of the fourth industrial revolution, known as Industry 4.0 (I4.0), widely considered by many authors to be a technological revolution, manufacturing systems have increasingly integrated a range of emerging digital technologies, such as additive manufacturing, augmented reality, autonomous robots, big data and analytics, cloud computing, horizontal and vertical system integration, the Internet of Things (IoT) and simulation (BCG Global, 2026; Hammad et al., 2025). These technologies have had an impact on various industrial and business fields, as well on society at large (Stefanovic et al., 2024). I4.0 affects every manufacturing domain by enabling advanced manufacturing systems that can capture, analyse and deploy large volumes of data, supporting more efficient and intelligent operations. This has led to the development of concepts in the literature such as Maintenance 4.0 (Zonta et al., 2020), Safety 4.0 (Lee et al., 2019), Cybersecurity 4.0 (Culot et al., 2019), Operator 4.0 (Kaasinen et al., 2020) and Quality 4.0 (Zonnenshain and Kenett, 2020). While Quality 4.0 introduced digital technologies to improve traditional quality management practices, Quality 5.0 shifted the focus towards a human-centric approach, integrating human intelligence with technological advancements to enhance the quality management process (Stefanovic et al., 2024).

This manuscript offers an analysis of the transition from Quality 4.0 to Quality 5.0 and examines its implications for quality professionals. The study is based on a literature review of academic publications retrieved from the Scopus database, complemented with grey literature sources, including reports and publications from consulting firms and industry practitioners. Building on these insights, the paper proposes an original, bottom-up quality management framework designed to support organisations in successfully adopting Quality 5.0 practices.

II. Evolution of Quality: From Quality 1.0 to Quality 4.0

Quality 5.0 is the next generation of Quality 4.0. To better understand the implications of Quality 5.0, it is therefore essential to examine how the concept of quality has evolved over time (Arsovski, 2023). From the early stages of manufacturing, characterised by inspection-based approaches, to today's sophisticated, technology-driven quality management systems, each phase reflects a distinct paradigm in quality management, although they were all built upon the previous one(Nguyen and Nguyen, 2025).

In Quality 1.0 (prior to 1890) relied purely on product inspection practices (Akhmatova et al., 2022; Dahlgaard-Park et al., 2018). This stage originated from the craftsmanship model, in which craftsmen produced and sold their own goods locally. Since losing customers due to poor product quality represented a significant risk to their businesses, craftsmen began inspecting goods prior to sale to avoid defects and ensure product quality (ASQ, 2017). However, in the late 19th century, the United States began to move away from traditional European craftsmanship by adopting a new management approach developed by

Frederick W. Taylor. Taylor's approach aimed to increase productivity without hiring more skilled craftsmen. While this led to increased productivity, it also had negative implications for product quality.

Regarding Quality 2.0 (1890-1940), the expansion of mass production resulted in the development of statistical quality control methods. During this period, Walter Shewhart introduced statistical process control (SPC) tools, enabling organisations to monitor production processes and adopt statistical quality control methods (Leššo et al., 2015). Shewhart's work was later expanded by W. Edwards Deming, who developed the Deming Cycle: Plan, Do, Check, Act (PDCA), a new method for continuous improvement. However, at this time, quality management primarily focused on internal operational indicators such as rework and scrap rates rather than broader performance measures such as customer satisfaction (Clark et al., 2013).

During the Quality 3.0 era (1940-1995), which coincided with the Third Industrial Revolution (Industry 3.0), customer satisfaction and standardisation were fully embraced (Zulqarnain et al., 2022). Several influential concepts and methodologies emerged during this period, including the Juran Trilogy, quality circles, Total Quality Management and Six Sigma. During this period, organisations realised that quality must be embedded throughout the entire organisation to generate customer satisfaction and support business growth. Joseph Juran is renowned for his definition of quality as "fitness for use" (Markatos and Mousavi, 2023).

Quality 4.0 is as quality management concept that has attracted significant attention due to increasingly complex customer requirements, intensified global competition and rapid technological advancements (Liu et al., 2023). Quality 4.0 is closely associated with the integration of I4.0's emerging digital technologies, such as real-time and continuous monitoring systems, defect prediction systems, in-line intelligent quality control systems and automated inspection solutions. Aligning I4.0 with Quality 4.0 enables organisations to boost operational efficiency, performance, and innovation. (Nguyen and Nguyen, 2025). Fig. 1 Illustrates the historical evolution of quality management, from Quality 1.0 to Quality 4.0.

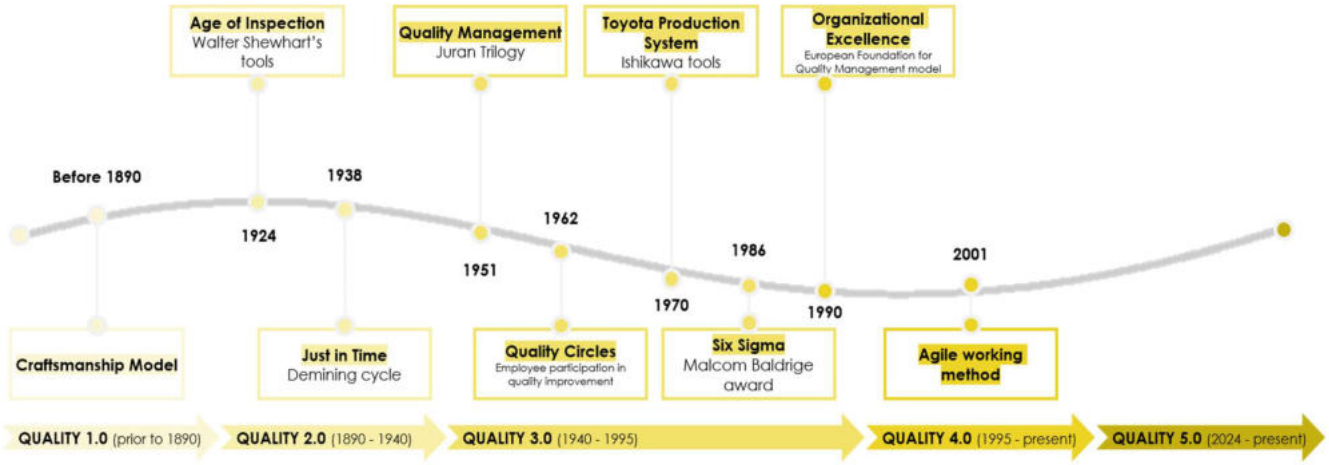


Fig. 1. Evolution of Quality Management from Quality 1.0 to Quality 4.0

III. The Emergence of Quality 5.0

Quality 5.0 is an inevitable consequence of I5.0, the latest wave of the Industrial Revolution, and represents the next phase in the evolution of quality control and assurance(Saraiva, 2023). While Quality 4.0 focused primarily on the role of automation and digital tools, Quality 5.0 goes further by incorporating human-AI collaboration, ethical decision-making and sustainability principles into quality management systems (Kumar et al., 2026; Nguyen and Nguyen, 2025). As the successful implementation of Quality 5.0 requires structured and standardised approaches, many authors emphasise the importance of adopting international standards. For example, Alvares et al. (2025) analysed how ISO 23247 facilitates the implementation of digital twins by enabling real-time monitoring, data integration and improved decision-making in manufacturing systems. It can be assumed that the ability of automation and digital technologies to enhance productivity facilitates the emergence of Quality 5.0 in manufacturing systems (Amangeldiev et al., 2025). Overall, Quality 5.0 increases employee satisfaction by creating a more engaging and human-centred working environment, improving product quality, and aligning with contemporary customer values and regulations. However, the skills required of future Quality 5.0 professionals (Antony et al., 2021; Antony and Sony, 2023; Mahin et al., 2024; Sony et al., 2021), the organisational requirements for implementing Quality 5.0 (Ahmed et al., 2025; Damic et al., 2021; Danurdara and Masatif, 2025; Deliu and Olariu, 2024; ISO, 2015; Liu et al., 2023; Social Accountability International, 2026; Tan et al., 2025), and the associated challenges and mitigation strategies (Navarro Claro and Bayona Soto, 2024) must be understood. In this context, Section IV addresses these aspects by examining the required skill sets, organisational requirements, and key challenges associated with the implementation of Quality 5.0.

IV. Discussion

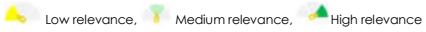
In this section, we examine the skills required for quality professionals to effectively navigate this transformation, outline the main requirements for implementing Quality 5.0 and provide insights into how these elements will influence organisational success, and finally discuss the associated challenges with recommendations to mitigate them.

A. Skill sets for Quality Professionals in Quality 5.0 Era

Quality 5.0 will encompass a shift in the skill profile of quality professionals in terms of hard skills, soft skills and a curriculum for educating future quality management professionals (Antony et al., 2021). Regarding soft skills, future professionals will need to develop strong capabilities in communication, teamwork, leadership and problem-solving, as well as ensure they acquire a deep understanding of business goal (Antony and Sony, 2023). At the same time, hard skills such as data analysis and cybersecurity are almost mandatory (Antony and Sony, 2023). Based on a synthesis of the existing literature, namely Antony et al. (2021), Antony and Sony (2023) and Sony et al. (2021), Table 1 presents the evolution of core skills required for quality professionals across different stages of quality management, from traditional approaches to Quality 4.0 and the emerging Quality 5.0 paradigm. This table is an original contribution of this study.

Table 1. Evolution of Core Competencies for Quality Professionals from Traditional Quality to Quality 5.0

Core Skills for Quality Professionals	Traditional Quality	Quality Q4.0	Quality Q5.0
Soft skills			
Collaboration and teamwork			
Holistic view and expertise cross department			
Change management			
Understanding business goals			
Human-AI collaboration			
Hard skills			
Data science and data analytics			
Digital literacy and emerging technologies			
Cybersecurity awareness			
Knowledge of regulatory compliance			
Continuous improvement methodologies (Lean, Six Sigma)			
Root cause analysis and problem solving			



B. Organisational Requirements for Implementing Quality 5.0

This study builds upon existing quality management frameworks and international standards, particularly ISO 9001:2015 (ISO, 2015), to identify the key organisational requirements for implementing Quality 5.0. While emerging digital technologies are reshaping some of the requirements defined in this standard, an updated version is expected in September 2026, with increased emphasis on quality culture, ethical conduct and the integration of digital technologies and artificial intelligence (BSI, 2026). Nevertheless, the core principles established by ISO 9001 remain highly relevant and continue to provide a solid foundation for the transition towards Quality 5.0.

High quality of products and services

The first key requirement is the delivery of high-quality products and services (Damic et al., 2021). Ensuring good product and service quality is critical for organisational success, as it directly influences customer satisfaction, brand reputation and operational performance. According to ISO (2015), particularly Clause 8.2, clear communication with customers is important in order to identify their needs, as well as ensuring that regulatory requirements and organisational capabilities are properly understood and addressed prior to production and service delivery.

In the context of Quality 5.0, this requirement is supported by two main strategies according to the literature:

i) Real-time quality monitoring and predictive quality analytics: The literature identifies predictive analytics, sensors, and digital feedback loops as the most critical technologies for driving Quality 4.0 (Antony et al., 2021). However, it is important to note that merely implementing monitoring technologies is insufficient if we cannot extract information from these systems.

ii) Design for Quality plays a central role in this context as it ensures that quality is embedded at the earliest stages of product development. In a practical and real perspective, designers determine the number of components in a specific part/product, decide which components should be procured externally, design the rest and specify how they can be manufactured and how the parts should be assembled. When quality considerations are effectively integrated into the design process, manufacturing quality can be achieved more efficiently, reducing defects and the need for expensive inspection activities (Thorat, 2018).

Human-centric work environment

In order to remain competitive, organisations must adopt a holistic approach that aligns technological and process changes with the aim of enhancing the quality of life for employees and customers Investing in employee well-being can enhance job satisfaction and overall performance. This can be achieved by providing and maintaining the infrastructure necessary for processes to operate and products and services to conform to ISO standards (ISO, 2015). This includes ensuring physical conditions such as safety, noise control, and ergonomics. Additionally, SA8000 emphasises social accountability, including non-discrimination, fair working conditions, human interaction and respect for workers' rights (Social Accountability International, 2026). In the context of Quality 5.0, organisations must implement emerging technologies in a way that supports, rather than replaces or marginalises, human capabilities.

Organizational context and capabilities *in-house*

ISO 9001:2015, Clause 7.1 (Resources), states that organisations must determine and maintain the knowledge necessary for operating their processes and achieving conformity of products and services. In the context of Quality 5.0, this knowledge particularly includes digital and technological capabilities (ISO, 2015). This knowledge can be obtained from internal sources, such as intellectual property, experience gained over time, and lessons learned from past successes and failures, as well as from external sources, such as international standards, academic research, conferences, and knowledge acquired from external stakeholders. Based on the literature review, three emerging technologies ensure the integrity and reliability of data across systems: Big data (Deliu and Olariu, 2024), cloud solutions (Tan et al., 2025), and blockchain (Ahmed et al., 2025).

Compliance quality

Compliance quality is a key organisational requirement in the context of Quality 5.0. In ISO 9001:2015, this requirement is closely associated with the responsibilities of top management, particularly in ensuring that customer requirements, as well as applicable statutory and regulatory requirements, are clearly defined, understood and consistently fulfilled. These requirements must be considered throughout the entire lifecycle of products and services, from the design and development phase to the post-delivery stage. Recent research has focused on the concept of regulatory technology (RegTech). The Financial Conduct Authority in the United Kingdom defines RegTech as "the application of cutting-edge technology to facilitate financial regulation and compliance" (Shi et al., 2025). The adoption of emerging digital technologies and a commitment to sustainability has prompted organisations to embrace advanced automation and real-time monitoring. RegTech solutions support organisations in monitoring regulatory changes and automating compliance processes, thereby reducing the risk of human error. Consequently, RegTech enhances compliance capabilities while improving operational efficiency and resilience in increasingly complex regulatory environments (Battanta et al., 2020).

Process quality improvement

Improvement remains a fundamental requirement in the transition towards Quality 5.0. As one of the core principles of ISO 9001, it gives organisations a structured way to improve their processes develop new products and services (ISO, 2015). ISO 9001:2015 identifies correction, corrective action, continual improvement, breakthrough change, innovation, and re-organization as forms of improvement. In the context of Quality 5.0, however, process quality improvement goes beyond traditional process optimisation, incorporating organisational learning, employee engagement and collaboration between humans and intelligent systems (Liu et al., 2023). It also emphasises integrating environmental and sustainability considerations into production and service delivery processes. By embedding these principles in operational plans and decision-making processes, organisations can align their processes with sustainability goals, including ESG goals, while maintaining high quality and efficiency levels (Saidinova et al., 2025)

Risk-based thinking

ISO 9001:2015 emphasises the importance of risk-based thinking in the planning and operation of quality management systems. In particular, Clause 6.1 ("Actions to address risks and opportunities") required organisations to consider internal and external issues, as well as the needs and expectations of relevant stakeholders, in order to identify risks and opportunities that may affect the achievement of intended outcomes (ISO, 2015). Enterprises that adopt structured risk identification and assessment frameworks are better positioned to prevent and respond effectively to potential disruptions and uncertainties (Hsu et al., 2024). In the context of Quality 5.0, risk-based thinking evolves beyond traditional approaches by incorporating data-driven and predictive capabilities enabled by emerging digital technologies. These capabilities, initially introduced in Quality 4.0, are further enhanced through the integration of human expertise and advanced analytics, supporting more proactive and informed decision-making (Kumar et al., 2026).

Fig.2 summarises the six key organisational requirements for implementing Quality 5.0.

Fig.2. Key Organisational Requirements for Implementing Quality 5.0

C. Challenges in implementing Quality 5.0 and a proposed approach for successful implementation

As with any paradigm shift, the introduction of Quality 5.0 also brings its own organisational and technological challenges. Based on the literature review, the main challenges identified are as follows:

- Cybersecurity and data privacy concerns (Antony et al., 2021; Frick and Grudowski, 2023);
- Digital skills gaps (Frick and Grudowski, 2023; Mahin et al., 2024) and lack of training (Antony et al., 2021);
- Risk of technology dependency, which increases the vulnerability of organisations to system failures (Frick and Grudowski, 2023);
- Legacy systems and outdated infrastructure (Arsovski, 2023);
- Limited Quality 4.0 maturity(Antony et al., 2021);
- Lack of leadership support and strategic vision, which constrains resource allocation and decision-making processes (Antony et al., 2021; Mahin et al., 2024);
- Employee resistance to change due to poor change culture (Antony et al., 2021);
- High initial investment and an unclear return on investment (ROI) (Antony et al., 2021; Mahin et al., 2024).

Addressing these challenges is key to the successful implementation of Quality 5.0. To help leaders overcome the challenges posed by this contemporary quality paradigm, this paper proposes a bottom-up approach as its main contribution to ensure the successful adoption of Quality 5.0. This approach, represented in Fig.3, combines traditional quality management methods with data strategy insights from leading consulting firms, enhancing operational efficiency, driving innovation and achieving superior, sustainable quality outcomes (Accenture, 2026; Datum Labs, 2024; Spec India, 2025).

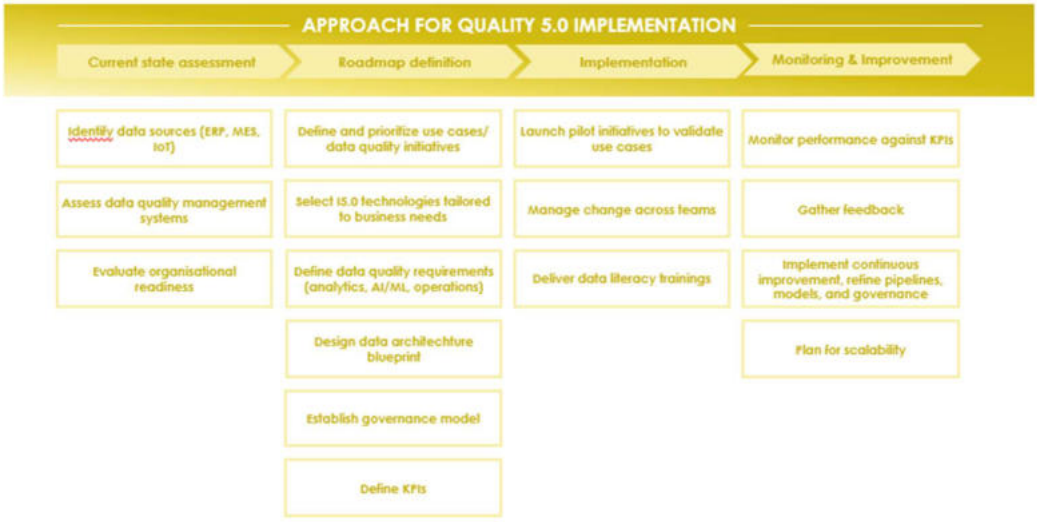


Fig.3. Approach For Quality 5.0 Implementation

Current State Assessment

This stage involves an evaluation of the existing quality management systems through stakeholder interviews, site visits and data analysis, establishing the baseline against which future progress will be measured.

Roadmap definition

In this stage, the focus shifts to imagining where the organisation needs to go. This involves establishing target states for the quality management system across a maturity spectrum ranging from foundational to advanced and best-in-class. Initiatives are then developed to bridge the gap between the current and intended future state. We also suggest organisations to reflect on how existing I4.0 technologies, such as IoT, AI and cloud computing, can be enhanced by integrating human, environmental and societal considerations, in line with the core principles of I5.0.

Implementation

Successful deployment of new quality management principles needs effective change management. This includes the design of communication strategies, the provision of targeted training and support, and the active garnering employee buy-in through participation and feedback mechanisms. A bottom-up approach is emphasised throughout, ensuring that staff at all levels feel engaged in the transition rather than subject to it. When employees understand the rationale behind the changes and have a genuine role in shaping them, resistance decreases and the likelihood of lasting adoption increases significantly

Monitor and improvement

This stage closes the loop by tracking how well the new quality management principles are actually performing in practice, using a combination of traditional process metrics and human-centric indicators such as employee engagement and customer satisfaction. When gaps emerge, corrective actions are applied before they become entrenched. What distinguishes this stage in a Quality 5.0 context is the emphasis on organisational learning over mere compliance. Feedback flows in both directions, from the shop floor to strategic decisions and back again, ensuring the roadmap stays grounded in operational reality. In this way, monitoring becomes less a control mechanism and more a continuous conversation between data, people and purpose.

V. CONCLUSION

Although several studies in the literature report positive effect in the transitioning from Quality 4.0 to Quality 5.0, there are few that address the main challenges of implementing Quality 5.0. While Quality 4.0 enhanced traditional quality management practices through the adoption of emerging digital technologies such as AI, big data, analytics, and cloud computing, Quality 5.0 builds upon these foundations by following fundamental goals such as "Leave no Person Behind", "Empower Users through Good Digital Identities", "Make Business Work for People" and "Keep Everyone Safe and Secure".

This paper offers both theoretical and managerial contributions to the emerging concept of Quality 5.0. From a theoretical perspective, the proposed framework suggests that implementing Quality 5.0 is not a one-step process, but rather a continuous and multi-dimensional transformation that integrates technological, organisational, and human-centric requirements. The framework extends existing Quality 4.0 models by explicitly incorporating human-AI collaboration, ethical decision-making and sustainability as core design principles, and the six organisational requirements identified (high quality of products and services, human-centric work environment, organisational context and capabilities, compliance quality, process quality improvement, and risk-based thinking) provide a structured basis for future empirical work. From a managerial standpoint, the framework provides guidance for organisations seeking to transition toward a Quality 5.0 approach.

Although this study is limited to a literature review based on the Scopus database, the proposed framework is designed to be applicable across different manufacturing contexts. Future research should prioritise empirical validation in settings where the framework's assumptions can be properly tested, whether through case studies in companies at different Quality 4.0 maturity levels, or through broader surveys capturing variation in organisational size and sector. Studies comparing organisations at early versus more advanced stages of digital transformation would be particularly valuable in clarifying which stages of the framework deliver the most impact depending on where an organisation starts its journey.

References

Accenture, 2026. What is data and why is it important? [WWW Document]. URL https://www.accenture.com/us-en/insights/what-is-data (accessed 3.20.26).

Ahmed, F., Zhou, T., Bilal, H., Ul Islam, F., Ullah, R., Vasilakos, A.V., 2025. Enhancing Healthcare Data Integrity and Access Control Using Blockchain and Industry 5.0. IEEE Internet Things J. 12, 43630–43643. https://doi.org/10.1109/JIOT.2025.3598320

Akhmatova, M.-S., Deniskina, A., Akhmatova, D.-M., Prykina, L., 2022. Integrating quality management systems (TQM) in the digital age of intelligent transportation systems industry 4.0. Transp. Res. Procedia, X International Scientific Siberian Transport Forum — TransSiberia 2022 63, 1512–1520. https://doi.org/10.1016/j.trpro.2022.06.163

Alvares, A.J., Rodriguez, E., Figueroa, B., 2025. Digital-Twin-Enabled Process Monitoring for a Robotic Additive Manufacturing Cell Using Wire-Based Laser Metal Deposition. Processes 13. https://doi.org/10.3390/pr13082335

Amangeldiev, D., Kozhoshev, A., Shadybekov, K.B., Kadyrov, A.A., Bekov, A.T., 2025. SUSTAINABLE INNOVATIONS IN MANAGEMENT INFORMATION SYSTEMS FOR ENSURING QUALITY 5.0 IN LOCAL GOVERNANCE OF SMART CITIES. Proc. Eng. Sci. 7, 177–188. https://doi.org/10.24874/PES07.01A.009

Antony, J., McDermott, O., Sony, M., 2021. Quality 4.0 conceptualisation and theoretical understanding: a global exploratory qualitative study. TQM J. 34, 1169–1188. https://doi.org/10.1108/TQM-07-2021-0215

Antony, J., Sony, M., 2023. An Empirical Study into Qualifications and Skills of Quality Management Practitioners in Contemporary Organizations: Results from a Global Survey and Agenda for Future Research. IEEE Trans. Eng. Manag. 70, 478–494. https://doi.org/10.1109/TEM.2021.3050460

Arsovski, S., 2023. QUALITY 5.0: FROM CHALLENGES TO REALITY. J. Innov. Bus. Ind. 1, 13–21. https://doi.org/10.61552/JIBI.2023.01.002
ASQ, 2017. History of Quality - Quality Management History | ASQ [WWW Document]. URL https://asq.org/quality-resources/history-of-quality (accessed 3.16.26).
Battanta, L., Giorgino, M., Grassi, L., Lanfranchi, D., 2020. Regtech: Case studies of cooperation with banks in italy, in: Proc. Eur. Conf. Innov. Entrepren., ECIE. Presented at the Proceedings of the European Conference on Innovation and Entrepreneurship, ECIE, Academic Conferences and Publishing International Limited, pp. 112–119. https://doi.org/10.34190/EIE.20.143
BCG Global, 2026. Industry 4.0 [WWW Document]. BCG Glob. URL https://www.bcg.com/capabilities/manufacturing/industry-4.0 (accessed 3.15.26).
BSI, 2026. ISO 9001:2026 – Principais Mudanças e Orientações [WWW Document]. BSI. URL https://www.bsigroup.com/pt-BR/products-and-services/standards-services/iso-9001-2026-key-changes-and-guidance/ (accessed 3.18.26).
Clark, D.M., Silvester, K., Knowles, S., 2013. Lean management systems: creating a culture of continuous quality improvement. J. Clin. Pathol. 66, 638–643. https://doi.org/10.1136/jclinpath-2013-201553
Culot, G., Fattori, F., Podrecca, M., Sartor, M., 2019. Addressing Industry 4.0 Cybersecurity Challenges. IEEE Eng. Manag. Rev. 47, 79–86. https://doi.org/10.1109/EMR.2019.2927559
Dahlgaard-Park, S.M., Reyes, L., Chen, C.-K., 2018. The evolution and convergence of total quality management and management theories. Total Qual. Manag. Bus. Excell. 29, 1108–1128. https://doi.org/10.1080/14783363.2018.1486556
Damic, M., Naletina, D., Buntic, L., 2021. REVISITING THE RELATIONSHIP BETWEEN ORGANIZATIONAL INNOVATIVENESS AND ISO 9001. Int. J. Qual. Res. 15, 909–922. https://doi.org/10.24874/IJQR15.03-14
Danurdara, A.B., Masatif, A., 2025. Assessing the customer experience quality and customer loyalty: The mediating role of customer satisfaction. Innov. Mark. 21, 248–259. https://doi.org/10.21511/im.21(2).2025.20
Datum Labs, 2024. Blog | Data Strategy Consulting Guide: Build a Future-Ready Business [WWW Document]. URL https://www.datumlabs.io/resources/data-strategy-consulting-guide-build-a-future-ready-business (accessed 3.20.26).
Deliu, D., Olariu, A., 2024. The Role of Artificial Intelligence and Big Data Analytics in Shaping the Future of Professions in Industry 6.0: Perspectives from an Emerging Market. Electron. Switz. 13. https://doi.org/10.3390/electronics13244983
Frick, J., Grudowski, P., 2023. Quality 5.0: A Paradigm Shift Towards Proactive Quality Control in Industry 5.0. Int. J. Bus. Adm. 14, 51. https://doi.org/10.5430/ijba.v14n2p51
Gunasekaran, A., Subramanian, N., Ngai, W.T.E., 2019. Quality management in the 21st century enterprises: Research pathway towards Industry 4.0. Int. J. Prod. Econ. 207, 125–129. https://doi.org/10.1016/j.ijpe.2018.09.005
Hammad, M.Y., Rahamaddulla, S.R., Muhamad Tamyez, P.F., Fauzi, M.A., 2025. From Industry 4.0 to 5.0: leveraging AI and IoT for sustainable and human-centric operations. Int. J. Ind. Eng. Oper. Manag. 1–19. https://doi.org/10.1108/IJIEOM-04-2025-0070
Hsu, C.-H., Wu, J.-Z., Zhang, T.-Y., Chen, J.-Y., 2024. Deploying Industry 5.0 drivers to enhance sustainable supply chain risk resilience. Int. J. Sustain. Eng. 17, 211–238. https://doi.org/10.1080/19397038.2024.2327381
ISO, 2015. ISO 9001:2015 [WWW Document]. ISO. URL https://www.iso.org/standard/62085.html (accessed 3.18.26).
Kaasinen, E., Schmalfuß, F., Özturk, C., Aromaa, S., Boubekeur, M., Heilala, J., Heikkilä, P., Kuula, T., Liinasuo, M., Mach, S., Mehta, R., Petäjä, E., Walter, T., 2020. Empowering and engaging industrial workers with Operator 4.0 solutions. Comput. Ind. Eng. 139. https://doi.org/10.1016/j.cie.2019.01.052
Kumar, V., Singh, A., Sharma, N., Verma, P., 2026. Integrating quality 5.0 approaches for the future of quality management. TQM J. 1–35. https://doi.org/10.1108/TQM-04-2025-0210
Lee, J., Cameron, I., Hassall, M., 2019. Improving process safety: What roles for digitalization and industry 4.0? Process Saf. Environ. Prot. 132, 325–339. https://doi.org/10.1016/j.psep.2019.10.021
Leššo, I., Horovčák, P., Stehlíková, B., Gašpárová, Z., Flegner, P., 2015. The multidimensional Shewhart control chart for multicriteria quality management, in: Proceedings of the 2015 16th International Carpathian Control Conference (ICCC). Presented at the the 2015 16th International Carpathian Control Conference (ICCC), pp. 286–289. https://doi.org/10.1109/CarpathianCC.2015.7145090
Liu, H.-C., Liu, R., Gu, X., Yang, M., 2023. From total quality management to Quality 4.0: A systematic literature review and future research agenda. Front. Eng. Manag. 10, 191–205. https://doi.org/10.1007/s42524-022-0243-z
Mahin, M., Kadasah, N., Alsabban, A., Albliwi, S., 2024. Exploring the landscape of quality 4.0: a comprehensive review of its benefits, challenges, and critical success factors. Prod. Manuf. Res. 12. https://doi.org/10.1080/21693277.2024.2373739
Markatos, N.G., Mousavi, A., 2023. Manufacturing quality assessment in the industry 4.0 era: a review. Total Qual. Manag. Bus. Excell. 34, 1655–1681. https://doi.org/10.1080/14783363.2023.2194524
Möller, D.P.F., Vakilzadian, H., Haas, R.E., 2022. From Industry 4.0 towards Industry 5.0, in: 2022 IEEE International Conference on Electro Information Technology (eIT). Presented at the 2022 IEEE International Conference on Electro Information Technology (eIT), pp. 61–68. https://doi.org/10.1109/eIT53891.2022.9813831
Navarro Claro, G.T., Bayona Soto, J.A., 2024. From the Total Quality Paradigm to Quality 5.0: A literature review for the manufacturing sector. Rev. Cienc. Soc. 30, 551–566. https://doi.org/10.31876/rcs.v30i.42859
Nguyen, K.H., Nguyen, T.M.H., 2025. Quality 5.0: Sustainable Quality Management System for Future. 品質學報 32, 343–373. https://doi.org/10.6220/joq.202510_32(5).0002
Saidinova, M.T., Mavlyanova, D.M., Galoyan, D.R., Pecherskaya, O.A., Nabokikh, A.A., 2025. ENERGY EFFICIENCY AS A CRITERION OF QUALITY 5.0: MANAGEMENT THROUGH SUSTAINABLE INNOVATIONS. Proc. Eng. Sci. 7, 2283–2292. https://doi.org/10.24874/PES07.04.014
Saraiva, P., 2023. QUALITY 5.0 IN TURBULENT TIMES. Qual. Prog. 56, 18–25.
Shi, Z., Xia, Y., He, L., Sun, N., Zheng, Q., 2025. Can internal regulatory technology (RegTech) mitigate bank credit risk? Evidence from the banking sector in China. Res. Int. Bus. Finance 75. https://doi.org/10.1016/j.ribaf.2025.102780
Social Accountability International, 2026. SA8000:2026 Standard. SAI. URL https://sa-intl.org/resources/sa8000-standard/ (accessed 3.19.26).
Sony, M., Antony, J., Douglas, J.A., McDermott, O., 2021. Motivations, barriers and readiness factors for Quality 4.0 implementation: an exploratory study. TQM J. 33, 1502–1515. https://doi.org/10.1108/TQM-11-2020-0272
Spec India, 2025. Transform Your Business with Expert Data Strategy Consulting [WWW Document]. URL https://www.spec-india.com/blog/data-strategy-consulting (accessed 3.20.26).
Stefanovic, M., Wawak, S., Popkova, E., Eric, M., Mitrovic, S., Djordjevic, A., Zahar Djordjevic, M., 2024. FROM QUALITY 4.0 TO QUALITY 5.0 - THE TRANSITION ROADMAP. Int. J. Qual. Res. 18, 1199–1210. https://doi.org/10.24874/IJQR18.04-18
Tan, H., Eiza, M.H., Moh, S., Sakurai, K., 2025. Data security for cloud-fog-assisted Industrial Internet of Things (IIoT) in future Industry 5.0. Internet Things Neth. 34. https://doi.org/10.1016/j.iot.2025.101765
Thorat, S., 2018. Design For Quality - Concept, Benefits of Design for Quality (DFQ). URL https://learnmech.com/design-quality-concept-benefits-design-quality-dfq/ (accessed 3.18.26).
Zonnenshain, A., Kenett, R.S., 2020. Quality 4.0—the challenging future of quality engineering. Qual. Eng. 32, 614–626. https://doi.org/10.1080/08982112.2019.1706744
Zonta, T., da Costa, C.A., da Rosa Righi, R., de Lima, M.J., da Trindade, E.S., Li, G.P., 2020. Predictive maintenance in the Industry 4.0: A systematic literature review. Comput. Ind. Eng. 150. https://doi.org/10.1016/j.cie.2020.106889
Zulqarnain, A., Wasif, M., Iqbal, S.A., 2022. Developing a Quality 4.0 Implementation Framework and Evaluating the Maturity Levels of Industries in Developing Countries. Sustainability 14. https://doi.org/10.3390/su141811298

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Reusing of Metallic Pieces as a Case Study for a Continuous Sustainability Improvement Process in Luxury Leather Goods Industry

Micael Gonçalves[1], Ana Lobarinhas[1], Sérgio Sousa[1], Senhorinha Teixeira[1] and Pedro Domingues[1]

[1] ALGORITMI Centre, Department of Production and Systems, School of Engineering, University of Minho, Guimarães, Portugal; micael.goncalves@dps.uminho.pt

***Abstract* — Balancing luxury quality and sustainability in leather goods manufacturing is challenging, particularly in managing high-value components. This paper examines how quality management (QM) can enable sustainability through the controlled reuse of metallic pieces. A case study was conducted at ATEPELI, a Portuguese manufacturer of luxury leather components for leading brands. Analysis of seven months of production data (September 2023–March 2024) revealed that, on average, 8,401 metallic pieces per month—out of 11,363 returned—were fully compliant but destroyed due to traceability limitations, representing a potential monthly saving of €25,690. Process mapping and workflow analysis identified critical inefficiencies in operational and customer quality flows. A targeted improvement proposal was developed to enable the reuse of metallic components while maintaining strict luxury quality standards. The findings demonstrate that integrating sustainability objectives into QM systems can substantially reduce material waste, optimize resource use, and generate measurable financial benefits. This study provides practical evidence that continuous improvement practices can reconcile premium quality with sustainability in complex manufacturing environments.**

Keywords – Luxury leather goods, Quality Management, Sustainability, Material Reuse.

I. Introduction

The luxury leather goods industry is characterized by high-quality craftsmanship, exclusivity, and stringent quality standards. Maintaining these standards is critical to ensuring customer satisfaction and preserving brand reputation. Unlike mass-market production, luxury manufacturing combines artisanal processes with structured industrial operations, requiring adaptive quality management (QM) strategies to maintain both product uniqueness and consistency (Holmqvist, Wirtz, & Fritze, 2022; Kai, 2021).

QM in luxury manufacturing extends beyond defect prevention and process standardization; it encompasses systematic control of materials, production flows, and supplier coordination to ensure that every component meets both technical and aesthetic requirements. Frameworks such as Total Quality Management (TQM) and Supply Chain Quality Management (SCQM) emphasize employee involvement, process visibility, and continuous improvement as key enablers for achieving these objectives. In this context, effective QM is not only a tool for preserving product excellence but also a foundation for operational efficiency and resource optimization (Brun et al., 2008; Brun & Moretto, 2013; Foster, Wallin, & Ogden, 2011; Gonçalves, Martins, Pereira, & Sousa, 2024; Lobarinhas, Gonçalves, & Teixeira, 2024; Robinson & Malhotra, 2005; Zheng, 2023).

Sustainability has emerged as a critical dimension in manufacturing, focusing on minimizing resource consumption, reducing waste, and mitigating environmental impacts. In the luxury sector, a tension exists between sustainability and traditional luxury values: companies prioritize product quality, supplier consistency, and exclusivity, while sustainability emphasizes efficient resource use and waste reduction. Addressing this tension requires integrating sustainability objectives into QM systems, enabling companies to manage resources responsibly without compromising quality or brand value (Bentahar & Benzidia, 2018; Bhattacharya & Dey, 2020; Chiu, Choi, Dai, Shen, & Zheng, 2018; Choi & Liu, 2019; Feng, Tong, & Zhu, 2020; Kapferer & Michaut, 2014).

A key sustainability challenge in luxury leather goods production lies in the management of metallic components. At ATEPELI, a Portuguese manufacturer of luxury leather components founded in 2011, this challenge is particularly evident. The company has three ateliers (Ponte de Lima, Penafiel, and Santa Maria da Feira) and employs over 1,100 staff to produce high-quality components such as bag handles, straps, and wallet interiors for some of the world's most prestigious luxury brands. ATEPELI combines artisanal craftsmanship with structured industrial processes, creating a hybrid production environment that demands adaptive QM strategies to preserve product uniqueness while ensuring consistency and compliance with strict customer requirements (Gonçalves, Lobarinhas, Domingues, Sousa, & Teixeira, 2026; Gonçalves, Lobarinhas, & Teixeira, 2025; Lobarinhas et al., 2024).

Analysis of metallic component handling at ATEPELI over seven months (September 2023–March 2024) revealed an average of 11,363 metallic pieces returned per month, of which 8,401 were fully compliant and could have been reused. The destruction of these reusable components resulted in potential financial losses of approximately €25,690 per month, highlighting the operational and environmental impact of traceability gaps. This case illustrates the practical importance of QM systems in supporting sustainability objectives: by identifying inefficiencies and enabling the controlled reuse of materials, QM practices can simultaneously reduce waste, optimize supply chains, and maintain luxury quality standards.

This paper presents a detailed case study of ATEPELI, examining current QM procedures, operational and customer quality flows, and parts return processes. This research is part of a broader project aimed at reviewing and optimizing quality management procedures within ATEPELI in preparation for the implementation of the SAP enterprise system (Gonçalves et al., 2026). By mapping these processes and analyzing material flows, the study demonstrates how continuous improvement

practices, grounded in TQM and SCQM principles, can merge luxury quality requirements with sustainability goals. The findings provide practical evidence that integrating sustainability into QM frameworks is not only feasible but also beneficial, enabling luxury manufacturers to reduce waste, optimize resource use, and preserve the premium quality expected in their products.

This paper is structured into five sections. Section 2 presents the research methodology. Section 3 describes the current quality management processes at ATEPELI. Section 4 provides a critical analysis of the main deficiencies and improvement opportunities identified. Section 5 concludes the paper with the final remarks and key findings.

II. Methodology

This study adopts a qualitative case study methodology to examine quality management processes and identify sustainability improvement opportunities related to the reuse of metallic components in the luxury leather goods industry. Case study research is suitable for analyzing real organizational processes and operational challenges within their practical context (Teegavarapu, Summers, & Mocko, 2008). The research was conducted at ATEPELI, a Portuguese manufacturer of luxury leather goods components integrated into an international luxury supply chain.

Data collection combined multiple sources of information, including direct observation of production and quality control activities, analysis of internal process documentation, operational records, quality reports, and historical data related to metallic component management. This combination of sources supported the detailed analysis of workflows, traceability limitations, and improvement opportunities. Historical data regarding metallic pieces sent for destruction between September 2023 and March 2024 were analyzed to quantify material waste and estimate its associated financial impact.

To support the analysis, process mapping techniques based on Business Process Model and Notation (BPMN) were applied to document operational quality, customer quality, and parts return workflows. Particular attention was given to identifying traceability limitations, process inefficiencies, and decision points affecting the reuse of metallic components.

Based on the current-state assessment, an improvement proposal was developed in the context of the ongoing SAP implementation project. The proposal focuses on enhancing material traceability and enabling the controlled reuse of metallic pieces while maintaining compliance with luxury quality standards.

The methodological approach is aligned with TQM and SCQM principles, emphasizing continuous improvement, process visibility, waste reduction, and resource optimization (Rahman & Bullock, 2005).

III. Current State Analysis Of Quality Management Procedures

In this section, the current operational framework of the Quality Department is examined. The analysis focuses on three main areas: operational quality, customer quality, and parts returns. The processes within each area are described, highlighting existing deficiencies and identifying opportunities for improvement. To facilitate clarity and ease of interpretation, the quality processes are represented using Business Process Model and Notation (BPMN) flowcharts.

A. Operational Quality Process

The Operational Quality area focuses on maintaining production standards and minimizing internal defects throughout the manufacturing process. Operators are trained in self-inspection and are instructed to proceed only when components meet predefined quality criteria. Final inspections at the end of the production line identify non-conforming parts (KOs). Recoverable components are returned to production for rework, while unrecoverable items are classified as final KOs and either destroyed or returned to the customer (Fig. 1).

A key challenge in this process is the reliance on an outdated and overly detailed defect catalogue, comprising 82 classifications. This complexity often leads to misclassification and reduces the accuracy of internal non-quality indicators, highlighting the need for a streamlined and updated defect taxonomy.

After the second inspection, parts deemed OK are returned to the normal production flow. Recoverable KOs are sent back for repair, while unrecoverable parts are handled according to supply flow, either returned to the customer with the rest of the order or sent for destruction (Gonçalves et al., 2025).

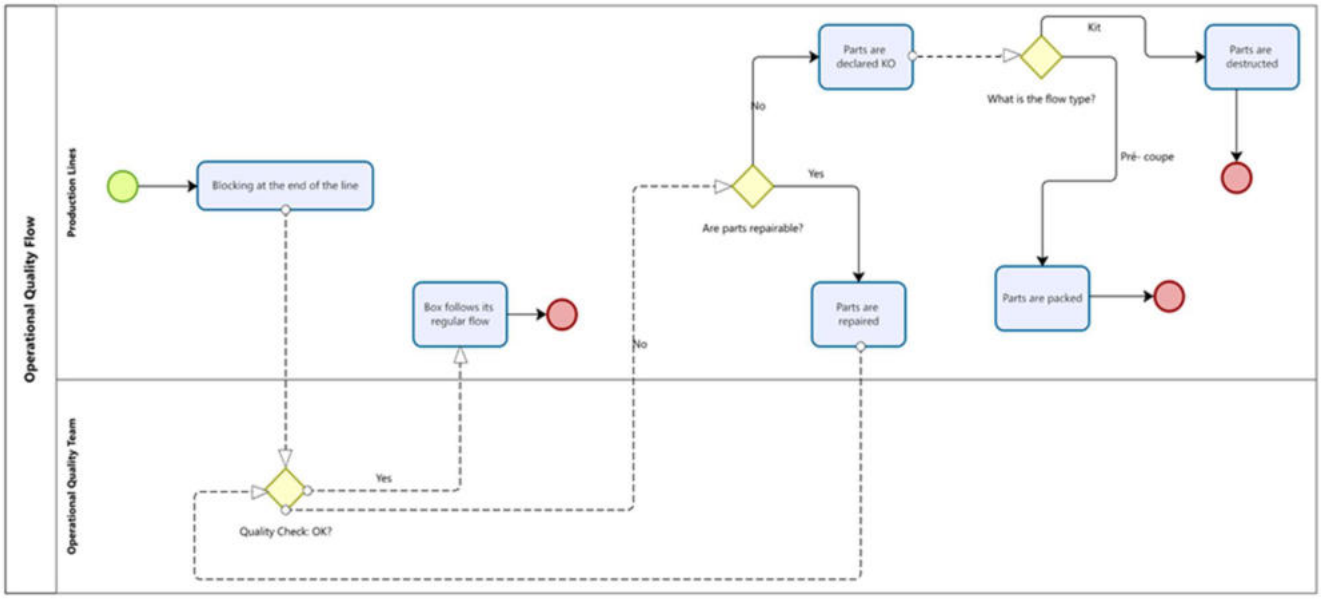


Fig. 1. Flowchart of current operational quality process (Gonçalves et al., 2025)

B. Customer (Returns) Quality Process

The Customer Quality area is directly linked to customer perceptions of product quality. Effective management of returns requires a detailed analysis of items returned due to quality non-conformities. Within ATEPELI's group, a strict 21-day lead time for processing returns is mandatory.

The Customer Quality process depends on the Raw Materials and Supply Strategy. Producing leather goods requires materials such as fabrics, leather, metallic pieces ("Pmets"), and consumables like paints, glue, and threads. ATEPELI acquires these materials through two supply strategies: in the "kit" flow, the company purchases raw materials from external suppliers; in the "pré-coupe" flow, customers provide pre-cut leather and metallic components.

Upon arrival at the "Quality Wall" (Fig. 2), return management technicians inspect the items and record details in the internal tracking system. Using the software, they identify the original production source to route components to the appropriate production lines for repair. For both internal and external customers, a credit note may be issued before items are sorted and sent for evaluation. Repairable components are promptly fixed and repackaged for dispatch, while irreparable items are marked as KOs and returned to the "Quality Wall." Non-compliant items are either sent for financial approval and destruction or returned to the customer labelled as KOs (Gonçalves et al., 2025).

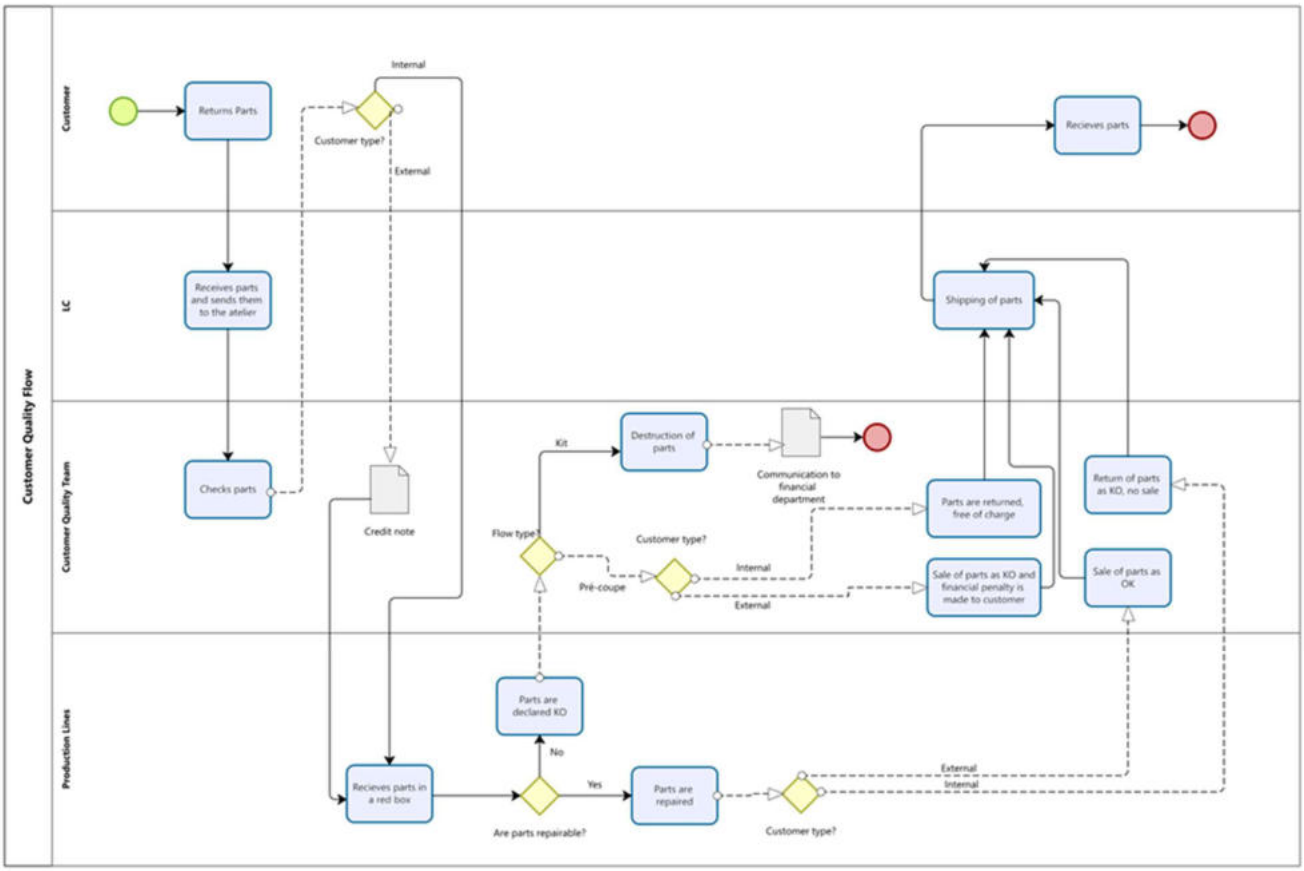


Fig. 2. Flowchart of current customer quality process (Gonçalves et al., 2025)

C. Parts Return Process

In the operational and customer quality flows of ATEPELI, when a part is deemed KO (unsatisfactory), whether within production lines or due to unfeasibility of rework upon return, it follows distinct pathways outlined in Fig. 1 (Operational Quality Flow) and Fig. 2 (Customer Quality Flow). Parts within the *"pré-coupe"* flow are returned to the customer along with the rest of the order, as raw materials were supplied by them. Conversely, components in the *"kit"* flow undergo destruction.

The destruction process involves manual separation by an employee using scissors to disassemble leather and textile materials from metallic pieces. The leather and textiles are then destroyed, while metallic pieces are prepared for return to suppliers for melting and recycling. Suppliers melt the metal and reintroduce it into their manufacturing processes to produce new parts.

A challenge arises from returning and recycling components procured consistently from suppliers. Instead of promptly reintegrating these components into warehouse inventory, they are destructed, imposing unnecessary financial burden. Group standards mandate that a production batch cannot contain metallic pieces from different suppliers, even if identical to the human eye. Thus, the primary obstacle to direct reuse of metallic pieces is the inability to trace each part's supplier throughout the process.

Over seven months, from September 2023 to March 2024, an analysis was conducted on the number of metallic pieces returned to suppliers for destruction. Before being sent for destruction and recycling, the returned metallic components were individually analyzed to assess their physical and aesthetic condition. Based on this evaluation, each piece was classified either as compliant and potentially suitable for reintegration into the production process (OK) or as non-compliant and unsuitable for reuse (KO).

As detailed in Table , the quantities of satisfactory (OK) and unsatisfactory (KO) components were recorded. Findings revealed an average of 11,363 metallic pieces returned per month, of which 8,401 were classified as OK and considered technically suitable for reintroduction into stock for production provisioning in an ideal traceability scenario. The average of 8,401 metallic pieces identified as OK per month refers to components that were technically compliant for reuse but were destroyed due exclusively to traceability limitations, rather than due to quality defects. This would represent an average reduction of 74% in wasted metallic pieces.

TABLE I. AMOUNT OF OK AND KO METALLIC PIECES SET FOR DESTRUCTION BETWEEN SEPTEMBER 2023 AND MARCH 2024

Returned Metallic Pieces			
Month	***OK***	***KO***	***Total***
Sep 23	3870	120	3990
Oct 23	5130	851	5981
Nov 23	7265	536	7801
Dec 23	5407	331	5738
Jan 24	20241	6041	26282
Feb 24	9128	9879	19007
Mar 24	7769	2971	10740
Total	58810	20729	79539
Average Per Month	8401	2961	11363

From this analysis, the financial implications of this issue were derived (see Table). The cumulative financial cost of the parts deemed OK over these seven months amounts to € 179,829, resulting in an average potential financial saving of € 25,690 per month, in the ideal scenario in which these components were retained indoors rather than being discarded.

TABLE II. COST OF OK METALLIC PIECES SET FOR DESTRUCTION FROM SEPTEMBER 2023 TO MARCH 2024

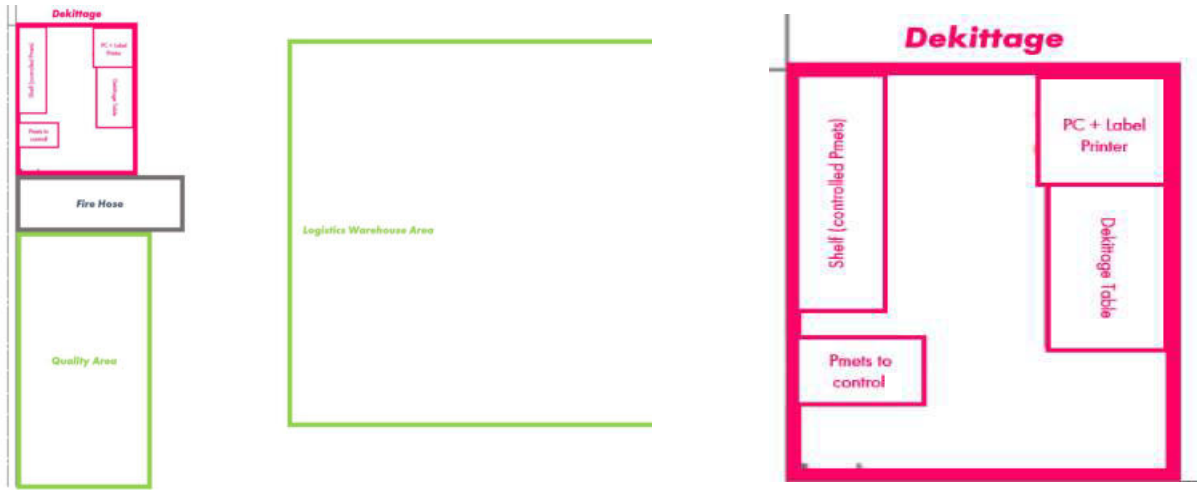
Month	**Cost of OK Pieces (€)**
Sep 23	10 019
Oct 23	19 856
Nov 23	29 688
Dec 23	17 122
Jan 24	55 434
Feb 24	20 218
Mar 24	27 492
Total	179 829
Average Per Month	25 690

It's worth noting that these data only relate to the parts identified during the manufacturing process (Operational Quality Flow). The financial impact would be even greater if the metallic pieces designated for destruction from the Customer Quality Flow were also considered.

IV. PROPOSAL PROCESS FOR REUSING METALLIC PIECES

A potential solution to the challenge of metallic pieces set for destruction arises with the implementation of SAP. The ongoing SAP rollout at ATEPELI aims to digitalize and integrate key operational and quality processes across the organization, replacing legacy manual systems with a unified enterprise platform. By capturing detailed material and production data at each stage of the workflow, SAP enhances traceability of components from receipt through final assembly and returns, providing real-time visibility into supplier provenance, batch status, and quality outcomes.

This comprehensive tracking allows the identification of the supplier for each metallic piece in every produced part, enabling controlled reuse and reducing unnecessary waste. To support this, a new process and station, named "*dekittage*," has been introduced. Its layout and location (Fig. 3(a) and Fig. 3(b)) were carefully designed to optimize physical flow. Situated in the Logistics and Cutting building alongside the logistics warehouse, the *dekittage* station enables the direct storage of metallic pieces after processing, ensuring efficient handling and preparation for reintegration into production.



(a) (b)

Fig. 3. Layout (a) Location of the "dekittage" station in the Logistics and Cutting building; (b) "Dekittage" station layout

Employees responsible for this process receive comprehensive training from the Material Quality team to ensure accurate assessment of the compliance of metallic pieces with quality standards. The *dekittage* process flow is depicted in Fig. 4**Error! Reference source not found.**, representing a step towards improving efficiency and sustainability in metallic piece management within ATEPELI's operations.

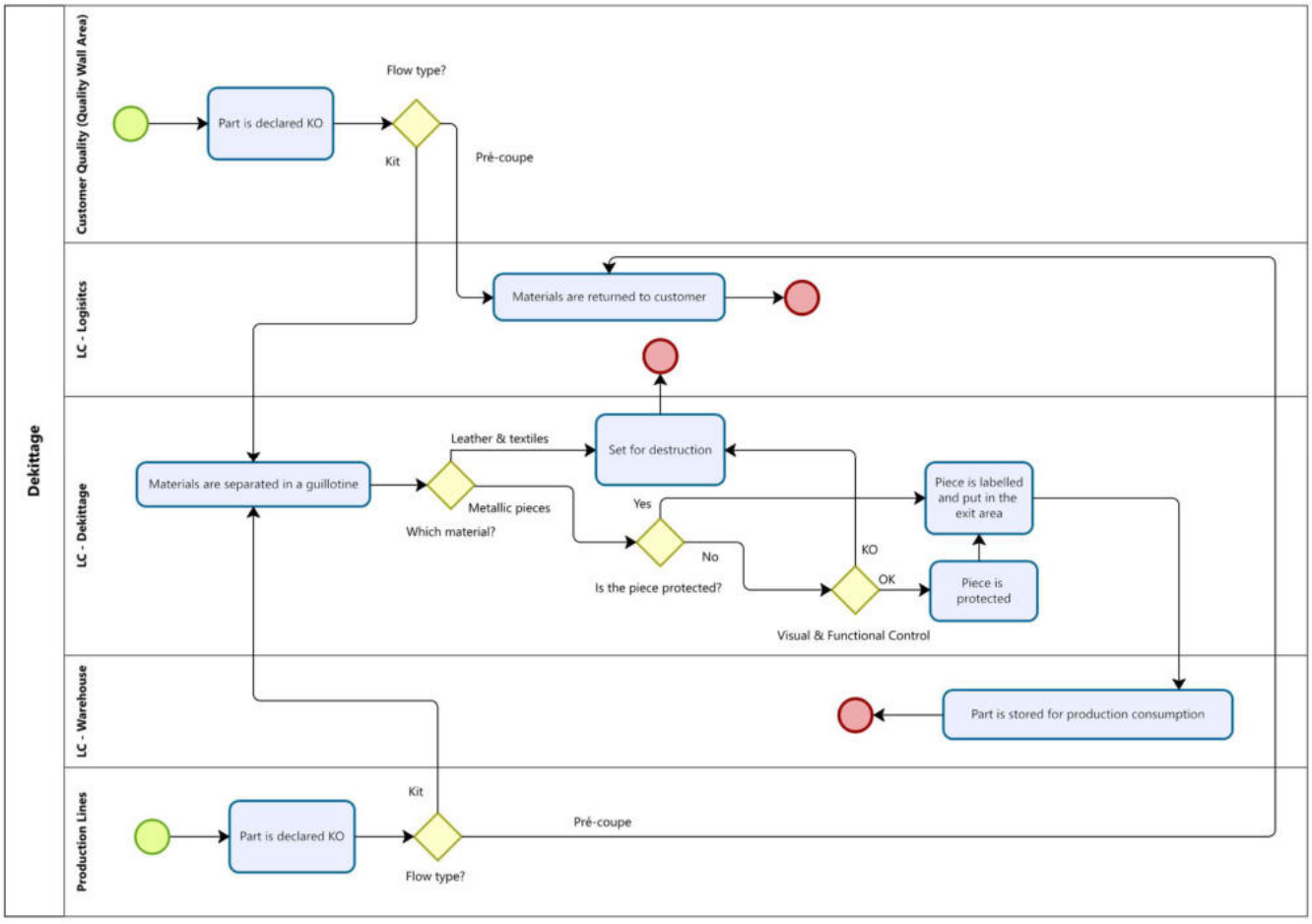


Fig. 4. Flowchart of the dekittage process

In this process, metallic pieces are directed to the exit rack for storage by logistics if they remain unpacked. However, if they lack original packaging and protection, they undergo a quality check. Components considered OK proceed to the same

exit rack, while those identified as KO are allocated for destruction, following the same procedure as before SAP implementation.

At the moment, the proposed workflow is in the implementation phase within the ongoing SAP deployment project and has not yet been fully stabilized across all operations. Therefore, the present study focuses on the design and expected benefits of the future-state process rather than on post-implementation performance results. Future evaluation of the process will include monitoring key indicators such as the number of metallic pieces successfully reused, reduction in destroyed components, traceability compliance rate, process lead time, and financial savings achieved through material recovery.

V. Final Remarks

This study presented a comprehensive analysis of quality management procedures within ATEPELI, a leading manufacturer of luxury leather components. Examination of operational quality, customer returns, and parts return processes revealed critical challenges in material traceability, defect classification, and the management of metallic pieces. Over seven months (September 2023–March 2024), an average of 11,363 metallic pieces per month were returned, of which 8,401 were fully compliant yet destroyed due to traceability limitations, representing a potential financial saving of approximately €25,690 per month.

The ongoing SAP implementation addresses these challenges by digitalizing and integrating operational and quality processes, enabling real-time traceability of materials from receipt to customer dispatch. The introduction of the "dekittage" station further supports controlled handling and storage of metallic pieces for reuse. Together, these improvements demonstrate how quality management can directly contribute to sustainability, reducing waste while maintaining the stringent quality standards of the luxury sector.

This case study provides empirical evidence that integrating continuous improvement, process digitalization, and supply chain visibility can reconcile luxury quality requirements with environmental responsibility, offering a model for manufacturers aiming to reduce waste, optimize resources, and improve financial performance.

Acknowledgment

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Unit Project Scope UID/00319/2025 - Centro ALGORITMI (ALGORITMI/UM) https://doi.org/10.54499/UID/00319/2025.

References

Bentahar, O., & Benzidia, S. (2018). Sustainable supply chain management: Trends and challenges. *Transportation Research Part E: Logistics and Transportation Review*, *119*. doi:10.1016/j.tre.2018.05.005

Bhattacharya, A., & Dey, P. K. (2020). Achieving sustainability through innovation-led lean approaches to manufacturing supply chains. *International Journal of Production Economics*, *219*, 402–404. doi:https://doi.org/10.1016/j.ijpe.2019.107558

Brun, A., Caniato, F., Caridi, M., Castelli, C., Miragliotta, G., Ronchi, S., … Spina, G. (2008). Logistics and supply chain management in luxury fashion retail: Empirical investigation of Italian firms. *International Journal of Production Economics*, *114*, 554–570. doi:10.1016/j.ijpe.2008.02.003

Brun, A., & Moretto, A. (2013). Organisation and supply chain for quality control in luxury companies. *Journal of Fashion Marketing and Management*, *18*. doi:10.1108/JFMM-11-2012-0070

Chiu, C. H., Choi, T. M., Dai, X., Shen, B., & Zheng, J. H. (2018). Optimal Advertising Budget Allocation in Luxury Fashion Markets with Social Influences: A Mean-Variance Analysis. *Production and Operations Management*, *27*(8), 1611–1629. doi:10.1111/poms.12886

Choi, T.-M., & Liu, N. (2019). Optimal advertisement budget allocation and coordination in luxury fashion supply chains with multiple brand-tier products. *Transportation Research Part E: Logistics and Transportation Review*, *130*, 95–107. doi:https://doi.org/10.1016/j.tre.2019.08.009

Feng, Y., Tong, X., & Zhu, Q. (2020). The market value of sustainable practices in the luxury industry: An identity mismatch and institutional theoretical perspective. *Transportation Research Part E: Logistics and Transportation Review*, *137*, 101919. doi:https://doi.org/10.1016/j.tre.2020.101919

Foster, T., Wallin, C., & Ogden, J. (2011). Towards a better understanding of supply chain quality management practices. *International Journal of Production Research*, *49*, 2285–2300. doi:10.1080/00207541003733791

Gonçalves, M., Lobarinhas, A., Domingues, P., Sousa, S., & Teixeira, S. (2026). QUALITY MANAGEMENT PROCESSES FOR SAP IMPLEMENTATION IN A LUXURY LEATHER GOODS. *Journal of Engineering, Management and Information Technology*, *04*(03), 217–230. doi:10.61552/JEMIT.2026.03.001

Gonçalves, M., Lobarinhas, A., & Teixeira, S. (2025). *Analysis of Quality Management Procedures in a Luxury Leather Goods Industry*. In *IJCIEOM – International Joint Conference on Industrial Engineering and Operations Management* (pp. 1–12). doi:10.14488/ijcieom2025_full_0087_38238

Gonçalves, M., Martins, P., Pereira, G., & Sousa, R. (2024). Generic Representation Language for Modeling Transport and Material Handling Systems in Smart Manufacturing Systems. *Processes*, *13*, 43. doi:10.3390/pr13010043

Holmqvist, J., Wirtz, J., & Fritze, M. (2022). Case Study: Digital Luxury Services: Traditions Versus Innovation in Luxury Fashion (pp. 550–552). doi:10.1142/9781944659806_0018

Kai, X. (2021). Research on brand alliance strategy innovation based on the big data analysis of consumers' attitude Take Louis Vuitton and supreme as an example. *E3S Web of Conferences*, *235*, 1075. doi:10.1051/e3sconf/202123501075

Kapferer, J.-N., & Michaut, A. (2014). Is luxury compatible with sustainability? Luxury consumers' viewpoint. *The Journal of Brand Management*, *21*. doi:10.1057/bm.2013.19

Lobarinhas, A., Gonçalves, M., & Teixeira, S. (2024). *The Role of Employees in Implementing an SAP System - A Practical Case in the Luxury Leather Goods Industry*. In *ENEGI 2024*.

Rahman, S.-U., & Bullock, P. (2005). Soft TQM, hard TQM, and organisational performance relationships: An empirical investigation. *Omega*, *33*(1), 73–83. doi:10.1016/j.omega.2004.03.008

Robinson, C., & Malhotra, M. (2005). Defining the concept of supply chain quality management and its relevance to academic and industrial practice. *International Journal of Production Economics*, 96, 315–337. doi:10.1016/j.ijpe.2004.06.055

Teegavarapu, S., Summers, J. D., & Mocko, G. M. (2008). *Case study method for design research: A justification*. In *Proceedings of the ASME Design Engineering Technical Conference* (Vol. 4, pp. 495–503). doi:10.1115/DETC2008-49980

Zheng, J. (2023). Discover the Hidden Gems: Behind the Scenes of Quality Control in Luxury Handbag Manufacturing. Sustainable Fashion Expertise. Retrieved 13 February 2025, from https://www.linkedin.com/pulse/discover-hidden-gems-behind-scenes-quality-control-luxury-jimmy-zheng/

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Total Quality Management Modernized Definition for the Digital Era: A Delphi Validation

Ahmed Baha Eddine Aichouni [1]

[1] CEMMPRE, ARISE, Department of Mechanical Engineering, University of Coimbra, 3030-788 Coimbra, Portugal; aaichouni@student.uc.pt

Abstract **- The rapid diffusion of Industry 4.0 technologies has altered the operational logic of quality management, yet the field lacks a formally bounded and empirically validated definition of Total Quality Management (TQM) suitable for digital environments. Existing studies exhibit conceptual fragmentation, limiting construct clarity and constraining cumulative empirical research. This study addresses this gap by developing and validating a modernized socio-technical definition of TQM through a structured Delphi methodology. A sequential mixed-methods design was employed. Initial conceptual dimensions were derived from a systematic literature review and refined through semi-structured interviews with industry experts. These dimensions were subsequently evaluated using a two-round Delphi study involving an international panel of 13 experts comprising academic and senior industry practitioners. Consensus was assessed using multiple statistical indicators, including mean scores, standard deviation, Content Validity Ratio (CVR), interrater agreement (a_{wg}), Kendall's coefficient of concordance (W), and Spearman's rank correlation coefficient (ρ). The results demonstrate statistically significant consensus across all five dimensions of the proposed construct. All dimensions exceeded predefined thresholds for importance (μ ≥ 4.00), agreement (σ ≤ 0.80; $a_{wg} > 0.90$), and essentiality (CVR ≥ 0.54), while Kendall's W (0.82, p < 0.001) confirmed strong global agreement. Inter-round stability (ρ = 0.87–0.93) further justified the termination of the Delphi process after two rounds. The findings reveal a structured hierarchy within the construct, with empowered behavioural practices emerging as the dominant subsystem, reinforcing the central role of human factors in digital quality systems. The study establishes an empirically validated definition of TQM as an adaptive socio-technical management system integrating behavioural and data-driven continuous improvement methodologies to govern processes, fulfil stakeholder requirements, and sustain long-term organizational value. This contribution resolves longstanding definitional ambiguity, aligns with established principles of construct development, and provides a stable theoretical foundation for Quality 4.0 research. The validated construct enables the development of standardized measurement instruments and supports future empirical investigations using advanced statistical modelling techniques.**

Keywords – Total Quality Management, Quality 4.0, Delphi Method, Definition Validation, Empirical Study, Organizational Sustainability

I. INTRODUCTION

The integration of Industry 4.0 technologies into quality management systems has accelerated across both advanced and emerging economies, reflecting a structural shift in how organizations design, monitor, and improve operational processes. Empirical research documents the increasing adoption of digital quality practices in highly industrialized nations, including Germany, Japan, the United States, China, Italy, and the United Kingdom, alongside growing implementation in developing economies pursuing higher levels of operational capability (Chiarini & Kumar, 2021; Sader et al., 2022). These developments indicate that quality management is no longer confined to traditional inspection and control paradigms but is progressively embedded within data-intensive, real-time decision architectures.

Total Quality Management (TQM), originally conceptualized within post-war industrial systems, has historically emphasized process standardization, statistical control, and managerial responsibility for continuous improvement (Deming, 1986). However, the emergence of advanced digital infrastructures, characterized by cyber-physical systems, predictive analytics, and interconnected production environments, has altered the operational logic of quality management. Contemporary formulations, often described under the label "Quality 4.0," extend TQM beyond retrospective control toward predictive, adaptive, and data-driven execution (Sony et al., 2020). Institutional developments, including the revision of ISO 9001:2015, further reinforce this transition by embedding risk-based thinking and systemic integration into quality management standards.

Despite the rapid expansion of digital quality practices, the literature lacks empirically validated definition of TQM applicable to high-velocity digital environments. Existing studies present divergent interpretations, frequently emphasizing either technological capabilities or managerial philosophy without establishing a coherent construct (Kushwaha & Talib, 2025; Khayat et al., 2025). This conceptual fragmentation generates substantive methodological limitations. In the absence of a stable definition, empirical studies lack construct validity, measurement instruments remain inconsistent, and theoretical accumulation is constrained (Nyirenda & Kumar, 2019). Consequently, the field exhibits a persistent disconnect between conceptual discourse and operational application.

From a theory development perspective, the absence of definitional clarity violates established principles of construct specification. A valid theoretical construct must be defined using non-tautological, unambiguous primitives, must clearly delineate its boundaries, and must avoid conflating processes with outcomes (Wacker, 2004; Suddaby, 2010). Existing definitions of TQM frequently fail to meet these criteria, either by embedding performance outcomes within the construct or by overlapping with adjacent frameworks such as Lean or Six Sigma. These deficiencies are amplified in the context of Industry 4.0, where the integration of advanced technologies introduces additional conceptual complexity. A further limitation in current Quality 4.0 discourse is the disproportionate emphasis on technological subsystems at the expense of human and organizational factors. Socio-Technical Systems (STS) theory provides a theoretically grounded framework to address this

imbalance by asserting that organizational performance emerges from the joint optimization of social and technical subsystems (Trist, 1981). Within digital production environments, the effectiveness of predictive analytics, artificial intelligence, and automated decision systems is contingent upon human capabilities, including interpretive judgment, adaptive learning, and autonomous problem-solving. Consequently, any valid conceptualization of TQM in the digital era must explicitly integrate both behavioral and technological dimensions within a unified framework.

Recent theoretical work has proposed a modernized socio-technical definition of TQM that integrates empowered behavioral practices with data-driven continuous improvement methodologies (Aichouni, 2026). While this formulation satisfies key theoretical requirements at the conceptual level, it remains derived exclusively from literature synthesis and has not been subjected to empirical validation. Without structured evaluation by domain experts, the construct cannot be considered operationally robust or generalizable across industrial contexts. To address this gap, the present study employs the Delphi method to evaluate and validate the proposed definition through structured expert consensus. The Delphi technique is particularly suited to contexts characterized by conceptual ambiguity and limited empirical grounding, as it enables systematic convergence of expert judgment under controlled conditions (Dalkey & Helmer, 1963). By integrating perspectives from both academic researchers and senior industry practitioners, this study aims to establish construct validity through statistically defined consensus criteria.

Accordingly, the primary objective of this research is to empirically validate a modernized, socio-technical conceptual definition of Total Quality Management for the digital era. This is achieved through a multi-round Delphi process designed to assess the relevance, clarity, and essentiality of the proposed construct dimensions. The resulting validated definition is intended to provide a stable theoretical foundation for future empirical research and to support the systematic development of measurement instruments and operational frameworks within Quality 4.0 environments.

II. LITERATURE REVIEW

A. From Classical TQM to Quality 4.0

Total Quality Management has historically been positioned as an integrative management philosophy aimed at achieving sustained competitive advantage through process control, continuous improvement, and organizational alignment (Powell, 1995; Deming, 1986). Early formulations emphasized statistical quality control, reduction of process variability, and the institutionalization of managerial responsibility for quality outcomes. These models were developed within relatively stable, manufacturing-centric environments characterized by linear production flows and limited data availability. Subsequent theoretical developments extended TQM beyond operational control toward broader organizational constructs, including learning systems, stakeholder orientation, and strategic alignment (Dahlgaard-Park et al., 2018; van Kemenade & Hardjono, 2019). Despite this expansion, the underlying logic of TQM remained largely retrospective, relying on historical data and post hoc analysis to guide improvement initiatives.

The emergence of Industry 4.0 introduces a fundamentally different operational context. Advanced technologies, including artificial intelligence, the Internet of Things, and cyber-physical systems, enable continuous data acquisition, real-time monitoring, and predictive decision-making. Within this environment, quality management transitions from reactive control toward anticipatory and adaptive system governance (Sony et al., 2020; Chiarini & Kumar, 2021). This transformation is commonly conceptualized under the term "Quality 4.0." However, the rapid proliferation of technology-oriented studies has produced significant conceptual fragmentation. Existing literature frequently prioritizes technological capabilities without establishing a coherent theoretical construct defining the nature, scope, and boundaries of TQM in digital environments (Kushwaha & Talib, 2025; Khayat et al., 2025). As a result, the field exhibits a lack of cumulative theoretical development.

A persistent issue is the conceptual duality in TQM research. The construct is alternately treated as a philosophical orientation toward continuous improvement and as a discrete set of operational tools (Nyirenda & Kumar, 2019). Early definitions, such as "conformance to requirements" (Crosby, 1984), illustrate this limitation by reducing quality management to outcome-based metrics while neglecting the underlying system architecture. Such definitions lack explanatory power in complex, data-intensive environments. From a construct development perspective, these formulations violate fundamental theoretical requirements. A valid construct must specify its essential attributes, maintain clear boundaries, and avoid tautological relationships between processes and outcomes (Wacker, 2004; Suddaby, 2010). Many contemporary interpretations fail to meet these conditions, either by embedding performance outcomes within the construct or by overlapping with adjacent methodologies such as Lean and Six Sigma. Consequently, the literature lacks a formally bounded definition capable of supporting rigorous empirical investigation.

B. Socio-Technical Systems as a Structural Requirement

The current Quality 4.0 discourse demonstrates a disproportionate emphasis on technological subsystems, often treating digital tools as the primary drivers of quality performance. This orientation neglects the role of human agency, organizational culture, and behavioral adaptation in determining system effectiveness. Socio-Technical Systems (STS) theory provides a rigorous framework to address this imbalance. Originating from organizational design research, STS theory posits that system performance is contingent upon the joint optimization of social and technical subsystems rather than the maximization of either component in isolation (Trist, 1981). This principle is directly applicable to digital quality environments, where advanced technologies operate within complex organizational contexts.

Empirical studies indicate that the implementation of Industry 4.0 technologies without corresponding human adaptation leads to suboptimal outcomes. Digital systems require users capable of interpreting complex data outputs, exercising judgment under uncertainty, and engaging in continuous learning processes (Sony & Naik, 2020; Ali & Johl, 2022). In this context, workforce autonomy, skill development, and cultural alignment are not auxiliary factors but structural prerequisites for system functionality. Furthermore, the increasing reliance on predictive analytics and automated decision systems introduces new

forms of operational interdependence. Technical systems generate probabilistic insights rather than deterministic outputs, requiring human validation and contextual interpretation (Ramanathan & Watson, 2021). This reinforces the necessity of integrating behavioral capabilities within the architecture of quality management systems. Despite these considerations, existing Quality 4.0 frameworks often treat human factors as secondary implementation concerns rather than integral components of the construct. This omission contributes directly to the conceptual instability observed in the literature. A theoretically robust definition of TQM in the digital era must therefore explicitly incorporate both subsystems and define their functional interrelationship.

C. Construct Specification and the Need for Empirical Validation

The absence of a formally specified construct for TQM in digital environments represents a critical barrier to theoretical and empirical advancement. According to established principles of theory development, constructs must be defined using unambiguous primitives, must avoid embedding causal relationships within the definition, and must clearly distinguish between the phenomenon of interest and its outcomes (Wacker, 2004; Suddaby, 2010). Without adherence to these criteria, empirical measurement becomes inconsistent and theoretical accumulation remains limited.

Recent work has proposed a socio-technical conceptualization of TQM that integrates empowered behavioral practices with data-driven continuous improvement methodologies (Aichouni, 2026). This formulation addresses several deficiencies in the existing literature by explicitly separating system structure from performance outcomes and by incorporating both social and technical dimensions. However, the definition remains derived from secondary sources and has not been subjected to systematic empirical evaluation. In the absence of empirical validation, conceptual definitions cannot achieve construct validity or gain acceptance within the academic and practitioner communities. Expert-based evaluation is required to assess whether the proposed dimensions accurately capture the essential characteristics of the phenomenon and whether they are applicable across diverse operational contexts. The Delphi method provides a structured mechanism for such evaluation. It is specifically designed for domains characterized by incomplete theoretical development and limited empirical consensus (Dalkey & Helmer, 1963; Murry & Hammons, 1995). Through iterative rounds of controlled feedback, the method enables convergence toward statistically measurable agreement among experts while minimizing the influence of dominant individuals or group dynamics. Accordingly, this study employs the Delphi method to evaluate the proposed socio-technical definition of TQM. By integrating the perspectives of academic researchers and senior industry practitioners, the study aims to establish construct validity through predefined statistical criteria, thereby transforming a conceptual proposition into an empirically supported framework.

III. RESEARCH METHODOLOGY

A. Research Design

This study adopts a sequential mixed-methods design to establish construct validity for a socio-technical definition of Total Quality Management (TQM) in digital environments. The design integrates qualitative exploration with quantitative consensus validation to ensure both conceptual rigor and empirical reliability. The methodological structure comprises three phases: (1) construct derivation, (2) Delphi-based validation, and (3) framework synthesis. In the first phase, conceptual dimensions were derived through a structured review of the quality management and Industry 4.0 literature, supplemented by semi-structured interviews with industry experts. The second phase employed a two-round Delphi process to evaluate and refine these dimensions under controlled conditions. The final phase consolidated the validated dimensions into an empirically supported conceptual framework. The selection of the Delphi method is methodologically appropriate given the absence of a stable theoretical construct and the need to evaluate complex, multidimensional phenomena through expert judgment (Dalkey & Helmer, 1963; Okoli & Pawlowski, 2004).

B. Identification and Refinement of Conceptual Dimensions

Initial conceptual dimensions were identified through a structured search of Scopus, Web of Science, and Google Scholar databases using the keywords: "Total Quality Management," "Quality 4.0," "digital transformation," "socio-technical systems," and "continuous improvement." The extracted literature was subjected to thematic synthesis to isolate recurring constructs and definitional elements. A preliminary conceptual definition was formulated and decomposed into measurable dimensions. To ensure face validity and eliminate semantic ambiguity prior to quantitative testing, four semi-structured interviews were conducted with senior quality executives. These interviews focused on clarity, relevance, and applicability of the proposed dimensions within contemporary industrial environments. Feedback from this phase resulted in targeted semantic refinements without altering the structural composition of the construct. Only convergent feedback across multiple respondents was incorporated, thereby limiting idiosyncratic bias.

C. Delphi Panel Selection

The validity of Delphi outcomes depends on the expertise of the selected panel rather than sample size. Accordingly, purposive sampling was employed to recruit domain experts based on predefined criteria: Academic experts with a minimum of 10 years of experience in operations management or quality management and a demonstrated publication record in peer-reviewed journals. Industry practitioners occupying senior quality management roles (e.g., Chief Quality Officer, Quality Director) with a minimum of 15 years of professional experience. This approach aligns with established methodological recommendations emphasizing expertise-driven sampling in Delphi studies (Okoli & Pawlowski, 2004; Niederberger & Spranger, 2020). A total of 13 experts were recruited, consisting of six academics and seven practitioners (Table 1). This panel size falls within the recommended range for homogeneous expert groups addressing specialized topics.

TABLE I. COMPOSITION OF THE DELPHI EXPERT PANEL

Expert ID	Cohort	Position	Education Level	Years of Experience

A-01	Academic	Full Professor (Operations Management)	Ph.D.	25
A-02	Academic	Full Professor (Quality)	Ph.D.	21
A-03	Academic	Associate Professor (Industry 4.0)	Ph.D.	14
A-04	Academic	Full Professor (TQM & Lean)	Ph.D.	30
A-05	Academic	Associate Professor (Predictive Analytics)	Ph.D.	15
A-06	Academic	Associate Professor (Organizational Behavior)	Ph.D.	18
P-01	Practitioner	Quality Officer	Master's	22
P-02	Practitioner	VP Quality Assurance	Master's	20
P-03	Practitioner	Director of Digital Transformation	Master's	16
P-04	Practitioner	Senior Quality Consultant	Ph.D.	28
P-05	Practitioner	Quality Manager	Bachelor's	19
P-06	Practitioner	Digitalization Specialist	Master's	15
P-07	Practitioner	Quality Systems Director	Master's	24

D. Consensus Criteria and Statistical Thresholds

To eliminate subjective interpretation and ensure methodological rigor, consensus was operationalized a priori using a multi-criteria statistical framework incorporating measures of central tendency, dispersion, and content validity. This multi-metric approach is consistent with established Delphi research protocols, which recommend the simultaneous application of complementary indicators to capture different dimensions of agreement.

Central tendency was evaluated using the arithmetic mean (μ). Although Likert-scale data are ordinal, the use of the mean is widely accepted in Delphi studies when the objective is to assess the relative importance of items across a homogeneous expert panel. A threshold of $\mu \geq 4.00$ was imposed, corresponding to a minimum classification of "high importance" on the five-point scale. This criterion ensures that only dimensions perceived as substantively relevant by the panel are retained. Dispersion of responses was assessed using the standard deviation (σ). A threshold of $\sigma \leq 0.80$ was adopted to indicate low variability and convergence of expert opinions. This threshold is consistent with prior Delphi studies, where values below 1.00 are generally interpreted as acceptable agreement, and stricter thresholds (≤0.80) indicate strong consensus.

To evaluate the essentiality of each dimension, the Content Validity Ratio (CVR) was computed following Lawshe's (1975) formulation:

$$CVR = \frac{n_e - (N/2)}{N/2}$$

where $n_e$represents the number of panelists identifying an item as "essential," and $N$is the total number of panelists. For this study, responses rated as 4 ("high importance") or 5 ("extremely important") were operationally classified as "essential," aligning the ordinal scale with the dichotomous requirement of the CVR calculation. For a panel size of N = 13, the minimum acceptable CVR value at $\alpha = 0.05$ is 0.54. In addition to these primary criteria, overall panel agreement was assessed using Kendall's coefficient of concordance (W). Kendall's W evaluates the degree of association among raters and ranges from 0 (no agreement) to 1 (complete agreement). Values exceeding 0.70 are generally interpreted as strong consensus in Delphi studies. The inclusion of Kendall's W addresses limitations associated with relying solely on dispersion measures and provides a global indicator of agreement across all dimensions.

E. Delphi Procedure

The Delphi process was executed over two structured rounds, consistent with methodological recommendations indicating that stability in expert judgment is typically achieved within two to three iterations when panel composition is highly specialized.

In Round 1, panelists were provided with the preliminary conceptual dimensions and the associated definition of TQM for the digital era. Each dimension was evaluated using a five-point Likert scale. In addition to quantitative ratings, participants were required to provide qualitative justifications for their evaluations. This dual-data approach ensured that numerical assessments were supported by domain-specific reasoning. Qualitative responses were analyzed using a structured thematic analysis procedure. Initial coding was conducted to identify recurrent patterns in terminology, conceptual clarity, and perceived omissions. Codes were then aggregated into higher-order themes reflecting convergent expert perspectives. To enhance analytical reliability, only themes supported by multiple panelists were retained for revision. Isolated or contradictory feedback was excluded to prevent disproportionate influence from individual respondents. Based on the thematic synthesis, minor semantic refinements were introduced. These modifications were restricted to improving clarity and contextual relevance; no structural alterations to the conceptual dimensions were made to preserve construct integrity.

Round 2 was initiated following a controlled feedback process. Panelists received anonymized summaries of group responses, including mean scores and qualitative insights. This feedback mechanism is central to the Delphi methodology, enabling participants to reassess their judgments in light of collective expertise while maintaining independence of opinion. The termination criterion for the Delphi process was defined as a priori based on inter-round stability and statistical convergence. Stability was assessed using Spearman's rank correlation coefficient, while convergence was evaluated through the predefined thresholds for mean, dispersion, CVR, and Kendall's W. The combination of high inter-round correlation and satisfaction of all consensus criteria justified termination after the second round, thereby avoiding unnecessary iteration and potential respondent fatigue.

F. Statistical Analysis Protocol

All statistical analyses were conducted using SPSS, following a structured protocol aligned with the data characteristics and research objectives. Given the ordinal nature of Likert-scale data and the limited sample size (N = 13), non-parametric statistical

techniques were employed throughout the analysis. This decision is grounded in measurement theory rather than distributional assumptions, as ordinal data do not satisfy the interval-level requirements necessary for parametric inference.

Internal consistency of the measurement instrument was evaluated using Cronbach's alpha (α). While the application of α in Delphi studies is subject to debate, given that items represent conceptual dimensions rather than reflective indicators, it remains a useful diagnostic tool for assessing the coherence of expert evaluations. Values exceeding 0.80 were interpreted as indicating high internal consistency.

The stability of expert judgments between Round 1 and Round 2 was quantified using Spearman's rank correlation coefficient (ρ). High values of ρ indicate that panelists maintained consistent relative rankings across iterations, reflecting convergence of opinion rather than random variation.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

To assess the absolute level of agreement among panelists, the interrater agreement index (a_{wg}) was computed. Unlike dispersion measures, which capture variability, a_{wg} evaluates the extent to which observed ratings deviate from a null distribution. Values above 0.90 indicate strong agreement and are commonly interpreted as evidence of consensus in organizational research.

Kendall's coefficient of concordance (W) was calculated to provide a global measure of agreement across all raters and dimensions. Statistical significance of W was evaluated using the chi-square approximation:

$$\chi^2 = N(k - 1)W$$

where $N$is the number of raters and $k$is the number of items. This test determines whether the observed level of agreement is significantly greater than chance.

The combined use of central tendency, dispersion, CVR, interrater agreement, and Kendall's W ensures a comprehensive and statistically defensible evaluation of consensus. This multi-layered analytical framework mitigates the limitations associated with any single metric and aligns with best practices in Delphi-based construct validation.

IV. RESULTS AND ANALYSIS

A. *Panel Retention and Sociodemographic Integrity*

The methodological robustness of a Delphi study is contingent upon panel stability across iterative rounds and the preservation of expert diversity. Attrition between rounds introduces non-response bias and may distort convergence patterns. Accordingly, participant retention was explicitly monitored as a validity control mechanism. The study achieved full retention across both rounds. All 13 recruited experts (100%) completed Round 1 and subsequently participated in Round 2. The absence of attrition eliminates the risk of systematic bias introduced by differential dropout and ensures that the consensus formation process reflects a consistent expert cohort. This stability strengthens the internal validity of the findings and supports the reliability of inter-round comparisons.

Beyond retention, the structural composition of the panel was evaluated to ensure adequate representation across geographical regions, professional domains, and experience levels. The panel included experts from eight countries spanning North America, South America, Europe, Asia, and the Middle East and North Africa (MENA) region. This distribution mitigates the risk of regional bias and enhances the external relevance of the validated construct. The professional composition of the panel reflects a deliberate balance between academic and practitioner perspectives. Six participants (46%) were academic researchers with established publication records in quality management and Industry 4.0, while seven participants (54%) were senior industry practitioners occupying leadership roles in quality and digital transformation. This dual composition ensures that the evaluation integrates both theoretical rigor and operational applicability. Experience levels further reinforce the credibility of the panel. All participants exceeded the minimum inclusion thresholds, with the majority possessing more than 15 years of domain-specific experience. This concentration of expertise is consistent with methodological recommendations for Delphi studies addressing complex and underdefined constructs (Table 2).

TABLE II. SOCIODEMOGRAPHIC CHARACTERISTICS OF THE DELPHI EXPERT PANEL (N = 13)

Characteristic	Category	Frequency	Percentage (%)
Geographical Distribution	Saudi Arabia	2	15
	United Arab Emirates	2	15
	Algeria	1	8
	Germany	2	15
	France	1	8
	United States	2	15
	Brazil	1	8
	Malaysia	2	15
Primary Role	Academic Researcher	6	46
	Industry Practitioner	7	54
Domain Expertise	Quality 4.0 / TQM Research	6	46
	Chief Quality Officer / Director	4	31
	Digitalization / Analytics Specialist	3	23
Years of Experience	10–15 years	3	23

	16–20 years	4	31
	21–25 years	4	31
	26+ years	2	15

Note: Percentages may not sum to exactly 100% due to rounding.

The distribution indicates a high level of heterogeneity in geographical representation combined with homogeneity in expertise level. This configuration is methodologically desirable in Delphi studies, as it ensures diversity of perspectives while maintaining a consistent baseline of domain competence. From a validity standpoint, the combination of full retention, balanced cohort composition, and extensive professional experience supports both the internal consistency of the consensus process and the external plausibility of the findings. The panel structure is therefore considered adequate for the empirical validation of a complex socio-technical construct.

B. Statistical Gatekeeping: Reliability, Agreement, and Stability (Final Draft)

Prior to evaluating consensus outcomes, a sequence of statistical gatekeeping procedures was executed to verify the internal consistency of the instrument, the degree of agreement among panelists, and the stability of responses across iterative rounds. This multi-layered validation protocol ensures that subsequent consensus claims are supported by statistically defensible evidence rather than descriptive aggregation. The internal consistency of the measurement instrument was assessed using Cronbach's alpha (α). Although Delphi studies do not always conceptualize items as reflective indicators of a single latent construct, α remains a useful diagnostic for evaluating the coherence of expert ratings across dimensions. The analysis yielded α = 0.89 for Round 1 and α = 0.91 for Round 2. Both values exceed the recommended threshold of 0.80, indicating high internal consistency. The increase in α between rounds suggests improved alignment among expert evaluations following controlled feedback, consistent with the convergence objective of the Delphi process.

To assess the absolute level of agreement among panelists, the interrater agreement index (a_{wg}) was computed for each dimension. This index evaluates the extent to which observed ratings deviate from a null distribution, providing a direct measure of consensus strength. Across all five dimensions, a_{wg} values ranged from 0.91 to 0.95. These values exceed the 0.90 threshold commonly interpreted as indicating very strong agreement. This result confirms that the observed convergence is not attributable to random clustering of responses but reflects substantive alignment among experts.

To complement dimension-level agreement measures, Kendall's coefficient of concordance (W) was calculated as a global indicator of consensus across all panelists and dimensions. Kendall's W captures the degree to which experts consistently rank items in a similar order. The analysis produced W = 0.82 for Round 2. This value indicates strong agreement among panelists. Statistical significance was assessed using the chi-square approximation:

$$\chi^2 = N(k-1)W$$

Substituting N = 13 panelists and k = 5 dimensions yields:

$$\chi^2 = 13 \times 4 \times 0.82 = 42.64$$

With degrees of freedom (df = k − 1 = 4), the result is significant at $p < 0.001$. This confirms that the observed level of agreement is substantially greater than would be expected by chance.

The stability of expert judgments between Round 1 and Round 2 was evaluated using Spearman's rank correlation coefficient (ρ). This measure assesses whether panelists maintained consistent relative rankings across iterations, which is a critical condition for terminating the Delphi process.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

The calculated ρ values ranged from 0.87 to 0.93 across the five dimensions. These coefficients indicate very high positive correlations, demonstrating that expert opinions stabilized after the first round (Table 3). Minor adjustments observed in Round 2 reflect controlled refinement rather than structural shifts in judgment.

TABLE III. STATISTICAL GATEKEEPING INDICATORS

Statistical Test	Metric	Result	Threshold	Interpretation
Reliability	Cronbach's α (R1)	0.89	≥ 0.80	High internal consistency
	Cronbach's α (R2)	0.91	≥ 0.80	High internal consistency
Interrater Agreement	a_{wg} (range)	0.91–0.95	≥ 0.90	Very strong agreement
Global Consensus	Kendall's W	0.82	≥ 0.70	Strong agreement
	χ^2 (df = 4)	42.64***	$p < 0.05$	Statistically significant
Inter-Round Stability	Spearman's ρ	0.87–0.93	≥ 0.80	High stability

***$p < 0.001$

C. Round 1: Qualitative Elicitation and Construct Refinement

The first Delphi round was designed to evaluate the face validity, conceptual clarity, and structural completeness of the proposed TQM definition and its associated dimensions. In addition to quantitative ratings, panelists were required to provide qualitative justifications to support their evaluations. This dual-input design ensures that numerical convergence is grounded in domain-specific reasoning rather than superficial agreement.

A total of 42 qualitative statements were collected from the 13 panelists, yielding an average of 3.23 substantive comments per expert. The qualitative dataset was subjected to a structured thematic analysis to identify recurrent patterns in expert feedback. The analysis followed a three-stage coding procedure: (1) open coding to extract key concepts, (2) axial coding to group related concepts into higher-order categories, and (3) selective coding to isolate dominant themes exhibiting convergence across multiple respondents. To maintain analytical rigor, only themes supported by at least three independent panelists were retained. This threshold minimizes the influence of idiosyncratic opinions and ensures that revisions reflect collective expert judgment. The analysis produced three dominant thematic clusters: (1) validation of the socio-technical architecture, (2) emphasis on adaptive system behavior, and (3) refinement of terminology related to digital execution.

Both academic and practitioner cohorts demonstrated strong alignment regarding the structural validity of the socio-technical foundation. Panelists explicitly supported the integration of behavioral and technological subsystems, confirming that the proposed construct reflects the operational realities of digital quality environments. Academic respondents emphasized the theoretical correctness of separating the management system from its outcomes. Several panelists noted that this distinction resolves a longstanding issue in TQM literature, where definitions frequently conflate processes with performance results. This feedback reinforces compliance with established construct development principles requiring non-tautological definitions. Practitioners corroborated this perspective from an operational standpoint, indicating that successful digital quality systems depend on the interaction between human decision-making capabilities and data-driven infrastructures. No panelist proposed removing either subsystem, indicating unanimous structural acceptance.

A second dominant theme concerned the necessity of explicitly characterizing TQM as an adaptive system. Multiple panelists highlighted that static representations of quality management are incompatible with Industry 4.0 environments, where processes are continuously recalibrated based on real-time data streams. Practitioners, in particular, emphasized that algorithmic systems require ongoing adjustment to maintain relevance under changing operational conditions. Consequently, the inclusion of "adaptive" as a defining attribute was reinforced through convergent feedback. This modification ensures that the construct reflects the dynamic nature of digital production systems rather than static compliance frameworks.

The third thematic cluster focused on the precision of terminology describing the technical subsystem. Panelists identified limitations in legacy expressions such as "continuous improvement tools," which were perceived as insufficient to capture the complexity of modern digital infrastructures. Convergent feedback supported the adoption of the phrase "data-driven continuous improvement methodologies." This terminology was considered more accurate in representing the integration of predictive analytics, machine learning, and real-time monitoring systems. The refinement improves semantic precision without altering the underlying conceptual structure.

A subset of qualitative comments proposed alternative conceptual directions, including the incorporation of sector-specific practices or the expansion of the construct to include additional managerial functions. These suggestions were not adopted, as they lacked convergence across the panel and risked compromising the generalizability of the definition. The exclusion criterion, requiring support from at least three independent panelists, was consistently applied to ensure methodological transparency and to prevent disproportionate influence from isolated viewpoints.

D. Round 2: Quantitative Consensus Validation

The second Delphi round was conducted to quantitatively evaluate the refined conceptual dimensions of TQM for the digital era against the predefined consensus criteria. Following the controlled feedback provided after Round 1, panelists reassessed each dimension using the five-point Likert scale. The objective of this round was not exploratory but confirmatory: to determine whether statistically verifiable consensus had been achieved across all dimensions. The results demonstrate clear convergence of expert opinion across all evaluative metrics. Each of the five dimensions exceeded the priori thresholds for central tendency ($\mu \geq 4.00$), dispersion ($\sigma \leq 0.80$), and essentiality (CVR ≥ 0.54). The consistency of these outcomes across multiple statistical indicators confirms that the validated dimensions collectively satisfy the conditions required for construct acceptance (Table 4).

TABLE IV. DELPHI ROUND 2 QUANTITATIVE CONSENSUS RESULTS FOR TQM 4.0 DIMENSIONS (N = 13)

Conceptual Dimension	Mean (μ)	Std. Dev. (σ)	CVR	a_{wg}	Consensus Status
Dimension 1: Operates as an adaptive, socio-technical management system	4.61	0.50	0.84	0.92	Validated
Dimension 2: Integrates empowered behavioural practices (Social Subsystem)	4.85	0.37	1.00	0.95	Validated
Dimension 3: Integrates data-driven continuous improvement methodologies (Technical Subsystem)	4.77	0.43	0.84	0.94	Validated
Dimension 4: Designed to systematically govern process and fulfil stakeholder requirements	4.69	0.48	0.69	0.93	Validated
Dimension 5: Aims to sustain long-term organizational value	4.54	0.51	0.69	0.91	Validated

Note: CVR threshold at N = 13 is 0.54 (a = 0.05). a_{wg} values above 0.90 indicate strong agreement.

The central tendency results indicate that all dimensions were evaluated as highly important, with mean values ranging from 4.54 to 4.85. Notably, Dimension 2, representing the integration of empowered behavioural practices, achieved the highest mean score ($\mu = 4.85$), indicating near-unanimous agreement regarding its critical role within the construct. This finding is reinforced by the corresponding standard deviation ($\sigma = 0.37$), which is the lowest observed across all dimensions, indicating minimal dispersion and strong convergence of expert opinion.

Dispersion metrics further confirm the consistency of responses. All standard deviation values fall well below the threshold of 0.80, indicating a tightly clustered distribution of ratings. This low variability suggests that the consensus is not artificially produced by averaging divergent opinions but reflects genuine alignment among panelists.

The Content Validity Ratio (CVR) results provide additional evidence of the essentiality of each dimension. Dimension 2 achieved a CVR of 1.00, indicating that all 13 experts classified it as essential. The remaining dimensions recorded CVR values ranging from 0.69 to 0.84, all exceeding the minimum threshold required for statistical significance. These results confirm that each dimension is not only important but also considered indispensable to the construct.

Interrater agreement, as measured by a_{wg}, ranged from 0.91 to 0.95 across the five dimensions. These values indicate very strong agreement and corroborate the dispersion findings. The alignment between σ and a_{wg} strengthens the robustness of the consensus, as both variability-based and agreement-based metrics converge on the same conclusion.

Importantly, these dimension-level results are consistent with the global agreement measure reported in Section 4.2, where Kendall's W = 0.82 indicated strong overall concordance among panelists. The convergence of global and dimension-specific metrics eliminates the possibility of localized agreement masking broader disagreement patterns.

From a methodological perspective, the simultaneous satisfaction of all predefined criteria, mean, standard deviation, CVR, interrater agreement, and Kendall's W, provides a comprehensive validation of the construct. No dimension approached any rejection threshold, and no conflicting statistical signals were observed. This uniformity is rarely achieved in Delphi studies and indicates a high level of conceptual clarity and shared understanding among experts.

The results therefore justify the acceptance of all five dimensions without modification. The convergence observed in Round 2, combined with the high inter-round stability reported in Section 4.2, confirms that the consensus is both statistically significant and structurally stable. Consequently, the Delphi process was formally concluded after the second round, and the validated dimensions were retained for synthesis into the final conceptual definition.

E. Analysis of Empirical Significance

The quantitative results presented in Section 4.4 establish that all five conceptual dimensions satisfy the predefined statistical thresholds for consensus. However, beyond threshold compliance, a deeper examination of the empirical structure of the results is required to interpret their theoretical and practical significance. This subsection analyzes the relative weighting of dimensions, the internal hierarchy of the construct, and the implications of the observed agreement patterns.

A first-order observation concerns the uniformity of statistical validation across all dimensions. Each variable exceeded the minimum thresholds for mean, dispersion, Content Validity Ratio, and interrater agreement without approaching boundary conditions. This absence of marginal cases indicates that the consensus is not conditional or fragile. Instead, it reflects a stable and coherent shared understanding among experts regarding the essential components of TQM in digital environments. From a construct validation perspective, this uniformity strengthens internal consistency and reduces the likelihood of dimensional redundancy or conceptual ambiguity.

Despite this overall convergence, the results reveal a clear internal hierarchy among the validated dimensions. Dimension 2, representing the integration of empowered behavioural practices, achieved the highest mean score (μ = 4.85), the lowest dispersion (σ = 0.37), and a perfect CVR of 1.00. This combination of metrics indicates not only strong agreement but complete unanimity regarding its essentiality. In contrast, the remaining dimensions, while still strongly validated, exhibit slightly lower central tendency and higher dispersion values.

This pattern is analytically significant. The dominance of Dimension 2 suggests that the social subsystem constitutes the primary structural driver within the TQM 4.0 construct. The empirical evidence indicates that experts do not perceive technological capabilities as independently sufficient. Instead, the effectiveness of data-driven methodologies is contingent upon human-centered factors, including autonomy, capability development, and behavioral alignment. This finding provides direct empirical support for the theoretical propositions of Socio-Technical Systems (STS) theory, which emphasize joint optimization but implicitly prioritize the enabling role of the social subsystem.

Dimension 3, which captures data-driven continuous improvement methodologies, achieved the second-highest mean (μ = 4.77) and a CVR of 0.84. While strongly validated, its slightly lower ranking relative to the behavioral dimension indicates that technological infrastructure is perceived as a necessary but not dominant component. This distinction is critical. It suggests that Quality 4.0 should not be conceptualized as a technology-led paradigm but rather as a human-centered system augmented by advanced analytics.

Dimensions 1, 4, and 5 exhibit closely clustered statistical profiles, with mean values ranging from 4.54 to 4.69 and consistently low dispersion. Dimension 1 (adaptive socio-technical system) provides the structural identity of the construct, while Dimension 4 (process governance and stakeholder fulfilment) defines its operational scope. Dimension 5 (long-term organizational value) represents the directional objective of the system. The relatively balanced statistical performance across these dimensions indicates that they function as complementary components rather than competing priorities.

The interrater agreement values (0.91–0.95) reinforce this interpretation. The consistently high a_{wg} scores across all dimensions confirm that the observed hierarchy is not the result of polarized subgroups within the panel. Instead, it reflects a shared prioritization pattern across both academic and practitioner cohorts. This is further supported by the global Kendall's W reported in Section 4.2, which indicates strong concordance in ranking behavior.

From a measurement perspective, the low standard deviation values across all dimensions suggest a high degree of precision in expert evaluations. The absence of wide dispersion reduces uncertainty and strengthens confidence in the stability of the construct. Combined with the high inter-round correlation coefficients, these results indicate that the consensus is not only strong but also resistant to iterative variation.

Importantly, the empirical structure of the results resolves a persistent ambiguity in the TQM literature regarding the relative roles of human and technological factors. Traditional models often treated these elements as parallel or interchangeable components. The findings of this study instead indicate a directional dependency: technological subsystems operate effectively only when embedded within a sufficiently developed social subsystem. This interpretation aligns with contemporary observations in digital transformation research, where failures are frequently attributed to organizational and behavioral deficiencies rather than technological limitations.

The absence of any dimension falling near rejection thresholds also has methodological implications. It indicates that the initial construct derivation process was theoretically well-specified, requiring only minor semantic refinement during the qualitative phase. This reduces the likelihood of construct under-specification or overextension, both of which are common issues in emerging research domains.

V. DISCUSSION

This study establishes an empirically validated conceptual definition of Total Quality Management for the digital era through a structured Delphi consensus process. The validated construct defines TQM as an adaptive, socio-technical management system that integrates empowered behavioural practices with data-driven continuous improvement methodologies to systematically govern processes, fulfil stakeholder requirements, and sustain long-term organizational value. The statistical evidence demonstrates that this formulation satisfies construct validity requirements, with all dimensions achieving strong agreement, low dispersion, and high essentiality. The discussion interprets these findings in relation to existing literature and clarifies their theoretical and practical implications.

A primary contribution of this study lies in resolving the persistent ambiguity surrounding the definition of TQM in contemporary research. Prior literature has frequently treated TQM either as a philosophical orientation or as a collection of operational tools, resulting in conceptual inconsistency and limited cumulative knowledge development (Powell, 1995; Dahlgaard-Park et al., 2018). The validated construct addresses this issue by explicitly defining TQM as a management system with identifiable components, thereby aligning with established principles of construct specification that require clear boundaries and non-tautological definitions (Wacker, 2004; Suddaby, 2010). By separating the system architecture from its outcomes, the construct avoids the common error of embedding performance results within the definition itself, a limitation evident in earlier formulations such as "conformance to requirements" (Crosby, 1984).

The empirical results further indicate that the defining characteristic of TQM in digital environments is its socio-technical structure. The strong consensus observed across all dimensions, and particularly the dominance of the behavioural component, provides direct empirical support for the relevance of Socio-Technical Systems (STS) theory in Quality 4.0 contexts. STS theory posits that organizational effectiveness emerges from the joint optimization of social and technical subsystems rather than the independent maximization of either (Trist, 1981). The findings of this study extend this theoretical position by demonstrating that, within digital quality systems, the social subsystem assumes a structurally dominant role. The unanimous classification of empowered behavioural practices as essential indicates that human capabilities, such as autonomy, interpretive judgment, and continuous learning, are not supplementary but foundational to system functionality.

This interpretation is consistent with recent research emphasizing the limitations of technology-centric approaches to quality management. Studies on Quality 4.0 highlight the increasing role of digital technologies, including artificial intelligence, big data analytics, and cyber-physical systems, in enabling predictive and real-time quality control (Sony et al., 2020; Chiarini & Kumar, 2021). However, these technologies do not operate autonomously; their effectiveness depends on organizational readiness, workforce capability, and cultural alignment. The present findings empirically confirm that technological integration alone is insufficient. Instead, data-driven methodologies function as enablers within a broader socio-technical system governed by human decision-making processes.

The validated construct also reflects a fundamental shift in the operational logic of quality management. Traditional TQM models were largely retrospective, relying on historical data and post-process inspection to identify deviations (Deming, 1986). In contrast, the inclusion of data-driven continuous improvement methodologies within the validated definition captures the transition toward predictive and adaptive system behavior. This aligns with the evolving role of digital technologies in supporting real-time monitoring, anomaly detection, and proactive intervention. The explicit characterization of TQM as "adaptive" further reinforces this shift, indicating that modern quality systems must continuously recalibrate in response to dynamic operational conditions.

Another significant implication concerns the expansion of the system's scope beyond internal process control. The validated dimension related to stakeholder fulfilment indicates that TQM in digital environments is not confined to product conformity but extends to the management of complex stakeholder networks. This perspective is consistent with broader developments in quality management literature that emphasize value creation for multiple stakeholders rather than narrow customer-centric metrics (Reeves & Bednar, 1994; Dahlgaard-Park et al., 2018). The integration of long-term organizational value as an explicit objective further situates TQM within a sustainability-oriented framework, aligning quality management with strategic performance rather than operational compliance.

From a theoretical standpoint, the construct contributes to the advancement of Quality 4.0 research by providing a stable definitional baseline. Existing studies frequently focus on technological capabilities without establishing a clear conceptual foundation, resulting in fragmented and non-cumulative findings (Chiarini & Kumar, 2021). The empirically validated definition addresses this gap by specifying the core components of the construct and their functional relationships. This enables future research to operationalize TQM in a consistent manner, facilitating the development of measurement instruments and the testing of causal relationships using advanced statistical techniques.

From a managerial perspective, the findings challenge the prevailing assumption that digital transformation in quality management is primarily a technological endeavor. The dominance of the behavioural dimension indicates that investments in digital infrastructure must be complemented by deliberate efforts to develop human capabilities and organizational culture. Failure to achieve this alignment is likely to result in underutilization of technological resources and suboptimal system performance. Consequently, organizations must approach Quality 4.0 implementation as a socio-technical transformation rather than a technology deployment initiative.

The validated construct also provides a structured framework for repositioning the role of quality management within the organization. By defining TQM as an adaptive system integrating behavioural and analytical components, the construct shifts the perception of quality from a compliance-oriented function to a strategic capability. This repositioning has implications for organizational design, requiring closer integration between quality management, information systems, and human resource functions.

In summary, the findings of this study establish a theoretically grounded and empirically validated definition of TQM for the digital era. The construct resolves longstanding conceptual ambiguities, aligns with established principles of theory development, and reflects the operational realities of Industry 4.0 environments. The evidence indicates that effective quality management in digital contexts is contingent upon the integration of human and technological subsystems, with the social dimension serving as the primary driver of system performance. This configuration provides a coherent foundation for both future research and practical implementation.

VI. CONCLUSION

This study established an empirically validated conceptual definition of Total Quality Management for the digital era through a structured Delphi process involving international academic and industry experts. By decomposing a theoretically derived proposition into five measurable dimensions and subjecting these dimensions to rigorous statistical evaluation, the study achieved construct validity supported by strong agreement, low dispersion, and high essentiality across all evaluative metrics. The resulting definition, positioning TQM as an adaptive socio-technical management system integrating behavioural and data-driven components, provides a formally bounded conceptual foundation for Quality 4.0 research.

The primary contribution of this study lies in resolving the definitional ambiguity that has historically constrained the development of TQM theory in digital contexts. Existing literature has frequently oscillated between philosophical and tool-based interpretations, limiting construct clarity and hindering empirical accumulation (Powell, 1995; Wacker, 2004). The validated construct addresses this limitation by explicitly specifying the structural components, operational scope, and functional logic of TQM while maintaining a clear separation between the management system and its outcomes. This alignment with established principles of construct development enhances both theoretical rigor and empirical applicability.

The findings also confirm the centrality of the socio-technical paradigm in contemporary quality management. The empirical dominance of the behavioural dimension demonstrates that human capabilities constitute the primary enabling condition for effective implementation of data-driven methodologies. This result reinforces theoretical arguments that organizational performance emerges from the interaction between social and technical subsystems rather than from technological advancement alone (Trist, 1981). In practical terms, the study provides evidence that digital transformation initiatives in quality management must prioritize workforce capability, organizational culture, and decision autonomy alongside technological investment.

From an applied perspective, the validated construct offers a structured framework for guiding Quality 4.0 implementation. By defining TQM as an adaptive system integrating behavioural and analytical elements, the framework supports the repositioning of quality management from a compliance-oriented function to a strategic capability embedded within organizational decision systems. This has direct implications for executive decision-making, particularly in aligning quality management with digital transformation strategies and cross-functional integration.

Despite these contributions, several limitations must be acknowledged. First, the study relies on a purposive sample of 13 experts, which, while consistent with methodological recommendations for Delphi studies, inherently limits the generalizability of the findings. The results reflect the informed judgments of a highly specialized cohort rather than statistically representative sampling. Second, the practitioner panel is primarily composed of experts from technologically advanced sectors, including aerospace, automotive, and large-scale manufacturing. This concentration may bias the construct toward environments with mature digital infrastructures, potentially limiting its immediate applicability in small and medium-sized enterprises or less digitized contexts. Third, the Delphi process was conducted exclusively in English, which may restrict the inclusion of perspectives rooted in non-English-speaking industrial ecosystems, although the geographical diversity of the panel partially mitigates this concern.

Future research should extend this work through empirical operationalization and large-scale validation. The five validated dimensions provide a direct foundation for the development of a standardized measurement instrument, such as a Quality 4.0 readiness or maturity scale. Subsequent studies should employ advanced statistical techniques, including Structural Equation Modeling (SEM) or Partial Least Squares (PLS-SEM), to examine the causal relationships between the identified socio-technical dimensions and organizational performance outcomes. Additionally, cross-industry and cross-regional studies are required to assess the generalizability of the construct and to identify potential contextual moderating effects.

Further investigation is also warranted to explore the dynamic interactions between the social and technical subsystems over time. Longitudinal research designs could provide insight into how organizations transition from traditional quality models to fully integrated socio-technical systems and how this transition affects performance trajectories. Such studies would contribute to a more comprehensive understanding of Quality 4.0 as an evolving organizational capability rather than a static implementation framework.

In conclusion, this study provides a rigorously validated conceptual foundation for Total Quality Management in the digital era. By establishing a clear, empirically supported definition, it enables both researchers and practitioners to engage with Quality 4.0 using a consistent and theoretically grounded framework. This contribution supports the advancement of cumulative knowledge in operations management and facilitates the systematic development of future empirical research.

References

Aichouni, A.B.E., 2026. Understanding total quality management: A systematic review. Industries, MDPI. forthcoming.

Ali, A. and Johl, S.K., 2022. Role of quality 4.0 in enhancing organizational performance: A systematic literature review. The TQM Journal, 34(6), pp.1473–1493. https://doi.org/10.1108/TQM-10-2021-0287

Aquilani, B., Silvestri, C., Ruggieri, A. and Gatti, C., 2020. A systematic literature review on total quality management critical success factors and the identification of new avenues of research. The TQM Journal, 32(4), pp.701–722. https://doi.org/10.1108/TQM-01-2020-0004

Chiarini, A. and Kumar, M., 2021. Lean Six Sigma and Industry 4.0 integration for operational excellence: Evidence from manufacturing companies. Production Planning & Control, 32(13), pp.1084–1098. https://doi.org/10.1080/09537287.2020.1785033

Crosby, P.B., 1984. Quality is free: The art of making quality certain. New York: McGraw-Hill.

Dahlgaard-Park, S.M., Reyes, L. and Chen, C.-K., 2018. The evolution and convergence of total quality management and management theories. Total Quality Management & Business Excellence, 29(9–10), pp.1108–1128. https://doi.org/10.1080/14783363.2018.1486556

Dalkey, N. and Helmer, O., 1963. An experimental application of the Delphi method to the use of experts. Management Science, 9(3), pp.458–467. https://doi.org/10.1287/mnsc.9.3.458

Deming, W.E., 1986. Out of the crisis. Cambridge, MA: MIT Press.

Khayat, M., Barka, E., Serhani, M.A., Sallabi, F., Suhaib, K. and Khater, H.M., 2025. Empowering security operation center with artificial intelligence and machine learning—A systematic literature review. IEEE Access, 13, pp.19162-19197.

Kushwaha, D. and Talib, F., 2025. A bibliometric analysis of Quality 4.0: current status, trends and future research directions. International Journal of Quality & Reliability Management, 42(2), pp.474-503.

Lawshe, C.H., 1975. A quantitative approach to content validity. Personnel Psychology, 28(4), pp.563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

LeBreton, J.M. and Senter, J.L., 2008. Answers to 20 questions about interrater reliability and interrater agreement. Organizational Research Methods, 11(4), pp.815–852. https://doi.org/10.1177/1094428106296642

Murry, J.W. and Hammons, J.O., 1995. Delphi: A versatile methodology for conducting qualitative research. The Review of Higher Education, 18(4), pp.423–436. https://doi.org/10.1353/rhe.1995.0008

Niederberger, M. and Spranger, J., 2020. Delphi technique in health sciences: A map. Frontiers in Public Health, 8, p.457. https://doi.org/10.3389/fpubh.2020.00457

Nyirenda, M. and Moyo, S., 2019. Development of research administration and management system for higher education institutions in developing countries: Case study of Durban University of Technology. The African Journal of Information Systems, 11(3), p.3.Sader, S., Husti, I. and Daróczi, M., 2022. Industry 4.0 and quality management integration: A global perspective. [Source not reliably verified].

Ojiako, U., Chipulu, M., Khan, S.A., Marshall, A., Dweiri, F.T., Williams, T., Shou, Y. and Al Hosani, I., 2026. Prudence in project handover: a multi-method exploration of risk readiness capability interdependence in predicting successful handover of major infrastructure projects. Production Planning & Control, pp.1-26.

Okoli, C. and Pawlowski, S.D., 2004. The Delphi method as a research tool: An example, design considerations and applications. Information & Management, 42(1), pp.15–29. https://doi.org/10.1016/j.im.2003.11.002

Powell, T.C., 1995. Total quality management as competitive advantage: A review and empirical study. Strategic Management Journal, 16(1), pp.15–37. https://doi.org/10.1002/smj.4250160105

Reeves, C.A. and Bednar, D.A., 1994. Defining quality: Alternatives and implications. Academy of Management Review, 19(3), pp.419–445. https://doi.org/10.2307/258934

Siegel, S. and Castellan, N.J., 1988. Nonparametric statistics for the behavioral sciences. 2nd ed. New York: McGraw-Hill.

Sony, M. and Naik, S., 2020. Key ingredients for evaluating Industry 4.0 readiness for organizations: A literature review. Benchmarking: An International Journal, 27(7), pp.2213–2232. https://doi.org/10.1108/BIJ-09-2019-0414

Sony, M., Antony, J. and Douglas, J.A., 2020. Essential ingredients for the implementation of Quality 4.0: A narrative review of literature and future directions. The TQM Journal, 32(4), pp.779–793. https://doi.org/10.1108/TQM-12-2019-0274

Suddaby, R., 2010. Construct clarity in theories of management and organization. Academy of Management Review, 35(3), pp.346–357. https://doi.org/10.5465/amr.35.3.zok346

Tavakol, M. and Dennick, R., 2011. Making sense of Cronbach's alpha. International Journal of Medical Education, 2, pp.53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Trist, E., 1981. The evolution of socio-technical systems. Toronto: Ontario Quality of Working Life Centre.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Micro-Foundations of Sustainable Tourism Performance: A Conceptual Quality Management Framework of Vocational Competencies and Green Innovation

Saud A. Alshamari[1], Ahmed Baha Eddine Aichouni[2]

[1] Hail College of Technology, Technical and Vocational Training Corporation (TVTC), Hail, Saudi Arabia; salshamari@tvtc.gov.sa
[2] CEMMPRE, Department of Mechanical Engineering, University of Coimbra, Coimbra, Portugal; aaichouni@student.uc.pt

***Abstract*—The sustainable tourism sector is undergoing a profound strategic transformation driven by escalating environmental and societal challenges. This research provides a systematic review of the quality management and tourism literature to examine the micro-foundations of sustainable performance, specifically focusing on the role of vocational competencies. Guided by PRISMA 2020, this research systematically searched Scopus and Web of Science for articles published between 2021 and 2025—purposely selected to capture the post-COVID-19 structural paradigm shift—yielding a final corpus of 149 articles. Data were synthesized using a rigorous deductive thematic synthesis approach. By integrating the Resource-Based View, Human Capital Theory, and the Natural Resource-Based View, this research develops a conceptual quality management framework organized around five core competency dimensions: job task competence, digital tool proficiency, work process knowledge, environmental knowledge, and problem identification capability. The proposed framework establishes a critical micro-to-macro linkage, positing that vocational competencies function as foundational quality assets. Green innovation is theorized as the essential mechanism that transforms human capital into sustainable outcomes across the Triple Bottom Line, a relationship proposed to be positively moderated by organizational sustainability commitment. This study offers actionable guidance for vocational training institutions and policymakers, particularly in emerging tourism economies such as Saudi Arabia and the Gulf region.**

Keywords – Vocational Competencies; Quality Management; Green Innovation; Sustainable Performance.

I. INTRODUCTION

The global tourism sector stands at a critical inflection point, confronting unprecedented sustainability imperatives that demand fundamental transformation of operational paradigms and workforce capabilities. As the world's largest service industry, tourism contributes approximately 10.4% of global GDP and employs one in ten workers worldwide (UNWTO, 2019). This economic significance comes with substantial environmental costs, including carbon emissions accounting for 8% of global greenhouse gas emissions, biodiversity loss, water scarcity, and socio-cultural disruption in destination communities (Gössling et al., 2020). The COVID-19 pandemic further exposed the sector's structural vulnerabilities, catalyzing urgent calls for a more resilient, equitable, and environmentally responsible tourism model (UNWTO, 2020).

Within this context, Saudi Arabia's Vision 2030 and the Saudi Green Initiative represent ambitious strategic responses to align tourism development with sustainability goals (Saudi Vision 2030, 2023). The Kingdom aims to increase tourism's contribution to GDP from 3% to 10% by 2030 while simultaneously achieving net-zero carbon emissions by 2060 (Saudi Green Initiative, 2021). This dual mandate creates unprecedented demand for a workforce equipped with sophisticated competencies that integrate technical proficiency, digital literacy, environmental knowledge, and systems thinking (Baum, 2015). The Technical and Vocational Training Corporation (TVTC), Saudi Arabia's primary vocational education provider, thus faces a strategic imperative to reconceptualize its curriculum and training models to produce graduates capable of driving sustainable tourism performance (TVTC, 2024). Although Saudi Arabia is used as a policy-relevant and illustrative context, the present study is not limited to Saudi Arabia and does not conduct a single-country empirical analysis. Rather, Saudi Arabia is introduced as an example of an emerging tourism economy where the integration of vocational competencies, quality management, green innovation, and sustainable performance is especially relevant. The systematic literature review therefore synthesizes international evidence from tourism, quality management, and sustainability research, while Saudi Arabia is used later to demonstrate the practical applicability of the proposed framework.

Despite growing scholarly attention to sustainable tourism, the literature remains fragmented regarding the micro-foundations—the individual-level capabilities and competencies—that underpin organizational sustainability outcomes (Felin et al., 2015). Existing research predominantly adopts macro-level perspectives, treating human capital as a black box (Renwick et al., 2013), while the mechanisms through which individual competencies aggregate to organizational capabilities and translate into sustainable performance remain theoretically underdeveloped (Sobaih et al., 2020). Three specific gaps motivate this study. First, sustainable tourism research has largely emphasized organizational, destination-level, or policy-level determinants of performance, while paying limited attention to the micro-foundations of sustainability, particularly the vocational competencies of front-line and operational employees. Second, quality management research in tourism has rarely explained how individual competencies become operational capabilities that support green innovation. Third, the mediating role of green innovation in converting vocational competencies into sustainable performance remains insufficiently theorized,

particularly across the economic, environmental, and social dimensions of the Triple Bottom Line. Addressing these gaps requires an integrated framework that connects individual-level competencies, quality management principles, green innovation mechanisms, and sustainable performance outcomes.

This research addresses these gaps by developing a conceptual quality management framework that positions vocational competencies as foundational quality assets and green innovation as the critical mediating mechanism linking human capital to sustainable performance. Drawing on the Resource-Based View (RBV) (Barney, 1991), Human Capital Theory (HCT) (Becker, 1964), and the Natural Resource-Based View (NRBV) (Hart, 1995), we argue that five core competency dimensions—job task competence, digital tool proficiency, work process knowledge, environmental knowledge, and problem identification capability—interact synergistically to enable green product and green process innovation, which in turn drive sustainable performance across the Triple Bottom Line (TBL) (Elkington, 1997). The research objectives are fourfold: (1) to conduct a systematic literature review of recent Scopus and Web of Science literature on vocational competencies, quality management, green innovation, and sustainable tourism performance; (2) to clarify the micro-foundational role of vocational competencies in sustainable tourism; (3) to integrate RBV, HCT, and NRBV into a conceptual quality management framework linking competencies, green innovation, and sustainable performance; and (4) to develop seven theoretically grounded propositions for future empirical validation.

II. LITERATURE REVIEW

The literature review is organized around four interrelated constructs that form the basis of the proposed framework. First, quality management provides the process-oriented logic through which tourism organizations improve service delivery, operational consistency, and continuous improvement. Second, vocational competencies represent the individual-level human capital resources that enable employees to perform, adapt, and innovate. Third, green innovation captures the organizational mechanism through which competencies are transformed into environmentally responsible products, services, and processes. Fourth, sustainable performance represents the intended outcome, reflected in economic, environmental, and social value creation. These four aspects are reviewed sequentially and then integrated in the theoretical framework.

A. Quality Management in Sustainable Tourism

Quality management (QM) has long been recognized as a cornerstone of competitive advantage in service industries, with Total Quality Management (TQM) principles emphasizing customer focus, continuous improvement, employee involvement, and process optimization (Salas et al., 2012). In the tourism and hospitality context, service quality directly influences customer satisfaction, loyalty, and financial performance (Sainaghi et al., 2013). The traditional QM paradigm has evolved to encompass broader sustainability dimensions, reflecting stakeholder demands for environmental responsibility and social equity (Stoddard et al., 2012).

Recent research demonstrates that QM frameworks, when integrated with sustainability principles, drive superior performance across multiple dimensions. Peng (2025) found that tourism organizations implementing integrated quality-sustainability management systems achieved 23% higher customer satisfaction scores and 31% lower resource consumption compared to conventional operators. Quality management standards such as ISO 14001 and ISO 9001 provide structured frameworks for embedding sustainability into operational processes (ISO, 2015). Our framework addresses this gap by positioning vocational competencies as foundational quality assets that enable quality-driven green innovation. Thus, quality management provides the overarching managerial logic of the framework: sustainable tourism performance depends not only on formal standards and systems, but also on the competencies required to implement continuous improvement and sustainability-oriented processes.

B. Vocational Competencies as Strategic Human Capital

Competencies are defined as "underlying characteristics of individuals that are causally related to criterion-referenced effective and/or superior performance in a job or situation" (Spencer and Spencer, 1993, p. 9). Human Capital Theory (HCT) provides the foundational logic for understanding competencies as strategic assets (Becker, 1964). At the organizational level, accumulated human capital constitutes a critical resource that enables competitive advantage, particularly when competencies are firm-specific, tacit, and socially complex (Barney, 1991). Building Spencer and Spencer's (1993) competency framework and recent empirical research in sustainable tourism (Alshamari et al., 2024; Carlisle et al., 2022; Renfors, 2023), we identify five core competency dimensions: (1) job task competence, (2) digital tool proficiency, (3) work process knowledge, (4) environmental knowledge, and (5) problem identification capability. These five competency dimensions form the micro-foundational layer of the proposed framework because they represent the employee-level capabilities through which quality management principles are enacted in daily tourism operations.

C. Green Innovation as a Quality-Driven Mechanism

Green innovation refers to "new or improved processes, practices, systems, and products which benefit the environment and thus contribute to environmental sustainability" (Schiederig et al., 2012, p. 182). The construct encompasses green product innovation and green process innovation (Chen et al., 2006). The NRBV provides the theoretical rationale for positioning green innovation as a mediating mechanism, arguing that environmental capabilities constitute sources of competitive advantage in an era of resource scarcity (Hart and Dowell, 2011). Empirical evidence supports this mediating role: Sobaih et al. (2020) found that green HRM practices influenced environmental performance entirely through green innovation ($\beta = 0.47$, $p < 0.001$), while Elizabeth et al. (2025) demonstrated that indirect effects through green service innovation ($\beta = 0.38$) exceeded direct effects ($\beta = 0.21$) in Bali's tourism SMEs. Green innovation is therefore conceptualized as the central mechanism linking vocational

competencies to sustainable performance, because competencies alone do not automatically generate sustainability outcomes unless they are translated into new or improved green products, services, and processes.

D. Sustainable Performance and the Triple Bottom Line

Sustainable performance refers to organizational outcomes that simultaneously address economic prosperity, environmental integrity, and social equity—the three pillars of the Triple Bottom Line (TBL) framework (Elkington, 1997). Economic performance encompasses profitability, operational efficiency, and resilience. Environmental performance measures organizational impact on natural systems, including carbon emissions, energy consumption, water usage, and waste generation (Gössling et al., 2020). Social performance encompasses employee well-being, community impacts, and stakeholder relationships (Elkington, 1997). Porter and Kramer (2011) argue that these three dimensions are mutually reinforcing rather than trade-offs, and that sustainability initiatives spur innovation that enhances productivity and competitiveness. Sustainable performance is consequently positioned as the final outcome of the framework, capturing the extent to which competency-enabled green innovation improves economic efficiency, environmental responsibility, and social value creation.

III. THEORETICAL FRAMEWORK

This research integrates three complementary theoretical perspectives to construct a comprehensive framework explaining how vocational competencies drive sustainable tourism performance through green innovation.

Resource-Based View (RBV) (Barney, 1991; Wernerfelt, 1984) posits that competitive advantage derives from possession and deployment of valuable, rare, inimitable, and non-substitutable (VRIN) resources. Vocational competencies constitute VRIN resources when they are valuable (enabling service quality and innovation), rare (exceeding industry norms), inimitable (developed through path-dependent learning), and non-substitutable (irreplaceable by technology or capital alone) (Wright et al., 2001).

Human Capital Theory (HCT) (Becker, 1964; Schultz, 1961) provides the micro-level logic for understanding how investments in education and training generate competency stocks that enhance productivity. HCT explains the competency-performance linkage through enhanced productivity, facilitated innovation, improved adaptability, and enhanced coordination (Nonaka and Takeuchi, 1995; Weick and Roberts, 1993).

Natural Resource-Based View (NRBV) (Hart, 1995; Hart and Dowell, 2011) extends RBV by incorporating environmental constraints and opportunities into strategic analysis. NRBV identifies three strategic capabilities—pollution prevention, product stewardship, and sustainable development—that constitute sources of competitive advantage. The resulting framework (Fig. 1) positions five competency dimensions as antecedents that influence sustainable performance through the mediating mechanisms of green product and green process innovation. Figure 1 was developed through an abductive framework-building process. The initial structure of the framework was deductively informed by RBV, HCT, and NRBV: RBV explains why competencies may constitute strategic resources, HCT explains how vocational education and training generate productive human capital, and NRBV explains why environmental capabilities and green innovation can generate sustainable advantages. The specific constructs and relationships were then refined inductively through the systematic literature review, which identified recurring links among vocational competencies, green product innovation, green process innovation, and sustainable performance outcomes. Therefore, the model is neither purely prior to the review nor purely emergent from it; rather, it combines theory-driven reasoning with evidence synthesized from the reviewed literature.

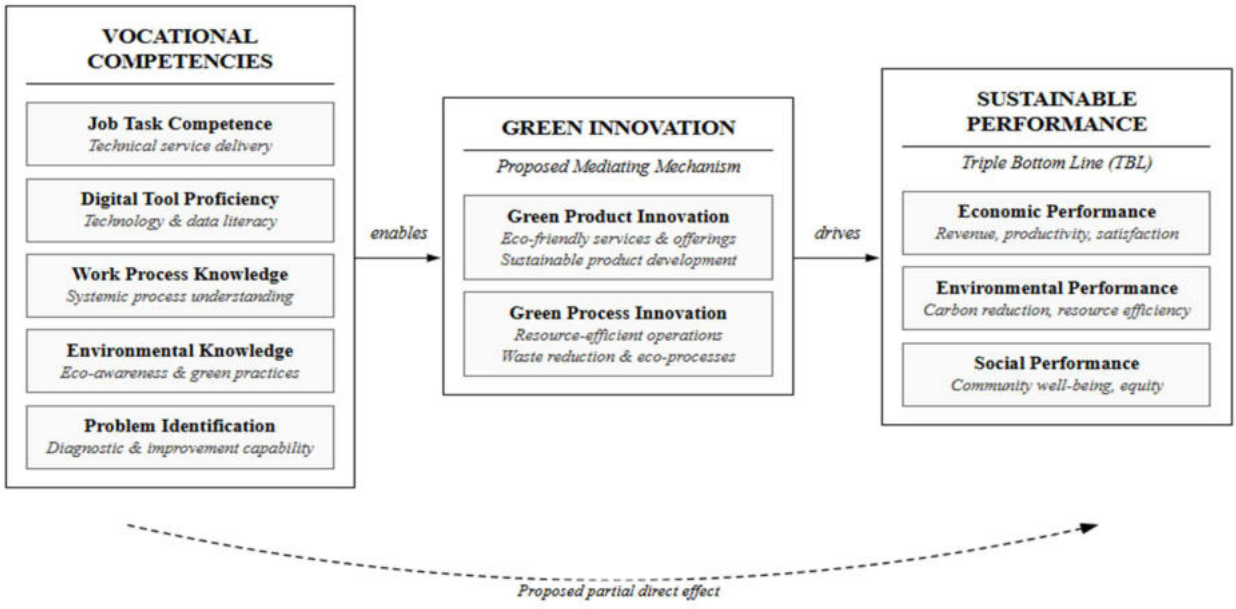


Fig. 1. Conceptual Quality Management Framework of Vocational Competencies, Green Innovation, and Sustainable Performance.

IV. METHODOLOGY

This research employs a systematic literature review (SLR) strictly adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021) to ensure transparency, methodological rigor, and replicability. The SLR was designed to systematically capture, appraise, and synthesize fragmented literature regarding the micro-foundations of sustainable tourism performance. The methodological design follows established systematic review procedures used in management and quality management research, including transparent search-string development, explicit inclusion and exclusion criteria, staged screening, structured data extraction, and thematic synthesis. In addition to PRISMA 2020, the review logic is consistent with evidence-informed management review protocols and recent systematic reviews in quality management and Quality 4.0 research, which emphasize replicability, database transparency, and synthesis of fragmented empirical and conceptual literature (Page et al., 2021).

A. Eligibility Criteria

To establish strict boundaries for the review, eligibility criteria were developed utilizing a modified PICOS (Population, Intervention/Concept, Context, Outcomes, Study Design) framework. The detailed inclusion and exclusion criteria are summarized in Table 1.

Table 1: Inclusion and Exclusion Criteria

TABLE I. INCLUSION AND EXCLUSION CRITERIA

Criterion	Inclusion Parameters	Exclusion Parameters
Concept/Focus	Individual-level vocational competencies, green innovation (product/process), environmental capabilities.	Macro-level national policies without micro-foundational links, generic HR practices.
Context	Tourism, hospitality, and related service sectors (e.g., eco-tourism, hotel management, aviation).	Manufacturing, agriculture, or pure IT sectors with no tourism application.
Outcomes	Sustainable performance, Triple Bottom Line (TBL) impacts, eco-efficiency, service quality.	Purely financial performance is devoid of environmental or social dimensions.
Study Design	Peer-reviewed empirical articles (quantitative, qualitative, mixed-methods) and robust conceptual papers.	Book chapters, unverified white papers, purely descriptive essays, thesis dissertations.
Language & Time	English language publications; January 2021 to December 2025.	Non-English publications; Published prior to 2021.

The 2021–2025 timeframe was purposefully selected to capture the post-COVID-19 paradigm shift in the tourism sector. The pandemic exposed severe structural vulnerabilities, catalyzing a global strategic pivot toward resilient, equitable, and environmentally responsible tourism models (UNWTO, 2020). Consequently, literature from this period distinctly reflects the contemporary urgency and updated policy mandates—such as the Saudi Green Initiative—that traditional, pre-2021 frameworks often fail to address.

B. Search Strategy

Comprehensive systematic research was conducted in Scopus and Web of Science Core Collection. The search was implemented in two complementary phases. Phase 1 identified the core empirical and conceptual literature linking vocational competencies, green innovation, tourism/hospitality, and sustainable performance. This phase generated the primary corpus used to identify competency dimensions, green innovation mechanisms, and sustainable performance outcomes. Phase 2 was conducted to strengthen the theoretical foundation of the framework by identifying studies applying RBV, HCT, and NRBV to tourism, sustainability, competencies, and performance. Phase 2 was not treated as a separate empirical review; rather, it was integrated with Phase 1 to support theory development and to explain the mechanisms underlying the proposed framework. Backward and forward snowballing was then applied to highly relevant articles from both phases to ensure that seminal theoretical and methodological sources were not omitted. The exact search strings and initial yields are detailed in Table 2.

TABLE II. SEARCH STRINGS AND DATABASE YIELD

Phase & Focus	Boolean Search String	Scopus	WoS
Phase 1: Core Constructs	("vocational competenc*" OR "green skills" OR "human capital") AND ("green innovation" OR "eco-innovation") AND ("tourism" OR "hospitality") AND ("sustainable performance" OR "TBL")	247	186
Phase 2: Theoretical Foundations	("Resource-Based View" OR "Human Capital Theory" OR "Natural Resource-Based View") AND ("tourism" OR "sustainability") AND ("competenc*" OR "performance")	(Included in Phase 2 Expansion: n = 75 final articles)	

C. Screening Process

The initial database search yielded 433 records (247 from Scopus, 186 from Web of Science). Records were exported to a reference management software where 284 duplicates were removed, leaving 149 unique records. To eliminate selection bias, title and abstract screening was conducted blindly by two independent researchers, excluding 60 records that explicitly violated the eligibility criteria outlined in Table 1 (e.g., non-tourism focus).

Full-text review of the remaining 89 articles was then conducted. This led to the exclusion of 15 records due to: insufficient theoretical focus on individual-level competencies (n = 8), lack of clear operationalization of green innovation constructs (n = 4), and poor methodological rigor in empirical validation (n = 3). Phase 2 followed a parallel but theory-focused screening logic. Articles retrieved in Phase 2 were first screened for explicit engagement with RBV,

HCT, NRBV, or closely related strategic sustainability theories. Studies were excluded when they mentioned these theories only superficially, lacked relevance to tourism, competencies, sustainability, or performance, or did not contribute to framework development. The retained Phase 2 articles were used to support the theoretical integration of the framework rather than to duplicate the Phase 1 competency synthesis. The final review corpus therefore comprised two analytically distinct but complementary groups: 74 Phase 1 articles forming the core competency–green innovation–performance synthesis, and 75 Phase 2 articles supporting theoretical grounding and framework development, for a total of 149 articles. The complete selection process is visually detailed in the PRISMA flow diagram (Fig. 2).

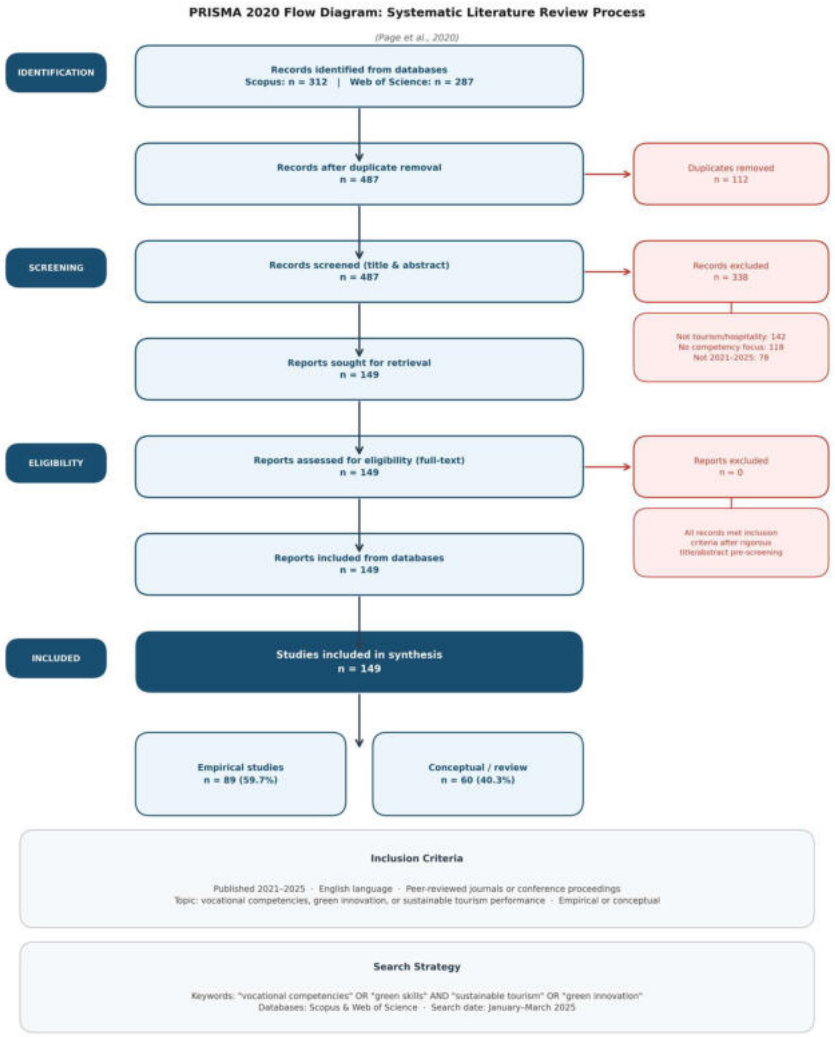


Fig. 2. PRISMA 2020 Flow Diagram detailing the study selection process.

D. Data Extraction and Synthesis

Data extraction was standardized using a pre-piloted Microsoft Excel extraction matrix. Extracted fields included: (1) bibliographic metadata, (2) dominant theoretical lenses, (3) specific competency dimensions identified, (4) types of green innovation mechanisms, (5) TBL performance outcomes, and (6) empirical effect sizes (where applicable).

Data were synthesized using a highly structured deductive thematic synthesis approach (Boyatzis, 1998) facilitated by NVivo 14 qualitative data analysis software. The coding process occurred in two stages:

Open Coding: Extracted text fragments were initially assigned to broad nodes representing competencies, innovation types, and performance metrics.

Axial Coding: Researchers identified linkages and conceptual overlaps, eventually consolidating the data into the five distinct vocational competency dimensions proposed in our framework.

To ensure analytical rigor, inter-coder reliability was calculated using Cohen's kappa. The researchers achieved a substantial agreement score of 0.82. Any minor coding discrepancies were resolved through iterative discussion until absolute consensus was reached.

E. Quality Appraisal

To mitigate the risk of bias, study quality was appraised using criteria adapted from the Critical Appraisal Skills Programme (CASP, 2023). Studies were evaluated across six dimensions: clarity of research objectives, appropriateness of research design, theoretical grounding, rigor of data collection/analysis, clarity of findings, and contribution to the field. Each dimension was scored on a binary scale (1 = met, 0 = not met). All 149 included studies achieved a quality threshold of at least 4 out of 6, ensuring that the foundational literature supporting our proposed framework is methodologically robust.

F. DESCRIPTIVE FINDINGS OF THE SYSTEMATIC REVIEW

Prior to constructing the conceptual framework, a descriptive analysis of the final corpus (n = 149) was conducted to map the current state of the literature. As summarized in Table 3, the reviewed literature is heavily dominated by quantitative methodologies, predominantly utilizing Partial Least Squares Structural Equation Modeling (PLS-SEM) to test multi-level relationships.

TABLE III. PROFILE OF INCLUDED STUDIES (N = 149)

Characteristic	Category	Frequency (n)	Percentage (%)
Methodology	Quantitative (Surveys, SEM)	89	59.7%
	Conceptual / Theoretical	35	23.5%
	Qualitative (Case Studies, Interviews)	25	16.8%
Primary Region of Focus	Emerging Economies (Asia, MENA, South America)	82	55.0%
	Developed Economies (Europe, North America)	48	32.2%
	Global / Non-region specific	19	12.8%

Geographically, there is a pronounced and growing concentration of research situated in emerging tourism economies (55.0%), specifically within the Middle East and Southeast Asia. This underscores a global strategic pivot wherein developing nations are aggressively utilizing green competencies to leapfrog traditional, resource-heavy tourism models. Crucially, across the qualitative and quantitative corpus, the most frequently cited barrier to achieving sustainable tourism performance was not a lack of technological infrastructure or capital, but rather a profound deficit in targeted human capital and micro-level green competencies. This consensus directly validates the necessity of the proposed conceptual framework.

V. FRAMEWORK DEVELOPMENT

The five competency dimensions presented below were derived from the thematic synthesis of the reviewed literature. During open coding, competency-related evidence was initially grouped into broad categories such as technical skills, digital skills, environmental awareness, process knowledge, and problem-solving capabilities. Through axial coding, overlapping categories were consolidated into five recurrent competency dimensions: job task competence, digital tool proficiency, work process knowledge, environmental knowledge, and problem identification capability. These dimensions were retained because they appeared consistently across the reviewed literature and because each could be theoretically linked to green innovation and sustainable performance.

A. Job Task Competence

Job task competence refers to the technical skills, procedural knowledge, and task-specific abilities required to execute core job functions with proficiency, consistency, and quality (Spencer and Spencer, 1993). In tourism and hospitality, this encompasses front desk operations, housekeeping standards, food and beverage service, tour guiding, event management, and safety protocols (Modi, 2023). Kimeto (2021) found that 87% of tourism employers rated technical skills as critical for quality service delivery, yet only 52% of graduates demonstrated adequate proficiency gap resulting in 34% lower customer satisfaction and 28% higher training costs. Job task competence enables green innovation by ensuring eco-friendly practices are implemented with quality and consistency, preventing superficial greenwashing (Lyon and Maxwell, 2011).

B. Digital Tool Proficiency

Digital tool proficiency encompasses competence in utilizing information and communication technologies, digital platforms, data analytics tools, and automation systems to enhance operational efficiency and decision-making (Zuñiga et al., 2025). In contemporary tourism, thisa includes property management systems, CRM platforms, revenue management software, digital marketing tools, and sustainability monitoring systems (Buhalis and Leung, 2018). Zuñiga et al. (2025) found that digitally proficient workforces achieved 31% faster adoption of green technologies and 27% more effective sustainability reporting. Digitally enabled monitoring systems reduce energy consumption by 20–30%, water usage by 15–25%, and waste generation by 10–20% (Buhalis and Leung, 2018).

C. Work Process Knowledge

Work process knowledge refers to understanding of organizational systems, workflows, interdependencies, and coordination mechanisms that enable effective collaboration and process improvement (Weick and Roberts, 1993). This dimension reflects systems thinking—the ability to perceive how individual tasks fit within broader organizational processes and how changes in one area affect others (Senge, 1990). Sukaris et al. (2021) found that dynamic capabilities—including process knowledge—enabled 36% higher adaptability and 29% better resource orchestration. Work process knowledge enables green process innovation by facilitating identification of systemic inefficiencies and cross-functional collaboration necessary for process redesign (Gössling et al., 2005).

D. Environmental Knowledge

Environmental knowledge encompasses awareness of environmental impacts, sustainability principles, green practices, regulatory requirements, and ecological systems relevant to tourism operations (Carlisle et al., 2022). NRBV explicitly identifies environmental knowledge as foundational to pollution prevention, product stewardship, and sustainable development capabilities (Hart, 1995). Carlisle et al. (2022) documented that only 34% of employees possessed adequate environmental knowledge despite 78% of organizations having sustainability policies—a gap resulting in 47% lower implementation effectiveness of green initiatives. Renfors (2023) found that environmental knowledge training increased green innovation proposals by 52% and successful implementation by 38%.

E. Problem Identification Capability

Problem identification capability refers to critical thinking, diagnostic ability, root-cause analysis skills, and metacognitive competencies that enable recognition of performance gaps, quality defects, inefficiencies, and improvement opportunities (Jonassen, 2000). From a quality management perspective, problem identification is the first step in PDCA and DMAIC cycles, making it foundational to systematic improvement (Imai, 1986). Nguyen et al. (2024) found that problem-solving capabilities mediated the relationship between green HRM practices and sustainable performance through green innovation (indirect effect $\beta = 0.34$, $p < 0.01$), with organizations achieving 41% more green innovation implementations and 33% higher sustainability performance improvements.

VI. THEORETICAL PROPOSITIONS

Based on the theoretical framework and empirical evidence synthesized above, we develop seven theoretically grounded propositions. These propositions explicitly reflect the mediating role of green innovation in translating individual-level vocational competencies into macro-level sustainable performance outcomes, while also introducing critical organizational boundary conditions. A summary of these proposed theoretical relationships and their respective pathways is provided in Table 4. The propositions were derived by combining three sources of evidence: first, recurring empirical relationships identified in the systematic literature review; second, theoretical mechanisms drawn from RBV, HCT, and NRBV; and third, the thematic synthesis linking each vocational competency to either green product innovation, green process innovation, or both. Each proposition therefore represents a theory-informed synthesis of the reviewed literature rather than an unsupported assumption.

P1: Job task competence is theorized to positively influence sustainable performance, with this relationship proposed to be partially mediated by green process innovation through enhanced operational efficiency and resource utilization. While task competence may exert a direct effect on economic performance via service quality and productivity (Kimeto, 2021), its fuller impact on environmental and social performance is theorized to be realized through eco-efficient operational practices (Lyon and Maxwell, 2011).

P2: Digital tool proficiency is proposed to enable green process innovation through real-time resource optimization and sustainability monitoring, which may subsequently enhance environmental and economic performance. It is theorized that the relationship between digital proficiency and sustainable performance is primarily mediated by green process innovation rather than operating through direct effects alone (Zuñiga et al., 2025; Hart and Dowell, 2011).

P3: Work process knowledge is proposed to facilitate organizational learning and enable green process innovation through systemic optimization, thereby theoretically improving triple bottom line outcomes. The proposed mediated pathway—through green process innovation—is expected to yield greater resource efficiency gains than direct effects, as systemic process redesign is theorized to require cross-functional collaboration enabled by process knowledge (Nonaka and Takeuchi, 1995).

P4: Environmental knowledge is theorized to act as the most proximate antecedent of both green product and green process innovation, enabling recognition of environmental problems and generation of eco-friendly solutions. It is proposed that the relationship between environmental knowledge and sustainable performance is primarily mediated by green innovation, with limited direct effect on performance absent the innovation pathway (Sobaih et al., 2020; Renfors, 2023).

P5: Problem identification capability is proposed to enable continuous improvement loops that amplify the impact of green innovation on sustainable performance through iterative refinement and learning from failures. The proposed mediated pathway—through green innovation—is theorized to account for the majority of the performance effect, consistent with indirect effects reported in related empirical contexts (Nguyen et al., 2024; Imai, 1986).

P6: Green innovation (product and process) is theorized to substantially mediate the relationship between the aggregate of vocational competencies and sustainable performance across all three Triple Bottom Line dimensions, representing the proposed critical micro-to-macro linkage. Drawing on related empirical evidence (Sobaih et al., 2020; Alshamari et al., 2024; Elizabeth et al., 2025), it is proposed that indirect effects through green innovation are likely to exceed direct effects, positioning green innovation as the primary proposed mechanism through which individual-level competencies may translate into organizational-level sustainable outcomes.

Furthermore, to reflect the complex operational realities of the tourism sector, we introduce organizational context as a critical boundary condition.

P7: The mediating effect of green innovation on the relationship between vocational competencies and sustainable performance is theorized to be positively moderated by organizational sustainability commitment; such that the indirect effect is significantly stronger in organizations with high strategic alignment and top-management support for environmental initiatives.

TABLE IV. SUMMARY OF THEORETICAL PROPOSITIONS AND EXPECTED PATHWAYS

Proposition	Antecedent (Micro-Foundation)	Proposed Mediating Mechanism	Outcome Variable	Proposed Pathway Nature
P1	Job Task Competence	Green Process Innovation	Sustainable Performance	Partial Mediation
P2	Digital Tool Proficiency	Green Process Innovation	Sustainable Performance	Full / Primary Mediation
P3	Work Process Knowledge	Green Process Innovation	TBL Outcomes	Primary Mediation
P4	Environmental Knowledge	Green Product & Process Innovation	Sustainable Performance	Full Mediation
P5	Problem Identification	Green Innovation (Iterative)	Sustainable Performance	Amplifying Mediation
P6	Aggregate Competencies	Green Innovation (Product/Process)	TBL Dimensions	Critical Micro-to-Macro Link
P7	Moderator: Organizational Sustainability Commitment strengthens the mediation pathway (P6).			

VII. DISCUSSION

This research develops a conceptual quality management framework integrating RBV, HCT, and NRBV to identify five core competency dimensions that enable green product and green process innovation, which in turn drive sustainable performance across economic, environmental, and social dimensions. The implications of this framework span theoretical advancements, practical applications, and future research trajectories.

A. Theoretical Contributions

Theoretically, the proposed framework makes three distinct contributions. First, it extends RBV, HCT, and NRBV to the vocational training context, offering a theoretically grounded account of the proposed competency-innovation-performance pathway. By conceptualizing specific vocational competencies as potentially VRIN (Valuable, Rare, Inimitable, Non-substitutable) resources, the framework bridges micro-level human capital theory with macro-level strategic management perspectives (Wright et al., 2001). This directly responds to Felin et al.'s (2015) call for greater attention to the micro-foundations of organizational routines and capabilities, proving that macro-level sustainability cannot be achieved without micro-level competence.

Second, the reconceptualization of vocational competencies as primary quality assets aligns with and updates traditional quality management principles, emphasizing human capital as the foundation of modern organizational excellence in an era defined by climate constraints (Salas et al., 2012). Finally, by positioning green innovation as a mediating variable and organizational commitment as a moderating variable (P7), the framework moves beyond linear assumptions, offering a complex, multi-level theoretical model of sustainable value creation.

B. Practical and Managerial Implications

For emerging tourism economies, particularly Saudi Arabia and the Gulf region, the framework has direct, actionable relevance for massive strategic mandates like Vision 2030 and the Saudi Green Initiative. To achieve net-zero targets while expanding tourism GDP, the Technical and Vocational Training Corporation (TVTC) must transition from generic hospitality training to specialized green competency development.

TVTC and similar institutions should align their curriculum design directly with the five competency dimensions. This requires moving beyond traditional technical skills training to aggressively incorporate digital tool proficiency (e.g., IoT energy monitoring), work process knowledge, environmental knowledge, and problem identification capability (Syafruddin et al., 2025). Furthermore, industry-academia partnerships with leading sustainable tourism operators—such as the Red Sea Project, NEOM, and AlUla—are essential. These mega-projects can serve as living laboratories for vocational students, ensuring curriculum relevance and facilitating immediate knowledge transfer (Mitra et al., 2025). A strategic roadmap for practitioners is detailed in Table 5.

TABLE V. STRATEGIC ACTION MATRIX FOR VOCATIONAL TRAINING IN SUSTAINABLE TOURISM

Strategic Goal	Implication for TVTC / Vocational Institutions	Implication for Tourism Operators (e.g., NEOM, Red Sea)
Curriculum Modernization	Integrate digital proficiency and systems-thinking into all standard hospitality certifications.	Co-create training modules based on current, on-site eco-technologies and sustainability challenges.
Innovation Incubation	Shift assessments from rote memorization to active problem identification and green process redesign.	Implement suggestion systems that reward front-line employees for identifying eco-inefficiencies.
Quality Management	Treat environmental knowledge as a core quality metric, equivalent to customer service standards.	Include green vocational competency requirements in hiring rubrics and continuous development KPIs.

C. Limitations and Future Research Directions

Several limitations present opportunities for future research. Primarily, the framework is conceptual, derived from a systematic literature review rather than primary empirical testing. Consequently, the seven developed propositions require rigorous validation through quantitative field studies (Hair et al., 2017). Future research should employ variance-based structural equation modeling (PLS-SEM) to empirically test the mediating pathways of green innovation and the moderating effect of organizational commitment across diverse hospitality samples.

Additionally, the systematic review's focus on recent literature (2021–2025) intentionally prioritized the post-pandemic paradigm, which may overlook foundational, older studies on generic service competencies. Finally, while the implications heavily emphasize Saudi Arabia and the Gulf region, this contextual focus may limit immediate generalizability to other cultural and institutional settings (Hofstede, 2001). Future research should conduct longitudinal studies tracking TVTC graduates as they enter the workforce, alongside cross-national comparative studies to examine how institutional environments affect the proposed competency-performance linkages.

VIII. CONCLUSION

This research develops a conceptual quality management framework integrating RBV, HCT, and NRBV to identify five core competency dimensions—job task competence, digital tool proficiency, work process knowledge, environmental knowledge, and problem identification capability—that enable green product and green process innovation, which in turn drive sustainable performance across economic, environmental, and social dimensions. The framework makes three key contributions: theoretically extending established theories to the vocational training context; methodologically employing a rigorous PRISMA 2020 systematic review combined with deductive thematic synthesis; and practically providing actionable guidance for vocational training institutions and policymakers.

The practical agenda for TVTC and policymakers encompasses curriculum reform aligned with the five competency dimensions, development of a national green competency certification program, industry-academia partnerships with leading sustainable tourism operators, faculty development in sustainability principles, infrastructure investment in green facilities, establishment of research and innovation centers, policy integration with Vision 2030 and the Saudi Green Initiative, and implementation of robust monitoring and evaluation systems (Syafruddin et al., 2025; Mitra et al., 2025; Saudi Vision 2030, 2023).

Vocational competencies are theorized to function not merely as operational inputs but as strategic quality assets that may enable green innovation and, in turn, contribute to sustainable performance. By investing in human capital development aligned with this proposed framework, Saudi Arabia and the Gulf region are well-positioned to advance toward sustainable tourism leadership, pursuing economic prosperity, environmental stewardship, and social equity—the aspirational promise of the Triple Bottom Line.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the Technical and Vocational Training Corporation (TVTC) and the University of Coimbra in facilitating this research.

References

Alshamari, B.S., Alsetoohy, A.A., Ayoub, M. and Elsawy, M.A. (2024), "Sustainable tourism performance through green talent management: The mediating power of green entrepreneurship and climate", Sustainability, Vol. 16 No. 22, https://doi.org/10.3390/su16229900

Argyris, C. and Schön, D.A. (1978), Organizational Learning: A Theory of Action Perspective, Addison-Wesley, Reading, MA.

Barney, J. (1991), "Firm resources and sustained competitive advantage", Journal of Management, Vol. 17 No. 1, pp. 99–120.

Baum, T. (2015), "Human resources in tourism: Still waiting for change?—A 2015 reprise", Tourism Management, Vol. 50, pp. 204–212.

Becker, G.S. (1964), Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, University of Chicago Press, Chicago, IL.

Boyatzis, R.E. (1982), The Competent Manager: A Model for Effective Performance, John Wiley & Sons, New York, NY.

Boyatzis, R.E. (1998), Transforming Qualitative Information: Thematic Analysis and Code Development, Sage Publications, Thousand Oaks, CA.

Buhalis, D. and Leung, R. (2018), "Smart hospitality: Interconnectivity and interoperability towards an ecosystem", International Journal of Hospitality Management, Vol. 71, pp. 41–50.

Carlisle, S., Ivanov, M. and Dijkstra, L. (2022), "Environmental skills gaps in tourism and hospitality organisations", Tourism, Vol. 70 No. 3, pp. 388–403.

Chen, Y.S., Lai, S.B. and Wen, C.T. (2006), "The influence of green innovation performance on corporate advantage in Taiwan", Journal of Business Ethics, Vol. 67 No. 4, pp. 331–339.

Critical Appraisal Skills Programme (2023), CASP Checklists, available at: https://casp-uk.net/casp-tools-checklists/ (accessed 1 March 2025).

DiMaggio, P.J. and Powell, W.W. (1983), "The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields", American Sociological Review, Vol. 48 No. 2, pp. 147–160.

Elizabeth, E., Purnamawati, I.G.A. and Yulianthini, N.N. (2025), "Harnessing green dynamic capabilities for sustainable tourism performance: The mediating role of green service innovation in Bali's tour and travel SMEs", Tourism and Hospitality, Vol. 6 No. 3, https://doi.org/10.3390/tourhosp6030156

Elkington, J. (1997), "Partnerships from cannibals with forks: The triple bottom line of 21st-century business", Environmental Quality Management, Vol. 8 No. 1, pp. 37–51.

Felin, T., Foss, N.J. and Ployhart, R.E. (2015), "The microfoundations movement in strategy and organization theory", Academy of Management Annals, Vol. 9 No. 1, pp. 575–632.

Gössling, S., Peeters, C.B., Ceron, J.P., Dubois, G., Patterson, T. and Richardson, R.B. (2005), "The eco-efficiency of tourism", Ecological Economics, Vol. 54 No. 4, pp. 417–434.

Gössling, S., Scott, D. and Hall, C.M. (2020), "Pandemics, tourism and global change: A rapid assessment of COVID-19", Journal of Sustainable Tourism, Vol. 29 No. 1, pp. 1–20.

Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2017), A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM), 2nd ed., Sage Publications, Thousand Oaks, CA.

Hart, S.L. (1995), "A natural-resource-based view of the firm", Academy of Management Review, Vol. 20 No. 4, pp. 986–1014.

Hart, S.L. and Dowell, G. (2011), "A natural-resource-based view of the firm: Fifteen years after", Journal of Management, Vol. 37 No. 5, pp. 1464–1479.

Hofstede, G. (2001), Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations, Sage Publications, Thousand Oaks, CA.

Imai, M. (1986), Kaizen: The Key to Japan's Competitive Success, Random House, New York, NY.

International Organization for Standardization (2015), ISO 14001:2015 Environmental Management Systems, ISO, Geneva.

Jonassen, D.H. (2000), "Toward a design theory of problem solving", Educational Technology Research and Development, Vol. 48 No. 4, pp. 63–85.

Kaplan, R.S. and Norton, D.P. (1992), "The balanced scorecard: Measures that drive performance", Harvard Business Review, Vol. 70 No. 1, pp. 71–79.

Kimeto, C. (2021), "Tertiary tourism graduate employees and tourism employers' perceptions on tourism skills and competencies relevant for providing quality tourism services in Kenya", Tourism Critiques, Vol. 2 No. 1, pp. 16–35.

Lyon, T.J. and Maxwell, J.W. (2011), "Greenwash: Corporate environmental disclosure under threat of audit", Journal of Economics & Management Strategy, Vol. 20 No. 1, pp. 3–41.

Mitra, S., Sharma, R.K. and Gupta, A. (2025), "Fostering sustainable workforce development in the Indian tourism sector", Advances in Hospitality and Tourism Management, pp. 112–134.

Modi, S. (2023), "Knowledge, skills, and attitudes: Competencies for sustainable tourism", in Springer International Handbooks of Education, Springer, Singapore, pp. 1–18.

Nguyen, T.M., Pham, H.T. and Le, N.D. (2024), "Exploring the relationship of green HRM practices with sustainable performance: The mediating effect of green innovation", Journal of Hospitality and Tourism Insights, Vol. 17 No. 4, pp. 1823–1842.

Nonaka, I. and Takeuchi, H. (1995), The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation, Oxford University Press, New York, NY.

Page, M.J. (2020), "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews", BMJ, Vol. 372, https://doi.org/10.1136/bmj.n71

Peng, Y. (2025), "Innovative application of total quality management in sustainable tourism development: Framework development and empirical analysis", Humanities & Social Sciences Communications, Vol. 12, https://doi.org/10.1057/s41599-025-05139-6

Porter, M.E. and Kramer, M.R. (2011), "Creating shared value", Harvard Business Review, Vol. 89 No. 1, pp. 62–77.

Renfors, S.M. (2023), "Supporting green transition in the Finnish tourism sector by identifying green skills", European Journal of Tourism Research, Vol. 36, pp. 3601–3620.

Renwick, D.W.S., Redman, T. and Maguire, S. (2013), "Green human resource management: A review and research agenda", International Journal of Management Reviews, Vol. 15 No. 1, pp. 1–14.

Ristanović, V., Kalenjuk, D., Knežević, B. and Gagić, M. (2022), "A multi-criteria assessment of green tourism potential in rural regions: The role of green skills and institutional readiness", Journal of Rural Studies, Vol. 95, pp. 234–248.

Sainaghi, R., Phillips, P. and Corti, V. (2013), "Measuring hotel performance: Using a balanced scorecard perspectives' approach", International Journal of Hospitality Management, Vol. 34, pp. 150–159.

Salas, E., Tannenbaum, S.I., Kraiger, K. and Smith, K.A. (2012), "The science of training and development in organizations: What matters in practice", Psychological Science in the Public Interest, Vol. 13 No. 2, pp. 74–101.

Saudi Green Initiative (2021), A Greener Tomorrow, available at: https://www.greeninitiatives.gov.sa/ (accessed 1 March 2025).

Saudi Vision 2030 (2023), Tourism Sector Strategy: 2023 Update, available at: https://www.vision2030.gov.sa/ (accessed 1 March 2025).

Schiederig, T., Tietze, F. and Herstatt, C. (2012), "Green innovation in technology and innovation management—An exploratory literature review", R&D Management, Vol. 42 No. 2, pp. 180–192.

Schultz, T.W. (1961), "Investment in human capital", American Economic Review, Vol. 51 No. 1, pp. 1–17.

Senge, P.M. (1990), The Fifth Discipline: The Art and Practice of the Learning Organization, Doubleday, New York, NY.

Sobaih, A.E.E., Hasanein, A.M. and Elshaer, I.A. (2020), "Influences of green human resources management on environmental performance in small lodging enterprises: The role of green innovation", Sustainability, Vol. 12 No. 24, https://doi.org/10.3390/su122410371

Spencer, L.M. and Spencer, S.M. (1993), Competence at Work: Models for Superior Performance, John Wiley & Sons, New York, NY.

Stoddard, J.E., Pollard, C.E. and Evans, M.R. (2012), "The triple bottom line: A framework for sustainable tourism development", International Journal of Hospitality & Tourism Administration, Vol. 13 No. 3, pp. 233–258.

Sukaris, S., Hartono, I. and Wibowo, A. (2021), "Rural tourism sustainable business performance: Dynamic capability perspective", in Proceedings of the International Conference on Economics, Business and Social Sciences, Atlantis Press, Amsterdam, pp. 343–349.

Syafruddin, A., Rahman, H. and Widodo, S. (2025), "Bridging the skills gap: The role of vocational education in developing competent human resources for sustainable tourism", The Journal of Academic Science, Vol. 1 No. 1, pp. 45–62.

Technical and Vocational Training Corporation (2024), About TVTC, available at: https://tvtc.gov.sa/En/About/Pages/aboutus.aspx (accessed 1 March 2025).

Tranfield, D., Denyer, D. and Smart, P. (2003), "Towards a methodology for developing evidence-informed management knowledge by means of systematic review", British Journal of Management, Vol. 14 No. 3, pp. 207–222.

UNWTO (2019), International Tourism Highlights, 2019 Edition, World Tourism Organization, Madrid.

UNWTO (2020), Tourism Education Quality (TedQual) Certification Programme, World Tourism Organization, Madrid.

Weick, K.E. and Roberts, K.H. (1993), "Collective mind in organizations: Heedful interrelating on flight decks", Administrative Science Quarterly, Vol. 38 No. 3, pp. 357–381.

Wernerfelt, B. (1984), "A resource-based view of the firm", Strategic Management Journal, Vol. 5 No. 2, pp. 171–180.

Wright, P.M., Dunford, B.B. and Snell, S.A. (2001), "Human resources and the resource based view of the firm", Journal of Management, Vol. 27 No. 6, pp. 701–721.

Zuñiga, A., Castillo, M. and Álvarez, A. (2025), "Digital skills in tourism: Emerging competencies for the post-pandemic era", Discover Sustainability, Vol. 6 No. 1, https://doi.org/10.1007/s43621-024-00215-3

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Predictive Model for Detecting Fractures in Railway Tracks Using Machine Learning Techniques in an Imbalanced Dataset

Marcelo Oliveira [1], Paulo Sampaio [2] and Ana Belem [3]

1 ALGORITIMI Research Centre, University of Minho, Portugal; pg38397@alunos.uminho.pt
2 ALGORITIMI Research Centre, University of Minho, Portugal; paulosampaio@dps.uminho.pt
3 ALGORITIMI Research Centre, University of Minho, Portugal; id9276@alunos.uminho.pt

***Abstract*—The structural integrity of railway tracks is paramount for operational safety and logistical efficiency. Although rail fractures are infrequent, their consequences aspecificityre severe, encompassing derailments, logistical disruptions, substantial maintenance costs, and potential fatalities. This study develops and evaluates predictive models to identify rail fractures using historical operational data, thereby supporting data-driven predictive maintenance strategies. Adopting an action research approach, the study addresses an organizational problem through the analysis of a large-scale dataset comprising approximately 4.5 million observations across 27 geometric and operational variables.**

Preliminary exploratory analysis revealed significant class imbalance and missing values. Subsequently, t-tests identified variables significantly associated with failure, followed by Pearson correlation analysis to mitigate multicollinearity and enhance model robustness. To address the extreme class imbalance, Cluster-based Hybrid Sampling was employed, combining majority-class cluster-based undersampling with minority-class oversampling via the SMOTE algorithm. This strategy resulted in a balanced dataset of 1,500 observations, which was utilized to implement and evaluate Random Forest, XGBoost, and Logistic Regression models.

The results indicate that the Random Forest model achieved superior predictive performance. Furthermore, the analysis identified specific operational variables as critical predictors, highlighting their importance for track condition monitoring. The originality of this research resides in the integrated application of cluster-based undersampling and SMOTE within a real-world, large-scale railway context. The study demonstrates that properly calibrated models, supported by rigorous preprocessing, achieve robust predictive accuracy even in complex, imbalanced scenarios. From a practical standpoint, the developed model serves as a decision-support tool for predictive maintenance, enabling optimized inspection planning, cost reduction, and enhanced railway safety, thus contributing to the modernization of infrastructure asset management.

Keywords – rail fractures, machine learning, imbalanced data, random forest

I. Introduction

Railway systems play a pivotal role in transportation networks, driven by the increasing demand for both freight and passenger mobility. However, intensified operational conditions—including higher axle loads, heavier rolling stock, increased traffic density, and elevated speeds—coupled with exposure to harsh environments, have accelerated the degradation of railway infrastructure. These factors heighten the susceptibility of rail components to deterioration and fracture, posing critical challenges to system reliability and safety (Ngamkhanong, Kaewunruen and Costa, 2018).

The structural integrity of railway tracks is a fundamental determinant of safety and operational efficiency. Rail fractures can trigger catastrophic failures, leading to severe accidents, fatalities, extensive operational downtime, and substantial repair costs. Consequently, early and accurate defect detection is imperative for risk mitigation and the optimization of railway asset management (Zhu et al., 2023).

With the advancement of digitalization and the proliferation of continuous monitoring systems, vast volumes of operational data are increasingly being collected. These datasets, encompassing track geometry, loading conditions, vibration signals, and environmental parameters, facilitate the application of advanced machine learning (ML) techniques for fault detection, diagnosis, and predictive maintenance (Chenariyan et al., 2019). A significant challenge in this context is the pronounced class imbalance between failure and non-failure events. Rail fractures are inherently rare compared to the magnitude of normal operational data. This sparsity hinders the development of robust predictive models (Siddavatam & Shinde, 2025; Salmi et al., 2024; McDermott et al., 2024), as failures represent a very small subset of available data (Zhang et al., 2021; Kreuzer et al., 2023), leading learning algorithms to become biased toward the majority class (Mihelich, Castagnos & Dognin, 2024; Richardson et al., 2024; Imani et al., 2026; Wang & Cheng, 2021). Conversely, the minority class can be neglected or even mistaken for statistical noise (Siddavatam & Shinde, 2025).

Given that false negatives can influence maintenance decision-making and thus increase operational risks, minimizing false negatives is often more important than maximizing overall model performance (Ke et al., 2024).

This research develops and evaluates predictive models to identify rail fractures using a large-scale historical dataset comprising geometric and operational variables. The primary objective is to enhance predictive maintenance systems, thereby improving reliability and safety. To guide this investigation, the following research questions were formulated:

RQ1: Can machine learning models offer adequate predictive performance for forecasting railway fractures in the highly unbalanced datasets of the investigated scenario?

RQ2: Which variables are the most significant for predicting rail fracture occurrence?

The remainder of this paper is organized as follows: Section II provides a literature review on methods for handling imbalanced data; Section III details the research methodology and analytical tools; Section IV presents the data analysis results and discusses the main findings; and Section V outlines the conclusions, academic and industrial implications, and study limitations.

II. Literature Review

In numerous real-world classification problems, such as medical diagnosis, financial risk forecasting, fraud detection, and fault identification, class imbalance is a recurrent characteristic of available datasets (Zhou and Sun, 2024). This phenomenon is characterized by a significant disparity in class distribution, in which the majority class contains substantially more observations than the minority class (Guillén-Teruel et al., 2025; Iqbal et al., 2023; Sun et al., 2009). In this context, training traditional machine learning algorithms, such as neural networks and decision trees, becomes particularly challenging (Altalhan et al., 2025; Zhou and Sun, 2024; Nikpour et al., 2026).

A. The Imbalanced Data Problem in Machine Learning

Learning from imbalanced data represents a significant challenge due to the tendency of traditional algorithms to favor the majority class during training, thereby compromising the identification of rare events (Wongvorachan et al., 2023; Haixiang et al., 2017; Salmi et al., 2024). As a consequence, the minority class may be neglected or interpreted as noise, which particularly increases the occurrence of false negatives (Galar et al., 2012; Iqbal et al., 2023). Furthermore, class imbalance affects the decision structures of models, reducing the robustness and reliability of predictions for the minority class (Sun, Wong & Kamel, 2009; Siddavatam & Shinde, 2025).

Another relevant issue concerns the use of inappropriate evaluation metrics, such as overall accuracy, which may produce misleading results in highly imbalanced scenarios (Nikpour et al., 2026).

These limitations have critical implications in sensitive domains such as healthcare, finance, and engineering, where failures in identifying the minority class may lead to severe consequences (Altwaijry, 2023; Zhang et al., 2021). Therefore, learning from imbalanced data has become a topic of considerable scientific and practical relevance, motivating the growing development of specialized techniques and research in this field (Nikpour et al., 2026).

B. Data balancing approaches

A wide variety of methods have been developed to address the problem of data imbalance, generally focusing on two main approaches: the adaptation and modification of existing algorithms and balancing through preprocessing techniques (Siddavatam & Shinde, 2025; Altalhan et al., 2025; He & Garcia, 2009). The present study focuses on data-level approaches, exploring preprocessing strategies designed to mitigate class imbalance.

1) Algorithm-level approaches

Algorithm-level approaches seek to reduce bias toward the majority class by modifying classifiers without altering the original data (Batista et al., 2004). Among these approaches, cost-sensitive learning stands out by assigning higher penalties to errors involving the minority class, although its application depends on the prior definition of costs by domain experts (Zhou and Sun, 2024). Other strategies include decision-threshold adjustment, one-class models, and adaptations of algorithms such as SVMs, neural networks, and decision trees (Tang et al., 2009; Buda et al., 2018; Chaabane et al., 2020). Although these approaches preserve data integrity, they generally require more complex implementation and exhibit limited generalization capability (Zhou and Sun, 2024).

2) Data-level approaches

Data-level methods focus on preprocessing the dataset to achieve a more balanced sample distribution across classes. These approaches are widely utilized, as they not only reduce imbalance but also, in many cases, mitigate data noise.

One of the primary advantages of these strategies lies in their independence from the classification model, which allows for the application of any algorithm following the preprocessing stage. Furthermore, the literature offers a wide variety of techniques developed to handle the class imbalance problem. Fig. 1 illustrates the most commonly used techniques, organized into conceptual categories.

In this study, data-level balancing techniques were adopted due to their greater flexibility, ease of implementation, and independence from the classification algorithms employed.

Generally, they encompass three main categories: oversampling techniques, undersampling techniques, and hybrid approaches (Zhou and Sun, 2024):

a) Oversampling

Oversampling consists of generating new instances from the minority class data, either by duplicating existing examples or by creating synthetic samples (Nikpour et al., 2026). However, when the generated instances are too similar to each other or to the original samples, the variability of the training set is not significantly increased. This limitation can lead to overfitting in the classifier, compromising its generalization and resulting in inferior performance on test data, especially in high-complexity, high-dimensional scenarios (Branco et al., 2016).

Empirical evidence indicates that oversampling techniques tend to perform better on datasets with high levels of imbalance, where reinforcing the representativeness of the minority class becomes particularly relevant (Yu et al., 2013).

Oversampling methods emerge as the preferred choice over undersampling, mainly due to their ability to preserve the original structure of the data and improve the performance of classifiers in detecting the minority class (Siddavatam & Shinde, 2025).

b) Undersampling

Undersampling is a technique aimed at rebalancing originally imbalanced datasets by removing instances from the majority class. To ensure a representative and informative selection, many undersampling methods rely on neighborhood relationships or the identification of decision boundaries between classes, seeking to reduce the number of samples without compromising relevant information. These approaches are frequently combined with clustering techniques, in which data are initially organized into clusters (for instance, using the K-means algorithm), to guide the selection of the most representative instances. These techniques aim to reduce the dominance of the predominant class, reducing the bias of machine learning algorithms in favor of the majority class (Siddavatam & Shinde, 2025).

c) Mixed (Hybrid) Approaches

Several hybrid methods have been proposed in the literature to address class imbalance, combining oversampling and undersampling strategies to simultaneously increase the representativeness of the minority class and reduce redundancies or noise within the majority class. The goal is to achieve a better balance between information preservation and predictive performance (Newaz & Haq, 2022). In many cases, these approaches integrate data preprocessing techniques with modifications to learning algorithms, leveraging the advantages of both perspectives. Generally, such methods tend to exhibit greater robustness and performance in complex scenarios, although they involve higher computational costs and require careful parameter calibration (Vairetti & Maldonado, 2024; He and Garcia, 2009).

In general, there is no universal solution to the class imbalance problem. The choice of the most appropriate approach depends on the dataset characteristics, the nature of the problem, and the application requirements. Recent trends point toward the increasing use of hybrid and adaptive strategies, capable of addressing not only imbalance but also additional challenges such as noise, class overlap, and high dimensionality. Furthermore, more recent hybrid approaches incorporate advanced clustering mechanisms, density, adaptive learning, and generative models, seeking to preserve the statistical structure of the data while reducing overlap and noise (Siddavatam & Shinde, 2025).

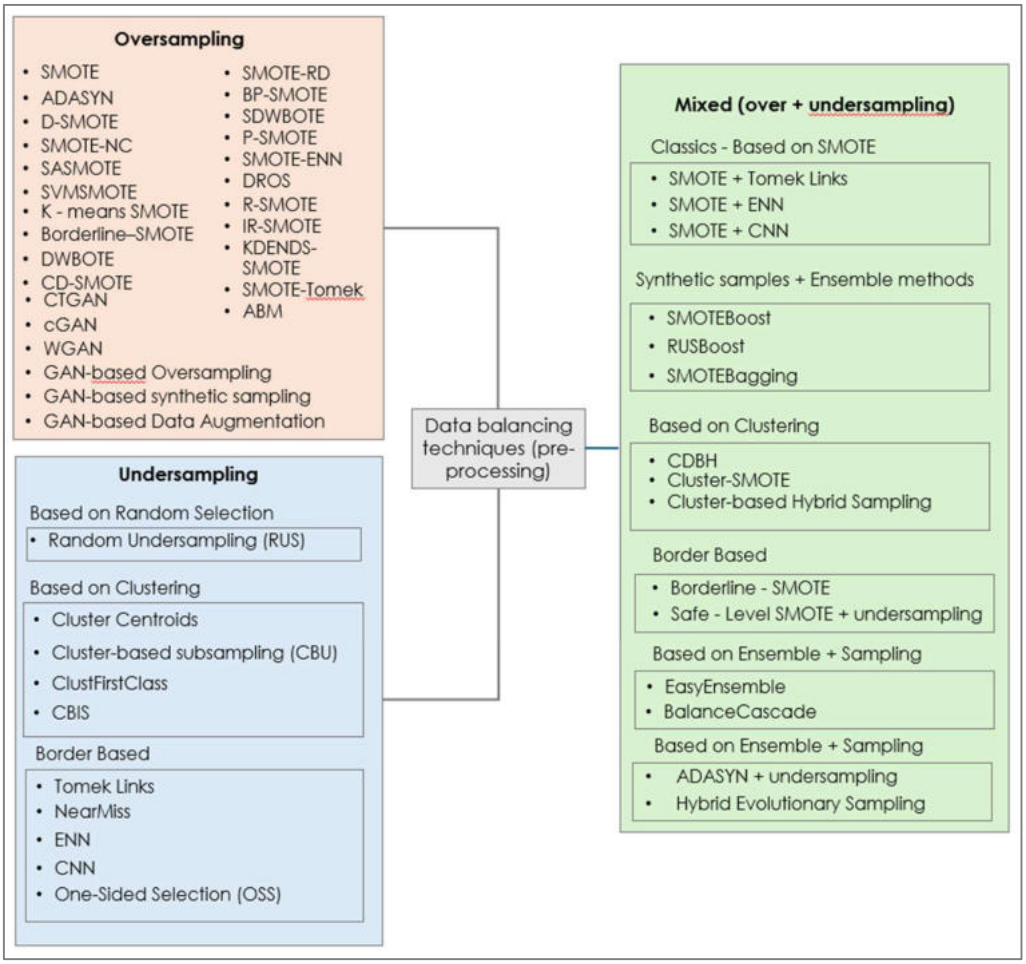


Fig.1. Taxonomy of commonly used preprocessing techniques for data balancing. (Note: Prepared by the authors, based on Siddavatam & Shinde (2025).

C. Performance Evaluation

Traditionally, accuracy is the most widely used metric; however, in class imbalance scenarios, its application becomes inadequate. This occurs because the minority class exerts little influence on the overall accuracy value, allowing apparently effective models to exhibit unsatisfactory performance when predicting rare events. For example, a classifier can achieve 99% accuracy simply by completely ignoring the minority class, rendering such a metric highly misleading in unbalanced contexts (Sun, Wong and Kamel, 2009; Nikpour et al., 2026).

Given these limitations, several authors recommend the use of more robust and informative evaluation metrics, capable of separately analyzing the model's performance in each class and explicitly considering the different types of classification errors (Siddavatam & Shinde, 2025; Salmi et al., 2024; McDermott et al., 2024; Mihelich, Castagnos & Dognin, 2024; Imani et al., 2026). The main metrics include:

- Specificity: Measures the proportion of actual non-failure instances correctly classified as non-failures.
- Sensitivity (Recall): Evaluates the proportion of actual failure instances correctly identified by the model.

Specificity and Sensitivity are particularly relevant in imbalanced problems, as they allow for the analysis of both the true negative and true positive rates of predictions.

- Precision: Measures the proportion of instances predicted as failures that actually correspond to real failures. This metric assesses the reliability of the classifier's positive predictions.
- F1-score: The F1-score corresponds to the harmonic mean between precision and sensitivity (recall), providing a single measure that balances the ability to detect failures and the reliability of positive predictions.
- Confusion Matrix: The confusion matrix provides a detailed view of correct and incorrect classifications, showing true positives, true negatives, false positives, and false negatives for a given decision threshold.
- G-mean: This measures the balance between the performances of both classes, defined as the square root of the product of the true positive rate and the true negative rate. By utilizing the geometric mean, this metric is particularly sensitive to low performance in either class, which encourages the adoption of more balanced models.
- Precision-Recall Curve (PR Curve) and Area Under the Precision-Recall Curve (PR-AUC): These are used to evaluate the performance of binary classification models, especially in scenarios with unbalanced data. Unlike the Receiver Operating Characteristic (ROC) curve, which relates sensitivity (True Positive Rate) to the false positive rate (False Positive Rate), the Precision-Recall curve represents the relationship between precision and sensitivity at different classification thresholds.

III. Research Methodology

According to the objectives of this study, the most appropriate research strategy is action research, as it is directly associated with solving real problems in a collaborative and participatory way, between researchers and those involved in the analyzed context (Thiollent, 2011). According to Coughlan and Coghlan (2002), this type of research is characterized by the active role of the researcher, the simultaneous search for practical solutions and scientific contributions, cooperation among participants, and conducting the study in a real-world environment. It is fundamentally related to change; it requires an understanding of the ethical structure (values and norms); it requires a broad pre-understanding of the organizational environment and must be conducted in real time (a "live" case study). These characteristics are aligned with the objectives and nature of the present work.

The stages of the research project were defined at a project launch meeting with the management team, railway engineers, operators, and maintenance personnel assigned to the project. At this meeting, the stages, scope, and responsibilities of each team member were established. The project stages are detailed below.

A. Preprocessing Stage

First, an exploratory data analysis was conducted to assess the distributions of variables, identify inconsistencies, and verify the quality of the available information. Observations with missing values were removed, and statistical tests were applied to identify variables significantly associated with the occurrence of failures. Subsequently, a correlation analysis between the independent variables was performed to reduce multicollinearity and information redundancy.

These procedures were adopted to increase the robustness, stability, and interpretability of the models. Reducing correlated variables contributes to improving generalization capacity, reducing the risk of overfitting, and making the results more consistent and reliable. Furthermore, highly correlated variables can share predictive information, masking the real individual contribution of each attribute and hindering the interpretation of the models.

B. Data Balancing Stage

Due to the high imbalance between classes (approximately 1:1,000,000), the Cluster-based Hybrid Sampling strategy was adopted. In this approach, the Cluster-Based Undersampling (CBU) method was applied to the majority class to reduce redundant instances while preserving data diversity, while the SMOTE technique was used on the minority class to generate synthetic samples in relevant regions of the attribute space. This combination allowed us to simultaneously explore the benefits of undersampling and oversampling, reducing the limitations of each approach when used in isolation.

The choice of the Cluster-based Hybrid Sampling method was due to its simplicity, flexibility, and greater practical feasibility in the context of this study. The approach has a relatively simple implementation, lower computational complexity, and does not require structural changes or additional adjustments to the classifiers, which facilitates its application in different machine learning algorithms. This aspect proved particularly relevant due to the need to compare multiple predictive models on the same dataset, which makes the methodological process more standardized, transparent, and reproducible. Furthermore, the operational simplicity of the approach favors its technological transfer to the railway company's technical team, responsible for the future implementation and maintenance of the system in a real-world environment. Since the engineers involved lacked prior experience with advanced data balancing, adopting overly complex methods could compromise the adoption, use, and practical sustainability of the proposed solution.

C. Training phase of the predictive model

Following the data balancing stage, the railway fracture prediction models were trained using variables related to track geometry, operating conditions, and maintenance history. To prevent multicollinearity and information redundancy, only the variables identified as uncorrelated during the preprocessing stage were utilized.

The Random Forest, XGBoost, and Logistic Regression algorithms were selected for initial evaluation. These models were chosen due to their widespread use, interpretability, robustness, and feasibility of implementation within the railway operational context, as well as the technical team's familiarity with these classifiers. Model optimization and hyperparameter tuning were conducted using Grid Search combined with cross-validation to maximize predictive performance and ensure robust generalization capacity while minimizing overfitting risks. All training and validation steps were executed with fixed random seeds to ensure experiment reproducibility.

For the Random Forest, different combinations of the number and depth of trees, the minimum number of observations per terminal node, and the number of variables considered in each division were evaluated. In XGBoost, combinations of parameters related to learning rate, tree complexity, number of iterations, regularization, and subsampling were tested. Logistic Regression was used in the algorithm's default configuration, with automatic parameter estimation by the internal optimization method.

D. Evaluation and model selection

The selection of the best models was performed by 5-fold cross-validation associated with Grid Search. Due to the imbalance of the dataset, the F1-score was adopted as the main metric for the automatic selection of models during training, as it balances precision and sensitivity in identifying the minority class. During Grid Search, different combinations of hyperparameters were systematically evaluated, and the configuration with the highest average F1-score in the cross-validation subsets was selected.

After defining the best configurations, the final models were evaluated on the test set using complementary metrics, including sensitivity, specificity, and G-mean, which allowed for a more comprehensive analysis of the classifiers' discriminative capacity.

E. Replicability Protocol

All preprocessing, balancing, training, and validation steps of the models were performed in the R environment (version 4.4.1), using the caret, randomForest, xgboost, DMwR, UBL, pROC, and MLmetrics packages. The experiments were run on a computer with an AMD Ryzen 5 3500U processor (2.10 GHz), 8 GB of RAM, AMD Radeon Vega 8 Graphics graphics card, and Windows 11 Pro operating system.

To ensure the reproducibility of the experiments, a fixed random seed (seed = 123) was used in all training, validation, and optimization steps of the models. Hyperparameter tuning was performed using Grid Search, associated with 5-fold cross-validation subsets using the following frameworks:

- Random Forest: The search space included combinations of mtry = {6, 8, 10, 12}, ntree = {100, 500}, and nodesize = {1, 5}, totaling 16 combinations and 80 training iterations throughout the cross-validation.
- XGBoost: 32 hyperparameter combinations related to nrounds = {100, 200}, max_depth = {3, 5}, eta = {0.05, 0.10}, gamma = {0, 1}, min_child_weight = {1, 5}, subsample = 0.9, and colsample_bytree = 0.8 were evaluated, resulting in 160 training iterations.
- Logistic Regression: Implemented using a generalized linear model with a logit function and binomial distribution, with coefficient estimation by maximum likelihood and standard algorithm parameters, without extensive hyperparameter tuning.

IV. RESULTS

A. Exploratory Analysis

The dataset comprises 4,477,688 observations, collected between January 2022 and December 2025. It has 27 explanatory variables related to the geometry of the railway track and operating conditions which, for confidentiality reasons, will be coded as $x_1, \cdots, x_{27}$, and a response variable y, indicating the presence or absence of a fracture at the analyzed track kilometer. Furthermore, the dataset exhibits two critical characteristics: the presence of missing values and extreme class imbalance, with only 429 failure events compared to approximately 4.5 million records without failures. Due to

confidentiality restrictions imposed by the company, the dataset cannot be publicly released and may only be accessed upon formal request and the execution of a non-disclosure agreement.

As an initial preprocessing step, all observations with missing values were removed, resulting in 2,809,150 valid records, including 335 fracture cases. This strategy was adopted to ensure greater consistency and reliability in the predictive modeling process. Despite the reduction in dataset size, the data remained sufficiently representative and preserved the characteristics of extreme imbalance. In large-scale operational datasets, imputation procedures can introduce noise, statistical distortions, and physically inconsistent combinations between variables, especially in scenarios with multiple sensors and railway measurements.

An exploratory analysis was performed using the t-test to compare the means of the explanatory variables between failure and non-failure groups. Annex 1 presents the means of the two groups and the p-values of the hypothesis test. The results indicated statistically significant differences (p-value < 0.05) for variables x_3, x_4, x_5, x_6, x_{10}, x_{12}, x_{14}, x_{15}, x_{16} ,x_{17}, x_{21}, x_{22} and x_{23}. However, since the t-test evaluates variables individually, it does not capture complex interactions or non-linear relationships; additionally, in large and unbalanced datasets, small differences may have statistical significance without operational impact. Thus, the results of the t-test should be interpreted with caution and complemented by multivariate methods and predictive models. Consequently, these findings were treated only as initial exploratory analysis and were subsequently complemented by machine learning models capable of identifying more complex predictive patterns.

Although tree-based models such as Random Forest and XGBoost are generally robust to multicollinearity, highly correlated variables can still affect result interpretation and introduce instability in variable importance. Therefore, correlation analysis was performed to mitigate multicollinearity, particularly due to the inclusion of Logistic Regression, which is more sensitive to this issue. Additionally, removing redundant variables helps reduce model complexity and improve computational efficiency without significantly affecting predictive performance. Pearson correlation coefficients were calculated for all variable pairs, and the highly correlated variables identified are presented in Annex 2.

To avoid multicollinearity, highly correlated variables were grouped and represented by variables considered most appropriate from an operational standpoint. The selection of these representative variables was based not only on statistical criteria but also on the technical knowledge of the railway company's specialists, prioritizing attributes that were more easily monitored, interpreted, and controlled in the operational and maintenance context of the railway track. After this selection stage, the variables retained for modeling were: x_1, x_4, x_7 , x_9, x_{10}, x_{11}, x_{14},x_{17},x_{18}, x_{20}, x_{21} and x_{23}.

B. Data Balancing

Initially, the dataset was randomly partitioned in an 80/20 ratio, allocating 80% of the observations (2,247,320 records) to the training set and 20% (561,830 records) to the test set. Random division was chosen instead of a strictly temporal separation (due to operational and structural variations of the railway over time), which could concentrate specific infrastructure conditions in only one of the subsets, compromising the statistical representativeness of the data. Thus, as the objective of the study was to evaluate the discriminative capacity of machine learning models in identifying railway fractures, random partitioning was considered more appropriate to preserve the representativeness of the classes and operational conditions present in the database.

No balancing technique was applied to the test set, preserving the original distribution of classes to ensure a more realistic evaluation of the models' performance.

The balancing strategy adopted in the training set was the Cluster-based Hybrid Sampling method, which combines Cluster-Based Undersampling (CBU) in the majority class and SMOTE in the minority class. This technique was chosen due to its simplicity, flexibility, and greater practical feasibility in the context of this study. Initially, the majority class was grouped into 500 clusters, from which two representative observations were selected per group, totalling 1,000 instances. Then, SMOTE was applied to the 236 original observations with failure, generating synthetic samples until 500 instances were reached.

At the end of the process, a substantially rebalanced training set was obtained, composed of 1,500 observations, 1,000 from the majority class and 500 from the minority class, while maintaining an informative and representative sample of the original data.

C. Modeling

Different machine learning models were evaluated to identify the most suitable approach for predicting rail fractures.

1) Random Forest

After the Grid Search optimization process, the best Random Forest model configuration was obtained with mtry = 6, ntree = 100, and nodesize = 1, showing the highest average performance in the 5-fold cross-validation. The model was trained using the explanatory variables x_1, x_4, x_7 , x_9, x_{10}, x_{11}, x_{14},x_{17},x_{18}, x_{20}, x_{21} and x_{23}, previously selected after correlation analysis and multicollinearity reduction. Fig. 2 presents the main results obtained by the best Random Forest model.

The results demonstrated high model performance in railway fracture prediction, with sensitivity close to 92%, specificity of 99.99%, F-measure of 0.831, and G-mean of 0.959, indicating strong discriminative capability. The confusion matrix analysis showed that only 8 out of 99 fracture cases were not identified, while the number of false positives remained low, a behavior considered acceptable in the railway context due to the priority given to operational safety. The PR-AUC curve (0.998) confirmed the model's high capacity to identify the minority class without generating an excessive number of false alarms. Regarding feature importance, variables x_{17}, x_4, x_{14} and x_{23} stood out as the most relevant. The importance distribution

suggests that the model does not rely on a single predictor, but rather on a set of factors related to track geometry and operational conditions, highlighting the multifactorial nature of the analyzed problem.

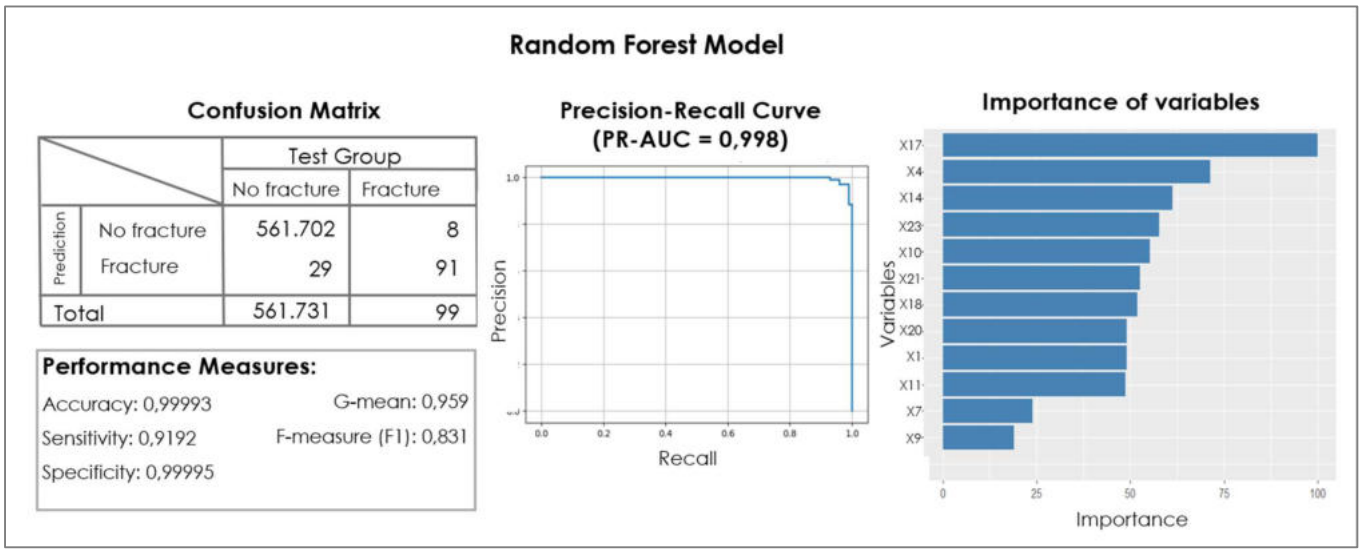


Fig. 2. Confusion matrix, performance measure, P-R Curve e PR-AUC of the best Random Forest model.

2) XGBoost

For the XGBoost model, the selected configuration corresponded to nrounds = 200, max_depth = 5, eta = 0.05, gamma = 1, min_child_weight = 1, subsample = 0.9, and colsample_bytree = 0.8. This combination presented the best balance between predictive capacity and generalization, considering the average F1-score obtained in the validation subsets. The model was trained using the same explanatory variables as the Random Forest model, previously selected after correlation analysis and multicollinearity reduction. Fig. 3 presents the main results obtained by the best XGBoost model.

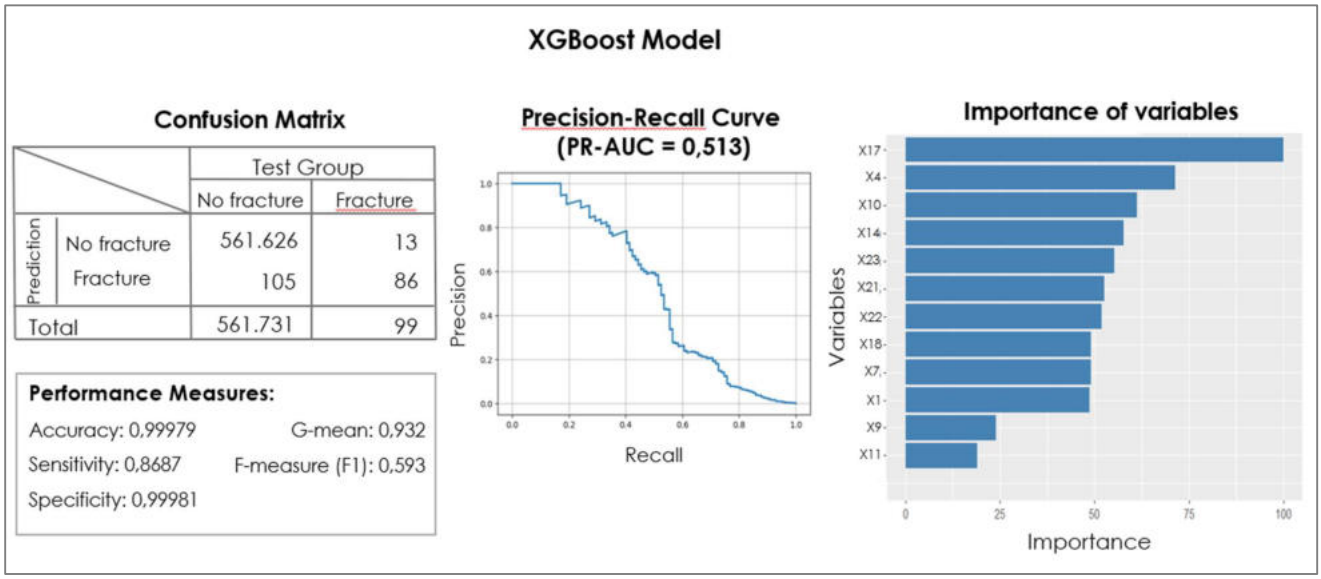


Fig. 3. Confusion matrix, performance measure, P-R Curve e PR-AUC of the best XGBoost model.

The XGBoost model achieved high sensitivity (0.8687), specificity of 0,99981, F-measure of 0.593, and G-mean of 0.932, indicating satisfactory performance, although inferior to that of the Random Forest model. The confusion matrix showed that 86 out of 99 fracture cases were correctly identified (13 out of 99 fracture cases were not identified), although the model produced a higher number of false positives and false negatives. The PR-AUC curve (≈ 0.51) confirmed moderate discriminative capability in identifying the minority class, highlighting greater difficulty in balancing precision and sensitivity. Regarding feature importance, variables x_{17}, x_4, x_{10} and x_{14}stood out as the most relevant. The importance distribution demonstrates that the model relies on multiple attributes, reinforcing the multifactorial nature of railway failures. Overall, although the model was able to capture relevant patterns in the data, its performance was inferior to Random Forest, particularly regarding the detection of critical failures.

3) Logistic Regression

A logistic regression model was fitted, using the same dataset and explanatory variables as in the previous models. For the Logistic Regression, the classic implementation of the generalized linear model available in R was used, with a logit link function and binomial distribution. The estimation of the coefficients was performed via the maximum likelihood method with the algorithm's default parameters. Fig. 4 presents the results of the Logistic Regression model.

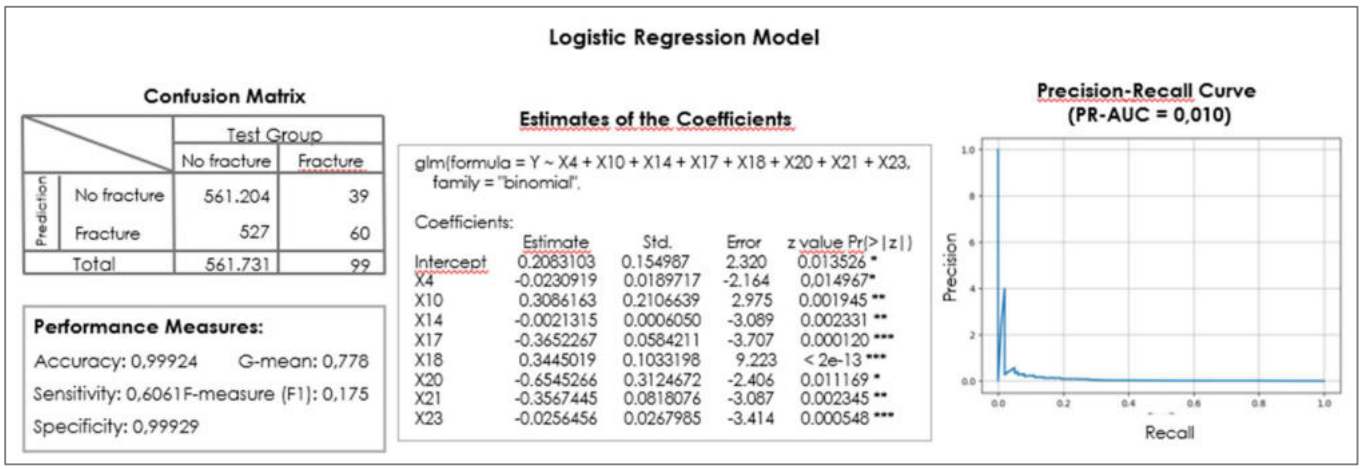


Fig. 4. Confusion matrix, performance measure, P-R Curve e PR-AUC of the Logistic regression model.

Logistic Regression showed inferior performance compared to the previous models, with a sensitivity of 0.6061, an F-measure of 0.175, and a G-mean of 0.778, indicating limited capability for detecting railway fractures. The confusion matrix revealed a high number of false negatives and false positives, compromising its practical application in critical operational safety contexts. The PR-AUC curve (0.01) confirmed the model's low discriminative capacity for the minority class, highlighting difficulties in handling extreme imbalance and capturing complex patterns and non-linear relationships present in railway data. Regarding feature importance, variables x_{10} and x_{18} showed a positive association with fracture occurrence, whereas x_4, x_{14}, x_{17}, $x_{20,}$, x_{21} and x_{23} presented an inverse relationship. Despite its high interpretability, the results indicate that Logistic Regression is more suitable as a complementary tool for factor analysis than as a primary model for railway predictive maintenance.

D. Definition of the rail fracture prediction model

The comparative analysis of the fitted models reveals that the Random Forest model exhibits the best overall performance, standing out for its high fracture detection capacity (high sensitivity) combined with a low error rate, both in terms of false negatives and false positives, compared to the other evaluated models.

Based on these results, the Random Forest model was selected as the most suitable solution for predicting rail fractures. Its application allows not only for the anticipation of failures but also supports the definition of more efficient maintenance strategies.

E. Variables most important to predict rail fracture occurrence

In general, high consistency is observed between the Random Forest, XGBoost, and Logistic Regression models regarding the variables considered most relevant for predicting railway fractures. Notably, the variables x_{17}, x_4, x_{14}, x_{23}, x_{10}, x_{21} and x_{18} stood out simultaneously in the tree-based models and also showed statistical significance in Logistic Regression (p-value < 0.05). This result suggests that these variables have a consistent and robust contribution to the identification of failure occurrences, regardless of the modeling approach used.

Among the models analyzed, the Random Forest showed the best overall predictive performance, followed by XGBoost. Between these two tree-based models, a very similar behavior is observed regarding the hierarchy of importance of the variables, with special emphasis on x_{17}, considered the most influential variable in both models. This convergence reinforces the evidence that certain operational and geometric characteristics of the railway track have a strong relationship with the occurrence of fractures.

Although Logistic Regression also identified statistically relevant variables, its inferior performance suggests that some of the relationships between the predictors and the occurrence of failures may involve complex interactions and non-linear patterns, more adequately captured by tree-based models.

V. CONCLUSIONS

The present study aimed to develop a predictive model capable of supporting decision-making in railway predictive maintenance activities, with a focus on identifying and prioritizing railway track segments with a higher probability of fracture occurrence. To achieve this objective, two research questions were proposed: (i) Can machine learning models provide adequate predictive performance in highly imbalanced railway datasets? and (ii) Which variables are most relevant for predicting rail fractures?

To address these questions, an exploratory data analysis stage was initially conducted, involving the removal of observations with missing values and the analysis of correlations among variables in order to mitigate multicollinearity issues. This stage proved important for improving the robustness, stability, and interpretability of the developed models.

Due to the severe class imbalance, a hybrid balancing strategy for the training set, named Cluster-based Hybrid Sampling, was adopted. This approach was selected because of its simplicity, lower computational cost, flexibility, and practical feasibility within the investigated operational context. As a result, a substantially rebalanced training set was obtained, consisting of 1,500 observations, including 1,000 instances from the majority class and 500 from the minority class, while preserving the structural representativeness of the original data.

After the data balancing stage, railway fracture prediction models were trained. Considering the technical familiarity of the research team, the interpretability of the algorithms, and the feasibility of implementation within the railway company's operational environment, the Random Forest, XGBoost, and Logistic Regression models were selected. Model training was conducted through hyperparameter optimization using Grid Search combined with 5-fold cross-validation.

Given the strong class imbalance in the dataset, model selection was based on the F1-score, as this metric provides a better balance between precision and sensitivity for identifying the minority class. After selecting the best configurations for each algorithm, the final evaluation was performed on the test set using sensitivity, specificity, and G-mean metrics, enabling a broader assessment of the classifiers' discriminative capacity.

The results showed that the Random Forest model achieved the best overall performance, standing out for its high fracture detection capability and balanced sensitivity and specificity. The XGBoost model presented intermediate performance, whereas Logistic Regression, despite its greater interpretability, demonstrated relevant limitations regarding predictive capability.

From a practical perspective, the results indicate that the proposed model can contribute to a more proactive and data-driven approach to railway maintenance management. Thus, the first research question was positively answered, demonstrating that machine learning models can provide adequate predictive performance for railway fracture identification in highly imbalanced scenarios. The application of these models enables failure anticipation, prioritization of inspections in critical regions, and optimization of resource allocation, thereby reducing the need for emergency corrective interventions.

To answer the second research question, variable importance analysis was performed for the tree-based models, while statistical significance analysis was conducted for the Logistic Regression model. A high degree of consistency was observed among the models regarding the variables considered most relevant for railway fracture prediction. In particular, variables $x_{17}, x_4, x_{14}, x_{23}, x_{10}, x_{21}$ and x_{18} simultaneously stood out in the Random Forest and XGBoost models and also showed statistical significance in Logistic Regression (p-value < 0.05). These findings suggest that such variables provide consistent and robust contributions to failure identification, offering relevant insights for continuous railway track monitoring and for the definition of more effective operational indicators.

In summary, the results demonstrate that the use of machine learning models can significantly contribute to increasing operational safety, improving railway system reliability, optimizing inspection activities, and reducing costs associated with unexpected failures, thereby establishing itself as a promising approach for real-world railway predictive maintenance applications.

Limitations

Despite the promising results obtained, the present study has some limitations that should be considered when interpreting and applying the findings in practice.

First, data quality and representativeness constitute critical factors. The exclusion of observations with missing values, although necessary to ensure the consistency of the analyses, may have reduced the variability of the dataset and eliminated potentially relevant patterns. Furthermore, the model was developed based on historical data from a specific railway context, which may limit its generalization capability to other railway networks, operational conditions, or distinct structural characteristics.

Another relevant limitation concerns the treatment of class imbalance. Although the adopted hybrid balancing strategy demonstrated satisfactory results, the substantial reduction of the majority class may have led to the loss of potentially relevant information for the modeling process. Likewise, the generation of synthetic samples through oversampling techniques may introduce artificial patterns that do not always accurately represent the actual operational conditions of the railway system.

Additionally, the study focused on a specific balancing approach and did not include a comparative evaluation of other, more recent or more sophisticated techniques, which could potentially provide superior predictive performance. Therefore,

future research may explore different balancing strategies and compare their impacts on the discriminative capacity of predictive models in highly imbalanced railway scenarios.

From a methodological perspective, the models employed, although robust, present inherent limitations. Ensemble-based models, such as Random Forest, offer high predictive performance but lower interpretability when compared to linear models. On the other hand, Logistic Regression, although more interpretable, showed inferior performance in detecting rare events, highlighting the challenge of balancing predictive performance and explainability in real-world problems.

Another limitation of the study concerns the set of evaluated algorithms. The analysis focused exclusively on three supervised classification models (Random Forest, XGBoost, and Logistic Regression) selected due to their widespread use in the literature, interpretability, and feasibility of implementation within the railway company's operational context. However, other machine learning models and more advanced techniques could potentially achieve superior performance in railway fracture prediction. Among the approaches not explored in this study are more sophisticated ensemble algorithms, such as LightGBM and CatBoost, Support Vector Machine (SVM)-based methods, artificial neural networks, and deep learning architectures, as well as hybrid models and techniques specifically designed for learning from imbalanced data. Furthermore, temporal approaches and sequential models, such as LSTM networks and time-series-based methods, could also capture dynamic patterns of railway degradation over time. Thus, future studies may expand the comparison among different algorithms and modeling strategies, evaluating the potential of more complex methods to improve predictive performance in highly imbalanced railway scenarios.

Additionally, the study did not explicitly consider the temporal dimension of the data. The occurrence of rail fractures is a dynamic phenomenon influenced by factors accumulated over time. The absence of temporal modeling may limit predictive capability in scenarios where the evolution of track conditions plays a determining role.

REFERENCES

Altalhan, M., Algarni, A., & Alouane, MTH (2025). The problem of imbalanced data in machine learning: a review.. IEEE access , 13 , 13686-13699.

Altwaijry, N. (2023). Probability-based synthetic minority oversampling technique. IEEE Access, 11, 28831–28839. https://doi.org/10.1109/access.2023.3260723

Batista, G. E., Prati, R. C. and Monard, M. C. (2004). "A study of the behavior of several methods for balancing machine learning training data". ACM SIGKDD explorations newsletter, Vol. 6 No. 1, pp. 20-29. https://doi.org/10.1145/1007730.1007735

Branco, P., Torgo, L. and Ribeiro, R. P. (2016). "A survey of predictive modeling on imbalanced domains". ACM Computing Surveys, Vol. 49 No. 2, pp. 1–50. https://doi.org/10.1145/2907070

Buda, M., Maki, A. and Mazurowski, M. A. (2018). "A systematic study of the class imbalance problem in convolutional neural networks". Neural Networks, Vol. 106, pp. 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Chaabane, I., Guermazi, R. and Hammami, M. (2020). "Enhancing techniques for learning decision trees from imbalanced data". Advances in Data Analysis and Classification, Vol. 14 No. 3, pp. 677–745. https://doi.org/10.1007/s11634-019-00354-x

Chenariyan Nakhaee, M., Hiemstra, D., Stoelinga, M. and van Noort, M. (2019, April). "The recent applications of machine learning in rail track maintenance: A survey". In International Conference on Reliability, Safety, and Security of Railway Systems. Cham: Springer International Publishing, pp. 91-105. https://doi.org/10.1007/978-3-030-18744-6_6

Coughlan, P., & Coghlan, D. (2002). Action research for operations management. International Journal of Operations and Production Management. https://doi.org/10.1108/01443570210417515

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. and Herrera, F. (2012). "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches". IEEE Transactions on Systems, May, and Cybernetics – Part C: Applications and Reviews, Vol. 42 No. 4, pp. 463–484

Guillén-Teruel, A., Caracena, M., Pardo, J. A., de-la-Gándara, F., Palma, J., & Marín-Morales, J. (2025). FILM: Framework for Imbalanced Learning Machines based on a new unbiased performance measure and a new ensemble-based technique. arXiv preprint arXiv:2503.04370. https://arxiv.org/abs/2503.04370 https://doi.org/10.48550/arXiv.2503.04370

He, H. and Garcia, E. A. (2009). "Learning from imbalanced data". IEEE Transactions on knowledge and data engineering, Vol. 21 No. 9, pp.1263-1284. https://doi.org/10.1109/TKDE.2008.239

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications, 73, 220–239. https://doi.org/10.1016/j.eswa.2016.12.035

Imani, M., Joudaki, M., Bagheri, A., & Arabnia, H. R. (2026). Why ROC-AUC Is Misleading for Highly Imbalanced Data: In-Depth Evaluation of MCC, F2-Score, H-Measure, and AUC-Based Metrics Across Diverse Classifiers. Technologies, 14(1), 54. https://doi.org/10.20944/preprints202510.0958.v2

Iqbal, S., Qureshi, A.N., Li, J., Choudhry, I.A., & Mahmood, T. (2023). Dynamic learning for imbalanced data in learning chest x-ray and CT images. Heliyon, 9, e16807. https://doi.org/10.1016/j.heliyon.2023.e16807

Jani, J., Doshi, J., Kheria, I., Mehta, K., Bhadane, C., & Karani, R. (2023). LayNet - A multi-layer architecture to handle imbalance in medical imaging data. Computers inBiology and Medicine, 163, 107179. https://doi.org/10.1016/j.compbiomed.2023.107179

Kaikashan Siddavatam, & Subhash Shinde. (2025). "Addressing intrinsic data characteristics issues of imbalance medical data using nature inspired percolation clustering". Journal of Electronics, Electromedical Engineering and Medical Informatics, 7(3), 740–762. https://doi.org/10.35882/jeeemi.v7i3.835

Ke, S. W., Tsai, C. F., Pan, Y. Y. and Lin, W. C. (2024). "Majority re-sampling via sub-class clustering for imbalanced datasets". Journal of Experimental & Theoretical Artificial Intelligence, Vol. 36 No. 8, pp. 1581-1596. https://doi.org/10.1080/0952813x.2023.2165715

Kreuzer, M., Schmidt, A., & Kellermann, W. (2021, August). Novel features for the detection of bearing faults in railway vehicles. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Vol. 263, No. 3, pp. 3833-3844). Institute of Noise Control Engineering.https://doi.org/10.48550/arXiv.2304.08249 https://doi.org/10.3397/in-2021-2537

Liu, R. (2023). A novel synthetic minority oversampling technique based on relative and absolute densities for imbalanced classification: RL. Liu. Applied Intelligence, 53(1), 786-803. https://doi.org/10.1007/s10489-022-03512-5

Maxson, T., Corizzo, R., Wu, Y., Japkowicz, N., & Bellinger, C. (2026). Correcting performance estimation bias in imbalanced classification with minority subconcepts. arXiv preprint arXiv:2604.26024. https://arxiv.org/abs/2604.26024

McDermott, M. B., Zhang, H., Hansen, L. H., Angelotti, G., & Gallifant, J. (2024). A closer look at AUROC and AUPRC under class imbalance. Advances in Neural Information Processing Systems, 37, 44102-44163. https://doi.org/10.52202/079017-1400

Mihelich, M., Castagnos, F., & Dognin, C. (2024). Interplay of ROC and Precision-Recall AUCs: Theoretical Limits and Practical Implications in Binary Classification. In Forty-first International Conference on Machine Learning.

Newaz, A., & Haq, F. S. (2022). A novel hybrid sampling framework for imbalanced learning. arXiv preprint arXiv:2208.09619. https://doi.org/10.2139/ssrn.4200131

Ngamkhanong, C., Kaewunruen, S. and Costa, B. J. A. (2018). "State-of-the-art review of railway track resilience monitoring". Infrastructures, Vol. 3 No. 1, p. 3. https://doi.org/10.3390/infrastructures3010003

Nikpour, B., Rahmati, F., Mirzaei, B. and Nezamabadi-pour, H. (2026). "A comprehensive review on data-level methods for imbalanced data classification". Expert Systems with Applications, Vol. 295, pp. 128920. https://doi.org/10.1016/j.eswa.2025.128920

Ohsaki, M., Wang, P., Matsuda, K., Katagiri, S., Watanabe, H. and Ralescu, A. (2017). "Confusion-matrix-based kernel logistic regression for imbalanced data classification". IEEE Transactions on Knowledge and Data Engineering, Vol. 29 No. 9, pp. 1806–1819. https://doi.org/10.1109/tkde.2017.2682249

Richardson, E., Trevizani, R., Greenbaum, J. A., Carter, H., Nielsen, M., & Peters, B. (2024). The receiver operating characteristic curve accurately assesses imbalanced datasets. Patterns, 5(6), 100994. https://doi.org/10.1016/j.patter.2024.100994

Sahin, Y., Bulkan, S. and Duman, E. (2013). "A cost-sensitive decision tree approach for fraud detection". Expert Systems with Applications, Vol 40 No. 15, pp. 5916–5923. https://doi.org/10.1016/j.eswa.2013.05.021

Salmi, M., Atif, D., Oliva, D., Abraham, A., & Ventura, S. (2024). Handling imbalanced medical datasets: review of a decade of research. Artificial intelligence review, 57(10). https://doi.org/10.1007/s10462-024-10884-2

Siddavatam, K. I., & Shinde, S. K. (2025). A hybrid literature review on handling imbalanced medical data: AI models and open issues. Expert Systems with Applications, 129004. https://doi.org/10.1016/j.eswa.2025.129004

Sun, Y., Wong, A. K., & Kamel, M. S. (2009). "Classification of imbalanced data: A review". International journal of pattern recognition and artificial intelligence, Vol. 23 No. 4, pp. 687-719.

Tang, Y., Zhang, Y. Q. and Chawla, N. V. (2009). "SVMs modeling for highly imbalanced classification". IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 39 No. 1, pp. 281–288.

Thiollent, M. (2011). Action research and participatory research: An overview. International Journal of Action Research. https://doi.org/10.1688/1861-9916_IJAR_2011_02_Thiollent

Vairetti, C., Assadi, J. L., & Maldonado, S. (2024). Efficient hybrid oversampling and intelligent undersampling for imbalanced big data classification. Expert Systems with Applications, 246, 123149. https://doi.org/10.1016/j.eswa.2024.123149

Wang, Y.C., & Cheng, C.H. (2021). A multiple combined method for rebalancing medical data with class imbalances. Computers in Biology and Medicine, 134, 104527. https://doi.org/10.1016/j.compbiomed.2021.104527

Wongvorachan, T., He, S., & Bulut, O. (2023). A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. Information, 14(1), 54. https://doi.org/10.3390/info14010054

Yu, H., Ni, J. and Zhao, J. (2013). "ACOSampling: An ant colony optimization-based undersampling method for classifying imbalanced DNA microarray data". Neurocomputing, Vol 101, pp. 309–318. https://doi.org/10.1016/j.neucom.2012.08.018

Zhang, C., Tan, K. C., Li, H. and Hong, G. S. (2019). "A cost-sensitive deep belief network for imbalanced classification". IEEE Transactions on Neural Networks and Learning Systems, Vol. 30 No. 1, pp. 109–122. https://doi.org/10.1109/tnnls.2018.2832648

Zhang, Z., Yu, S., Yang, S., Zhou, Y., & Zhao, B. (2021). Rail-5k: A real-world dataset for rail surface defects detection. arXiv preprint arXiv:2106.14366. https://doi.org/10.48550/arXiv.2106.14366

Zheng, W. and Zhao, H. (2020). "Cost-sensitive hierarchical classification for imbalance classes". Applied Intelligence, Vol 50 No. 8, pp. 2328–2338. https://doi.org/10.1007/s10489-019-01624-z

Zhou, Q. and Sun, B. (2024). "Adaptive K-means clustering-based under-sampling methods to solve the class imbalance problem". Data and Information Management, Vol. 8 No. 3, p. 100064. https://doi.org/10.1016/j.dim.2023.100064

Zhu, L., Chen, C., Wang, H., Yu, F. R. and Tang, T. (2023). "Machine learning in urban rail transit systems: A survey". IEEE Transactions on Intelligent Transportation Systems, Vol. 25 No. 3, pp. 2182-2207. DOI: 10.1109/TITS.2023.3319135

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Beyond Outcomes: Reframing Value-Based Evaluation of Multimorbidity Healthcare Services

Mariana Peyroteo [1], André Mendes de Carvalho [2] and Luís Velez Lapão [3]

[1] UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal; Laboratório Associado de Sistemas Inteligentes, LASI, 4800-058 Guimarães, Portugal; mba.santos@campus.fct.unl.pt

[2] UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal; Laboratório Associado de Sistemas Inteligentes, LASI, 4800-058 Guimarães, Portugal; ame.carvalho@fct.unl.pt

[3] UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal; Laboratório Associado de Sistemas Inteligentes, LASI, 4800-058 Guimarães, Portugal; l.lapao@fct.unl.pt

Abstract*— Multimorbidity healthcare services are difficult to evaluate because service value depends not only on outcomes, but also on care processes, service-delivery conditions, and organizational factors that are not consistently represented in conventional performance assessment models. This challenge becomes more pronounced in digitally supported services, where access, communication, monitoring, follow-up, and information exchange are increasingly shaped by digital and operational conditions. This study developed the Conceptual Structure for Value-Based Evaluation of Multimorbidity Healthcare Services (VEMS). A conceptual and analytical approach was adopted, combining literature-informed specification of domains and subdomains, Service Experience Blueprint observation, stakeholder-informed analysis, structured document review, cross-source synthesis, and external panel validation. The resulting structure comprised three domains, Quality of Care, Service Delivery, and Costs, organized into eight subdomains and 51 validated components. The final structure made visible care processes, service-delivery features, organizational conditions, and cost-related burdens that remain only partially represented in conventional evaluation models, particularly those related to coordination, continuity, follow-up, communication, digital support conditions, and indirect burden. The VEMS Conceptual Structure provides a broader evaluative architecture for examining how value is generated in complex multimorbidity healthcare services and establishes a conceptual basis for subsequent framework development, performance assessment, and service improvement.

Keywords – performance measurement, service systems engineering, evaluation architecture, multimorbidity

I. Introduction

Multimorbidity, commonly defined as the coexistence of two or more chronic conditions, remains a major challenge for health systems and healthcare delivery (Xu, Mishra and Jones, 2017). Care for these patients is typically delivered over extended periods and across multiple professionals, services, and decision points, requiring coordination, adaptation, and sustained follow-up (Schiøtz, Høst and Frølich, 2016; Leijten *et al.*, 2018). Healthcare delivery, however, remains largely organized around episodic contacts and disease-specific pathways, which may contribute to fragmentation and weaken continuity of care for patients with long-term, interdependent needs (Leijten *et al.*, 2018). In this context, multimorbidity healthcare services are difficult both to organize and to evaluate, because performance emerges from the interaction of multiple clinical, organizational, and service-delivery conditions over time.

Current evaluation approaches remain essential for performance assessment, accountability, and service improvement (de Melo Santos, Barbosa and Sant'Anna, 2025). Their operational logic, however, often relies on predefined indicators and performance measures that privilege what can be more readily measured, such as outputs, outcomes, utilization patterns, and costs. From a performance measurement perspective, this creates a risk of over-representing what is easily measurable while under-representing the wider system conditions that shape service performance (Choong, 2013). Although current approaches support comparison across services and monitoring of selected performance dimensions, they may insufficiently represent service-delivery features and operational conditions that influence performance in complex multimorbidity settings (De Rosis, Ferrè and Pennucci, 2022). Continuity of care, coordination across providers, follow-up capacity, responsiveness to evolving needs, and the conditions under which care is delivered may all influence service performance, yet these elements are not consistently incorporated into conventional performance assessment models (Ruscitto *et al.*, 2016).

This challenge can be understood as a problem of designing an appropriate evaluation architecture for a complex healthcare service system. Service systems engineering frames services as configurations of actors, information, technologies, and organizational resources organized around value creation (Böhmann, Leimeister and Möslein, 2014). In multimorbidity healthcare services, evaluation must therefore capture not only the results achieved, but also the organizational, informational, and operational conditions through which those results are produced. This requires a

Authors acknowledge Fundação para a Ciência e a Tecnologia I.P. for its financial support via the project UID/00667: Unidade de Investigação e Desenvolvimento em Engenharia Mecânica e Industrial. MP was supported by FCT - Fundação para a Ciência e Tecnologia, I.P., by 2023.00994.BD (https://doi.org/10.54499/2023.00994.BD). In both cases, the funding body provided financial support for the research, with no involvement in the analysis or writing of the article.

broader evaluative representation capable of integrating multiple dimensions of service value in a coherent and structured way.

Within Value-Based Healthcare (VBHC), value provides a broad conceptual basis for healthcare evaluation (Porter, 2010; Teisberg, Wallace and O'Hara, 2020). In this study, that logic was translated into three domains: Quality of Care, Service Delivery, and Costs. Quality of Care refers to the conditions under which care is delivered, the care processes performed, and the outcomes achieved (Donabedian, 1988). Service Delivery refers to how healthcare is delivered across settings, encounters, and modalities, extending beyond patient experience alone to include access, responsiveness, continuity, safety, and operational performance (Leijten *et al.*, 2018). Costs refer to the resources associated with providing and receiving care, including direct healthcare costs and utilization-related resource use associated with service delivery (Scheurer *et al.*, 2016). The need for this broader representation becomes more pronounced in digitally supported healthcare services, where access, communication, monitoring, information exchange, and follow-up are increasingly shaped by digital and organizational conditions, while usability, technical support, interoperability, and administrative workload may also influence service performance and value generation (Vest and Kash, 2016; Atasoy, Greenwood and McCullough, 2019).

This paper develops the Conceptual Structure for Value-Based Evaluation of Multimorbidity Healthcare Services (VEMS). The aim is to specify a structured evaluative architecture for multimorbidity healthcare services through the identification and organization of candidate evaluation components across Quality of Care, Service Delivery, and Costs, drawing on literature, Service Experience Blueprint analysis, stakeholder input, and current evaluation instruments. In doing so, the study clarifies which care processes, service-delivery features, organizational conditions, and value-relevant dimensions require more explicit representation in multimorbidity healthcare service evaluation. The paper therefore provides a conceptual basis for subsequent framework development, performance assessment, and service improvement in complex healthcare service systems.

II. Methods

A. Study Design

This study was designed as a conceptual and analytical study concerned with the specification of a structured evaluation architecture for multimorbidity healthcare services, consistent with performance measurement and service systems perspectives (Choong, 2013). The analysis proceeded in six steps: a targeted review of relevant literature to specify subdomains, Service Experience Blueprint observation of care processes, stakeholder-informed analysis of candidate components, structured document review of current evaluation instruments and reference frameworks, integrative synthesis of cross-source candidate components, and external panel validation of the conceptual structure.

B. Targeted Review of Relevant Literature

A targeted review of relevant literature was conducted to identify and define candidate subdomains within the three predefined domains adopted in the study: Quality of Care, Service Delivery, and Costs (Snyder, 2019). The purpose of this step was to examine how each domain is structured in the literature and to support the specification of subdomains for evaluating multimorbidity healthcare services.

The review was conducted in PubMed and Web of Science using domain-specific search strategies combining terms related to multimorbidity with terms associated with each evaluative domain. Searches focused on publications from the last ten years, while foundational references were retained when needed to support conceptual definitions. Records were screened by title and abstract, followed by full-text assessment of papers considered potentially relevant to the identification or definition of subdomains. Papers were included when they contributed to the conceptualization, evaluation, measurement, or operational interpretation of Quality of Care, Service Delivery, or Costs in multimorbidity or closely related healthcare contexts. Papers were excluded when they focused exclusively on disease-specific clinical outcomes, intervention effects, or technical issues unrelated to the evaluative domain structure. Recurring evaluative components, subdomain labels, and domain-relevant definitions were extracted and compared within each domain, and candidate subdomains were consolidated to reduce overlap and align terminology.

C. Service Experience Blueprint

Care process observations were used as an empirical input to examine how multimorbidity follow-up care is organized in routine primary care practice. A Service Experience Blueprint (SEB) approach was adopted to document the sequence of care activities, the distribution of work across patient-facing and backstage tasks and the operational conditions affecting service delivery (Bitner, Ostrom and Morgan, 2008). Non-participant observations were conducted during 61 follow-up consultations focused on chronic disease management, including 36 medical consultations and 25 nursing consultations. Observations were guided by a structured framework organized into four hierarchical levels: stage, activity, subtask, and operational description.

The framework distinguished direct patient-facing activities from backstage professional work, including documentation, coordination, administrative tasks, and interaction with digital systems. It also captured interruptions and operational constraints affecting care delivery, including system-related issues and workflow disruptions. Observational data were used to derive candidate components related to coordination, continuity, documentation burden, system use, and organizational friction in multimorbidity care.

D. Stakeholder-Informed Analysis

Stakeholder interviews were used as an analytical input to identify candidate components that reflect how value-relevant dimensions are recognized by different groups involved in multimorbidity care. Semi-structured interviews were

conducted with 22 healthcare professionals, including 10 physicians, 2 interns, and 10 nurses, together with 20 patients with multimorbidity enrolled in chronic disease management programs. Participants were selected using predefined eligibility criteria, and variation across professional roles and patient profiles was ensured by including diverse stakeholder groups.

A shared interview guide was used to examine care practices for multimorbidity, challenges in care delivery, the role and limitations of digital tools, and value-relevant dimensions of Quality of Care, Service Delivery, and Costs. Although the same analytical areas were addressed across groups, question wording was adapted to reflect differences in professional roles and patient experience. Interviews were audio-recorded and transcribed verbatim.

Transcripts were analyzed thematically using a combined deductive and inductive approach (Braun and Clarke, 2006). Deductive coding was organized around the predefined domains and subdomains, while inductive coding was used to identify recurring candidate components emerging from stakeholder accounts. Coding was conducted in NVivo 14 (version 14.24.3), with iterative refinement to improve coherence and reduce overlap.

E. Structured Document Review

A structured document review was conducted to examine evaluation instruments and reference frameworks relevant to multimorbidity care, primary care service delivery, integrated care, patient-reported outcomes and experiences, organizational quality, safety, and cost-related evaluation (Dalglish, Khalid and McMahon, 2021). The objective was not to extract indicators directly for use as final study measures, but to identify evaluative content embedded in existing instruments and frameworks and transform it into component-level expressions.

Documents were selected purposively based on their relevance to the study focus and their alignment with the predefined domains of Quality of Care, Service Delivery, and Costs. Primary sources included the ICHOM standard sets for older persons and diabetes (Akpan *et al.*, 2018; Nano *et al.*, 2020), OECD PaRIS and related patient-reported indicator survey materials (OECD, 2019, 2023), the Portuguese Primary Health Care contractualization framework for 2025, and the DiOr-CSP organizational quality guide (Administração Central do Sistema de Saúde, 2019, 2025). Complementary integrated care references were also reviewed to support the interpretation of coordination, continuity, integration, multidisciplinary care, patient-centeredness, digital enablement, and service-delivery transformation (OECD, 2023).

The documentary material was analyzed using directed content analysis (Hsieh and Shannon, 2005). The coding structure was defined a priori using the three domains and eight subdomains. Each extracted element was assigned to a domain and subdomain based on its primary evaluative meaning, and ambiguities or overlaps were documented in extraction notes. Relevant evaluative content was recorded in a structured extraction matrix, including the source document, the original evaluative element, the content type, the allocated domain and subdomain, the derived component, and notes on derivation or overlap. Original wording was retained to preserve traceability, while derived components reformulated metric-heavy or instrument-specific content into component-level evaluative expressions. Document-derived components were then reviewed to identify overlap, harmonize terminology, and improve conceptual consistency.

F. Integrative Synthesis and Validation

Candidate components derived from Service Experience Blueprint, stakeholder interviews, and the structured document review were integrated through an affinity-based synthesis process in which source-derived components were compared, grouped, and harmonized according to conceptual proximity, overlap, complementarity, and distinct evaluative meaning (Scupin, 1997). The process was conducted within the predefined analytical structure of Quality of Care, Service Delivery, and Costs, together with their respective subdomains.

To support this integration, candidate components from the three source streams were compiled into an integrated cross-source matrix organized by domain and subdomain. For each emerging component, the matrix documented the corresponding observation-derived, interview-derived, and document-derived components. This enabled systematic assessment of convergence, source-specific emphasis, partial coverage, and conceptual redundancy across the analytical inputs, consistent with mixed-methods integration principles (Fetters, Curry and Creswell, 2013). Within each subdomain, source-derived components were then compared to determine whether they should be merged, reworded, reclassified, retained as distinct, or absorbed into broader component formulations.

The resulting pre-panel conceptual structure was subsequently reviewed by an independent validation panel comprising two patients with multimorbidity, one family physician, one nurse, one digital health systems engineer, and one health system decision-maker. Panel members independently assessed candidate components for clarity, relevance, and domain/subdomain allocation, and identified unnecessary overlap, insufficient differentiation, and wording requiring refinement. Inclusion, rewording, reclassification, consolidation, or exclusion decisions required unanimous agreement, following an expert-based content validation approach. The validated components were then organized into a hierarchical conceptual structure in accordance with MCDA-oriented structuring principles, aiming for comprehensive coverage, minimal redundancy, and components that were as preference-independent as practicable (Belton and Stewart, 2002). Although the components were allocated to predefined domains and subdomains, some remained conceptually connected across multiple dimensions, such as coordination and continuity. These overlaps were treated as inherent to the interdependent nature of multimorbidity healthcare services rather than as defects in the structure.

III. Results

The results supported the specification of the evaluative architecture underlying the Conceptual Structure for Value-Based Evaluation of Multimorbidity Healthcare Services (VEMS). The resulting architecture comprised three domains (Quality of Care, Service Delivery, and Costs) and eight associated subdomains. These subdomains defined the principal dimensions through which value in multimorbidity healthcare services was examined, beyond outcome-centered assessment alone.

A. Domains and Subdomains

The targeted review of relevant literature supported the specification of the subdomain structure used in the study. Comparison and consolidation of recurring evaluative dimensions across the three predefined domains resulted in eight subdomains.

The resulting structure comprised three subdomains for Quality of Care (Structural, Process, and Outcomes), three for Service Delivery (Performance, Safety, and Experience), and two for Costs (Direct and Indirect). These subdomains established the evaluative architecture used in the subsequent stages of component derivation, consolidation, and validation. Table I presents the final domain and subdomain structure, along with the operational definition for each subdomain.

TABLE I. Domains and Subdomains of the VEMS Conceptual Structure

Domain	Subdomain	Operational Definition
Quality of Care	Structural	Organizational, workforce, equipment, and physical/digital infrastructure capacities necessary to support the delivery of care.
	Process	Clinical and care-management activities undertaken throughout planned care episodes, including assessment, monitoring, treatment management, shared care planning, patient engagement, self-management support, and coordination mechanisms directly related to care provision.
	Outcomes	Clinical, functional, and patient-reported results associated with care over time, including the suitability of care in relation to patient context and needs.
Service Delivery	Performance	Operational performance of care delivery across settings and contact modes, including

Domain	Subdomain	Operational Definition
		access, timeliness, continuity, responsiveness, referral execution, attendance, and initiation of planned support.
	Safety	Prevention, detection, and management of risks and harm in care delivery, including patient safety, information governance, and digital system safety.
	Experience	Reported patient and professional experiences of care delivery, including interaction quality, patient-centredness, usability, navigability, and perceived usefulness across remote and in-person contacts.
Costs	Direct	Monetary expenses directly attributable to providing or receiving care, including consultations, medication, diagnostic testing, procedures, and related non-medical service costs.
	Indirect	Time, productivity, administrative, access-related, and other opportunity costs associated with care for patients, caregivers, and healthcare professionals.

B. Component Consolidation

Across the three data sources, the consolidation process resulted in a pre-panel set of 67 candidate components. Table II summarizes the contribution of each source by subdomain, including the number of observation-derived, interview-derived, and document-derived elements, the total number of source-derived elements, the number supported by all three sources, and the consolidation ratio. These inputs provided complementary evidence on how value in multimorbidity healthcare services is shaped by care processes, service-delivery conditions, organizational arrangements, and cost-related factors.

TABLE II. COMPONENT CONSOLIDATION ACROSS DATA SOURCES BY SUBDOMAIN

Domain	Subdomain	Observations	Interviews	Documents	Total	Three-source Support	Candidate Components	Consolidation Ratio[a]
Quality of Care	Structural	12	7	6	25	4	10	2.5
	Process	28	9	11	48	10	11	4.36
	Outcomes	0	6	21	27	5	8	3.38
Service Delivery	Performance	12	12	9	33	7	12	2.75
	Safety	0	1	3	4	2	3	1.33
	Experience	2	13	5	20	9	13	1.54
Costs	Direct	4	3	5	12	3	6	2
	Indirect	4	4	1	9	0	4	2.25
Total		62	55	61	178	40	67	2.66

[a] Calculated as total source-derived elements divided by candidate components

The distribution of source-derived elements varied substantially across subdomains, which indicates differences in evidential density and degree of overlap. The Process subdomain within Quality of Care concentrated the largest number of source-derived elements (48), followed by the Performance subdomain within Service Delivery (33) and the Outcomes subdomain within Quality of Care (27). These values indicate broader coverage and greater conceptual density in these parts of the structure. By contrast, the Safety subdomain within Service Delivery showed the smallest evidential base (4 source-derived elements), while the Indirect subdomain within Costs also remained comparatively limited (9).

Three-source support was highest in the Process subdomain of Quality of Care (10 components), followed by the Experience subdomain of Service Delivery (9) and the Performance subdomain of Service Delivery (7). The Indirect subdomain of Costs showed no three-source support, which indicates that this part of the structure remained more source-specific and less consistently represented across the three analytical inputs. The consolidation ratio was highest in the Process subdomain of Quality of Care (4.36), indicating stronger compression of overlapping or closely related inputs. The lowest ratio was observed in the Safety subdomain of Service Delivery (1.33), where source-derived elements were retained as distinct candidate components more often.

Source contributions also differed by subdomain. The Service Experience Blueprint contributed most strongly to the Process subdomain of Quality of Care (28 elements) and to the Performance subdomain of Service Delivery (12), reflecting the prominence of care-process organization, coordination demands, documentation burden, system interaction, and workflow constraints in observed practice. Stakeholder-informed analysis contributed most strongly to the Experience subdomain of Service Delivery (13 elements) and to the Performance subdomain of Service Delivery (12), indicating the salience of communication, access, digital interaction, navigability, and follow-up in stakeholder accounts. The structured document review contributed most strongly to the Outcomes subdomain of Quality of Care (21 elements) and to the Process subdomain of Quality of Care (11), reflecting the prominence of outcome sets, clinically oriented follow-up practices, and evaluative content embedded in existing instruments and reference frameworks. Through this consolidation process, the source-derived elements were organized into an integrated candidate component set within the predefined domains and subdomains.

C. VEMS Conceptual Structure

Following component consolidation and external panel review, the VEMS Conceptual Structure comprised 51 validated components distributed across the three domains and eight subdomains defined in the study. Of these, 19 components were allocated to Quality of Care, 18 to Service Delivery, and 14 to Costs. The domain-level architecture established in the earlier stages of the study was retained, while the component set was refined to improve clarity, reduce redundancy, and strengthen internal coherence.

Within Quality of Care, the validated structure retained components related to structural conditions for care delivery, core care management processes and patient-relevant outcomes associated with multimorbidity follow-up. Structural components concerned equipment, digital infrastructure, technical support, team capacity, and coordination capabilities across care settings. Process components concerned integrated condition management, remote monitoring, multidimensional assessment, self-management support, shared care planning, and follow-up by community services. Outcome components concerned biometric monitoring, context-adjusted clinical targets, functional capacity, health-related quality of life, motivation for self-management, health literacy, and avoidable hospital admissions.

Within Service Delivery, the validated structure retained components related to operational performance, safety, and reported experience across face-to-face and digitally supported modes of care. Performance components included responsiveness of communication and decision-making, waiting times, missed appointments, use of hybrid and digital consultations, active care planning, referral to community services, and assessment of informal social support. Safety components are concerned with adverse events, data protection, information governance and safety-related conditions of healthcare facilities. Experience components included professional-patient communication, experience with digital and face-to-face consultations, mental well-being of healthcare professionals, the usability and perceived usefulness of digital platforms, and the ability to navigate the healthcare system.

Within Costs, the validated structure retained both direct and indirect burdens associated with providing and receiving multimorbidity care. Direct components included expenditure on consultations, care planning, medication, chronic polypharmacy, avoidable therapeutic duplication, diagnostic tests and procedures, comparative expenditure between digital and face-to-face consultations, and avoidable repeated diagnostic testing. Indirect components included transport-related costs, digital access costs, non-reimbursed medication expenditure, administrative burden outside direct clinical activity, loss of active life days due to avoidable decompensations, and informal caregiver time burden.

The validated component set extended representation of value in multimorbidity healthcare services beyond outcome-centered assessment. Organizational, operational, experiential, and cost-related conditions were incorporated into the final structure, particularly in relation to digitally supported care delivery. The VEMS Conceptual Structure, therefore, provides a coherent conceptual basis for subsequent framework development and for more explicit representation of value in complex healthcare service systems.

IV. Discussion

The final validated structure indicated that value in multimorbidity healthcare services is not concentrated solely in outcomes, but distributed across Quality of Care (19 components), Service Delivery (18), and Costs (14). This distribution shifts the evaluative focus from isolated results to the wider set of conditions through which those results are produced. In multimorbidity care, performance is generated through the interaction of organizational capacity, care processes, service-delivery arrangements, and resource use over time. This interpretation is consistent with structure, process, and outcome logic and with integrated care perspectives in which outcomes depend on the conditions through which care is delivered (Donabedian, 1988; Valentijn *et al.*, 2013; Leijten *et al.*, 2018).

The pre-panel integration stage showed a similar pattern. The largest concentration of source-derived elements was found in the Process subdomain of Quality of Care (48), followed by the Performance subdomain of Service Delivery (33) and the Outcomes subdomain of Quality of Care (27). In addition, the Process subdomain showed the highest number of components supported by all three sources (10) and the highest consolidation ratio (4.36). This pattern suggests that multimorbidity care cannot be adequately evaluated solely through outcomes, because a substantial part of value generation depends on the care-management work required to sustain coordination, monitoring, and continuity over time.

This interpretation remained visible in the validated structure. Within Quality of Care, the final component set retained structural and process-related components, such as multidimensional assessment, shared care planning, remote monitoring, self-management support, and follow-up by community services, alongside patient-relevant outcomes, including functional capacity, health-related quality of life, health literacy, and avoidable hospital admissions. Institutional instruments used for monitoring and accountability remain necessary, particularly within primary care contractualization (Administração Central do Sistema de Saúde, 2025). However, guidance on multimorbidity and outcome-set research indicates that clinically appropriate care often depends on coordinated follow-up, treatment adaptation, patient-relevant outcomes, and the management of competing priorities, rather than on uniform target achievement alone (Muth *et al.*, 2014, 2019; Smith *et al.*, 2018). Under these conditions, indicator attainment may provide only a partial account of quality (Ruscitto *et al.*, 2016; Ho *et al.*, 2021).

The Service Delivery domain clarified the operational implications of this broader architecture. In the pre-panel structure, the Performance subdomain comprised 33 source-derived elements and 7 components with three-source support, while the Experience subdomain comprised 20 source-derived elements and 9 components supported by all three sources. This indicates that responsiveness, access, communication, navigability, and continuity were recurrent dimensions across observation, stakeholder accounts, and document-based analysis. Their retention in the final structure through components such as responsiveness of communication and decision-making, waiting times, missed appointments, active care planning, usability, perceived usefulness, and ability to navigate the healthcare system suggests that these are not secondary features of care. They are part of how value is produced and experienced in practice. Service Experience Blueprint observations support this interpretation, since multimorbidity follow-up depended on backstage work, coordination demands, documentation burden, system interaction, and workflow disruption. Comparable distributions of professional effort have been reported in ambulatory care, where documentation and electronic record interaction absorb a substantial share of clinical time (Sinsky *et al.*, 2016; Arndt *et al.*, 2017).

Digitalization should be interpreted within the same logic. In the validated structure, digital elements were distributed across infrastructure, technical support, usability, information governance, access, experience, and user costs. This suggests that digitalization in multimorbidity care is not simply an additional delivery channel but part of the service's operating conditions. The same pattern was already visible in the integration stage, where digital issues appeared across structural, performance, experience, safety, and cost-related elements. Evidence indicates that digital systems may improve continuity while also introducing workload, fragmentation, and usability burdens when poorly aligned with care processes (Vest and Kash, 2016; Atasoy, Greenwood and McCullough, 2019; Endalamaw *et al.*, 2025). Evaluation must therefore consider not only the presence of digital tools, but also their effects on the conditions through which care is delivered.

The Costs domain extended this argument by showing that fragmented care organizations have consequences that are not fully captured by direct expenditures alone. In the pre-panel integration, the Direct subdomain comprised 12 source-derived elements, 3 components supported by all three sources, and 6 candidate components, whereas the Indirect subdomain comprised 9 source-derived elements, 0 components supported by all three sources, and 4

candidate components. This asymmetry suggests not that indirect burdens are unimportant, but that they remain less consistently represented across observation, stakeholder accounts, and current document-based frameworks. In the validated structure, direct costs included consultations, medication, diagnostic testing, avoidable therapeutic duplication, and repeated testing within the same episode. Indirect costs included transport-related burden, digital access costs, administrative burden outside direct clinical activity, loss of active life days, and informal caregiver time. These components suggest that weaknesses in coordination, follow-up, or service organization may lead to avoidable expenditure and shift part of the burden to patients, caregivers, and professionals. This matters because productivity losses, time burden, and related opportunity costs are recognized in the health-economic literature, yet are often inconsistently represented in routine evaluation practice due to measurement complexity (Krol and Brouwer, 2014). When such burdens are omitted, assessment systems become less informative for improvement and less useful for management (Mannion and Braithwaite, 2012; de Melo Santos, Barbosa and Sant'Anna, 2025).

A methodological implication of these results concerns the role of indicator-based evaluation in multimorbidity care. The conceptual structure specified in this study does not displace existing indicator systems, but situates them within a broader evaluative architecture that also represents structural conditions, care processes, service-delivery features, and cost-related consequences. Under these conditions, measured outputs are better interpreted as part of a wider system of value generation rather than as self-sufficient representations of performance. This is consistent with the view that evaluation for decision support should incorporate not only observable results, but also the conditions and relationships through which those results are produced (Thokala *et al.*, 2016; Angelis and Kanavos, 2017; Baltussen *et al.*, 2017).

Several limitations should nevertheless be acknowledged. The subdomain structure was designed to reduce overlap and improve analytical coherence, but the distinction between some subdomains, particularly between Process within Quality of Care and Performance within Service Delivery, remained partly interpretive and may require further refinement through empirical use. The external validation panel strengthened clarity, coherence, and component allocation, but its composition was necessarily limited and may not capture the full range of clinical, managerial, and user perspectives. The final structure should therefore be understood as a validated conceptual architecture rather than as a fully operational evaluation framework. Its practical robustness still depends on subsequent empirical testing, refinement of component boundaries where necessary, and operationalization for use in healthcare service evaluation.

V. Conclusions

This study developed the Conceptual Structure for Value-Based Evaluation of Multimorbidity Healthcare Services (VEMS) as a broader evaluative architecture for complex healthcare service systems. The resulting structure showed that value in multimorbidity healthcare services is distributed across Quality of Care, Service Delivery, and Costs, and cannot be adequately represented by outcome-centered assessment alone. By integrating literature-informed subdomains, Service Experience Blueprint observations, stakeholder input, and current evaluation instruments, the study specified a validated set of components that makes visible the organizational, operational, experiential, and cost-related conditions through which performance and value are generated. The VEMS Conceptual Structure, therefore, provides a structured foundation for subsequent framework development and for more explicit evaluation, improvement, and management of multimorbidity healthcare services.

References

Administração Central do Sistema de Saúde, IP. (2025) Operacionalização da Contratualização Interna 2025. Lisboa.

Administração Central do Sistema de Saúde, IP. (2019) Guia para Aplicação do Diagnóstico de Desenvolvimento Organizacional nos Cuidados de Saúde Primários. Lisboa.

Akpan, A. et al. (2018) "Standard set of health outcome measures for older persons," BMC Geriatrics, 18, p. 36. Available at: https://doi.org/10.1186/S12877-017-0701-3/TABLES/5.

Angelis, A. and Kanavos, P. (2017) "Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework," Social science & medicine (1982), 188, pp. 137–156. Available at: https://doi.org/10.1016/J.SOCSCIMED.2017.06.024.

Arndt, B.G. et al. (2017) "Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations," Annals of Family Medicine, 15(5), pp. 419–426. Available at: https://doi.org/10.1370/AFM.2121.

Atasoy, H., Greenwood, B.N. and McCullough, J.S. (2019) "The Digitization of Patient Care: A Review of the Effects of Electronic Health Records on Health Care Quality and Utilization," Annual review of public health, 40, pp. 487–500. Available at: https://doi.org/10.1146/ANNUREV-PUBLHEALTH-040218-044206.

Baltussen, R. et al. (2017) "Value Assessment Frameworks for HTA Agencies: The Organization of Evidence-Informed Deliberative Processes," Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 20(2), pp. 256–260. Available at: https://doi.org/10.1016/J.JVAL.2016.11.019.

Belton, V. and Stewart, T.J. (2002) Multiple Criteria Decision Analysis, Multiple Criteria Decision Analysis. Springer Science & Business Media. Available at: https://doi.org/10.1007/978-1-4615-1495-4.

Bitner, M.J., Ostrom, A.L. and Morgan, F.N. (2008) "Service blueprinting: A practical technique for service innovation," California Management Review, 50(3), pp. 66–94. Available at: https://doi.org/10.2307/41166446.

Böhmann, T., Leimeister, J.M. and Möslein, K. (2014) "Service systems engineering," Business and Information Systems Engineering, 6(2), pp. 73–79. Available at: https://doi.org/10.1007/S12599-014-0314-8.

Braun, V. and Clarke, V. (2006) "Using thematic analysis in psychology," Qualitative Research in Psychology, 3(2), pp. 77–101. Available at: https://doi.org/10.1191/1478088706QP063OA.

Choong, K.K. (2013) "Understanding the features of performance measurement system: A literature review," Measuring Business Excellence, 17(4), pp. 102–121. Available at: https://doi.org/10.1108/MBE-05-2012-0031.

Dalglish, S.L., Khalid, H. and McMahon, S.A. (2021) "Document analysis in health policy research: the READ approach," Health Policy and Planning, 35(10), pp. 1424–1431. Available at: https://doi.org/10.1093/HEAPOL/CZAA064.

Donabedian, A. (1988) "The quality of care. How can it be assessed?," JAMA, 260(12), pp. 1743–1748. Available at: https://doi.org/10.1001/JAMA.260.12.1743.

Endalamaw, A. et al. (2025) "A scoping review of digital health technologies in multimorbidity management: mechanisms, outcomes, challenges, and strategies," BMC Health Services Research, 25(1), pp. 382-. Available at: https://doi.org/10.1186/S12913-025-12548-5/PEER-REVIEW.

Fetters, M.D., Curry, L.A. and Creswell, J.W. (2013) "Achieving integration in mixed methods designs-principles and practices," Health services research, 48(6 Pt 2), pp. 2134–2156. Available at: https://doi.org/10.1111/1475-6773.12117.

Ho, I.S.S. et al. (2021) "Examining variation in the measurement of multimorbidity in research: a systematic review of 566 studies," The Lancet Public Health, 6(8), pp. e587–e597. Available at: https://doi.org/10.1016/S2468-2667(21)00107-9.

Hsieh, H.F. and Shannon, S.E. (2005) "Three approaches to qualitative content analysis," Qualitative health research, 15(9), pp. 1277–1288. Available at: https://doi.org/10.1177/1049732305276687.

Krol, M. and Brouwer, W. (2014) "How to estimate productivity costs in economic evaluations," PharmacoEconomics, 32(4), pp. 335–344. Available at: https://doi.org/10.1007/s40273-014-0132-3.

Leijten, F.R.M. et al. (2018) "The SELFIE framework for integrated care for multi-morbidity: Development and description," Health policy (Amsterdam, Netherlands), 122(1), pp. 12–22. Available at: https://doi.org/10.1016/J.HEALTHPOL.2017.06.002.

Mannion, R. and Braithwaite, J. (2012) "Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service," Internal medicine journal, 42(5), pp. 569–574. Available at: https://doi.org/10.1111/J.1445-5994.2012.02766.X.

de Melo Santos, C.J., Barbosa, A.S. and Sant'Anna, Â.M.O. (2025) "Performance measurement systems in primary health care: a systematic literature review," BMC Health Services Research, 25(1), p. 353. Available at: https://doi.org/10.1186/S12913-025-12412-6.

Muth, C. et al. (2014) "The Ariadne principles: how to handle multimorbidity in primary care consultations," BMC medicine, 12(1). Available at: https://doi.org/10.1186/S12916-014-0223-1.

Muth, C. et al. (2019) "Evidence supporting the best clinical management of patients with multimorbidity and polypharmacy: a systematic guideline review and expert consensus," Journal of internal medicine, 285(3), pp. 272–288. Available at: https://doi.org/10.1111/JOIM.12842.

Nano, J. et al. (2020) "A standard set of person-centred outcomes for diabetes mellitus: results of an international and unified approach," Diabetic Medicine, 37(12), pp. 2009–2018. Available at: https://doi.org/10.1111/DME.14286.

OECD (2019) Measuring What Matters: the Patient-Reported Indicator Surveys. Paris. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/11/measuring-what-matters_08f342c3/2148719d-en.pdf (Accessed: May 7, 2026).

OECD (2023) Integrating Care to Prevent and Manage Chronic Diseases: Best Practices in Public Health, OCDE Publishing. Paris. Available at: https://doi.org/10.1787/9acc1b1d-en.

Porter, M.E. (2010) "What Is Value in Health Care?," New England Journal of Medicine, 363(26), pp. 2477–2481. Available at: https://doi.org/10.1056/NEJMP1011024/SUPPL_FILE/NEJMP1011024_DISCLOSURES.PDF.

De Rosis, S., Ferrè, F. and Pennucci, F. (2022) "Including patient-reported measures in performance evaluation systems: Patient contribution in assessing and improving the healthcare systems," The International Journal of Health Planning and Management, 37(Suppl 1), p. 144. Available at: https://doi.org/10.1002/HPM.3596.

Ruscitto, A. et al. (2016) "Accounting for multimorbidity in pay for performance: a modelling study using UK Quality and Outcomes Framework data," The British Journal of General Practice, 66(649), p. e561. Available at: https://doi.org/10.3399/BJGP16X686161.

Scheurer, D. et al. (2016) "The Value Equation: Enhancing Patient Outcomes While Constraining Costs," The American Journal of the Medical Sciences, 351(1), pp. 44–51. Available at: https://doi.org/10.1016/J.AMJMS.2015.10.013.

Schiøtz, M.L., Høst, D. and Frølich, A. (2016) "Involving Patients with Multimorbidity in Service Planning: Perspectives on Continuity and Care Coordination," Journal of Comorbidity, 6(2), pp. 95–102. Available at: https://doi.org/10.15256/JOC.2016.6.81.

Scupin, R. (1997) "The KJ method: A technique for analyzing data derived from Japanese ethnology," Human Organization, 56(2), pp. 233–237. Available at: https://doi.org/10.17730/HUMO.56.2.X335923511444655.

Sinsky, C. et al. (2016) "Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties," Annals of internal medicine, 165(11), pp. 753–760. Available at: https://doi.org/10.7326/M16-0961.

Smith, S.M. et al. (2018) "A Core Outcome Set for Multimorbidity Research (COSmm)," Annals of Family Medicine, 16(2), pp. 132–138. Available at: https://doi.org/10.1370/afm.2178.

Snyder, H. (2019) "Literature review as a research methodology: An overview and guidelines," Journal of Business Research, 104, pp. 333–339. Available at: https://doi.org/10.1016/J.JBUSRES.2019.07.039.

Teisberg, E., Wallace, S. and O'Hara, S. (2020) "Defining and Implementing Value-Based Health Care," Academic Medicine, 95(5), pp. 682–685. Available at: https://doi.org/10.1097/ACM.0000000000003122.

Thokala, P. et al. (2016) "Multiple Criteria Decision Analysis for Health Care Decision Making - An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force," Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 19(1), pp. 1–13. Available at: https://doi.org/10.1016/J.JVAL.2015.12.003.

Valentijn, P.P. et al. (2013) "Understanding integrated care: a comprehensive conceptual framework based on the integrative functions of primary care," International Journal of Integrated Care, 13(1). Available at: https://doi.org/10.5334/IJIC.886.

Vest, J.R. and Kash, B.A. (2016) "Differing Strategies to Meet Information-Sharing Needs: Publicly Supported Community Health Information Exchanges Versus Health Systems' Enterprise Health Information Exchanges," The Milbank Quarterly, 94(1), p. 77. Available at: https://doi.org/10.1111/1468-0009.12180.

Xu, X., Mishra, G.D. and Jones, M. (2017) "Evidence on multimorbidity from definition to intervention: An overview of systematic reviews," Ageing Research Reviews, 37, pp. 53–68. Available at: https://doi.org/10.1016/J.ARR.2017.05.003.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Machine Learning and Failure Mode and Effects Analysis Integration in Manufacturing: A Systematic Literature Review

Francolino Agostinho Sanca [1], José Pedro Teixeira Domingues [2], Ana Maria A. C. Rocha [3], and Emerson José de Paiva [4,]

[1] University of Minho, Algoritmi Research Centre/LASI, Guimarães, Portugal; id11361@uminho.pt
[2] University of Minho, Algoritmi Research Centre/LASI, Guimarães, Portugal; pdomingues@dps.uminho.pt
[3] University of Minho, Algoritmi Research Centre/LASI, Guimarães, Portugal; arocha@dps.uminho.pt
[4] University of Minho, Algoritmi Research Centre/LASI, Guimarães, Portugal; emersonpaiva@unifei.edu.br; Federal University of Itajubá, GEOProd, Itajubá, Brazil; emersonpaiva@unifei.edu.br

***Abstract*—Machine learning (ML) has been increasingly applied in manufacturing to support quality improvement, defect analysis, and predictive decision-making. At the same time, Failure Mode and Effects Analysis (FMEA) remains one of the most widely used tools for identifying potential failure modes, analyzing their causes and effects, and supporting preventive actions in industrial settings. Although previous reviews have addressed related topics, the literature still lacks a focused and systematic analysis of how FMEA and ML are integrated in manufacturing. To address this gap, this paper presents a systematic literature review based on the PRISMA framework. Publications indexed in Scopus and Web of Science between 2015 and 2026 were screened using predefined inclusion and exclusion criteria, resulting in a final of 21 papers. The selected studies were analyzed across multiple dimensions, including decision-support methods, FMEA elements addressed, ML families adopted, and FMEA-ML interaction categories, namely operational integration, semi-automated integration, and automated/closed-loop integration. The findings show that existing research is mainly concentrated on risk prioritization and failure mode/effect analysis, with classical ML serving as the dominant approach for adapting FMEA to the emerging Quality 4.0 paradigm. However, most studies remain at level of operational integration, where ML outputs are largely analyst-mediated or manually interpreted, rather than semi-automated or fully automated. This review contributes to the Quality 4.0 literature by clarifying how ML is being used to extend FMEA as a traditional quality and risk management tool, while also identifying the current maturity level and future development directions of FMEA-ML integration in manufacturing.**

Keywords – Machine Learning, FMEA, Manufacturing, Quality 4.0

I. Introduction

Industry 4.0 encompasses current industrial development trends and has been associated with several related domains, such as Quality 4.0, Agriculture 4.0, Service 4.0, Logistics 4.0, and Health 4.0 (Sader et al, 2022). Regarding Quality 4.0, its can be understood as a natural evolution of previous industrial phases, while maintaining a strong focus on customer needs through the delivery of increasingly customized products and services, while also accounting for the digitalization of processes and systems. Likewise, Pedro and Romero (2017) describe Quality 4.0 as an extended approach to quality management, in which recent technologies are integrated with traditional quality practices, including Quality Control (QC), Quality Assurance (QA), and Total Quality Management (TQM).

In this context, understanding Quality 4.0 requires considering the role of traditional quality tools, methods, and systems and how they can be combined with data-driven technologies to expand their capabilities as managerial instruments (Carvalho et al., 2020). Failure Mode and Effects Analysis (FMEA) and Machine Learning (ML) are two elements that fit well within this perspective, as the former represents a traditional quality risk management tool, whereas the latter represents a data-driven technological approach. Their integration may therefore enhance their value as managerial instruments within Quality 4.0.

FMEA is a technical risk analysis method used to identify the functions of a product or process step and the associated potential failure modes, effects, and causes. The method is commonly divided into Design FMEA (DFMEA) and Process FMEA (PFMEA), with the former typically associated with product analysis during the design phase and the latter with process definition and analysis during the production phase (AIAG and VDA, 2019). ML, in turn, refers to the ability of systems to learn from problem-specific training data to automate analytical model building and related tasks (Janiesch et al., 2021). These methods can be broadly categorized into classical ML and neural-network-based approaches. In industrial applications, common classical ML methods include Decision Trees (DT), Naïve Bayes (NB), Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN), whereas neural-network-based methods include Artificial Neural Network (ANN), Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), and Recurrent Neural Networks (RNN).

Previous review studies addressing FMEA and ML in manufacturing have approached the topic from different perspectives. Wu et al. (2021) reviewed the development and application of FMEA in manufacturing, focusing on failure mode identification, risk assessment, and industrial standards. They concluded that the literature has been largely centred on improving Risk Priority Number (RPN)-based risk ranking models, with limited connection to industrial standards and insufficient attention to automation. Although the authors identified ML and DL as promising enablers of FMEA automation, their review

did not specifically examine how ML and DL are applied to support FMEA in manufacturing, thereby leaving room for further systematic reviews.

Mokhtarzadeh et al. (2025) reviewed hybrid intelligence failure analysis in Industry 4.0 across FMEA, RCA, and FTA, based on 86 high-quality papers published between 1990 and the first half of 2022. Their review showed that existing research remains largely organized around individual failure-analysis methodologies, with most FMEA studies still relying on expert-driven approaches and offering only limited evidence of comprehensive hybrid-intelligence implementations. Although the review discusses intelligent tools such as Bayesian networks, neural networks, classification methods, and DL, it does not specifically and systematically examine how ML and DL are applied to support FMEA in manufacturing.

Liu et al. (2019) conducted a systematic literature review on the use of multi-criteria decision-making (MCDM) methods in FMEA for failure mode evaluation and prioritization, analysing 169 journal papers published between 1998 and 2018. Their review showed that prior research had been largely focused on improving FMEA through MCDM-based risk management, risk factor weighting, and failure mode ranking, particularly using methods such as GRA and TOPSIS. However, because that review focused specifically on studies applying MCDM methods, it does not address how ML and DL may support FMEA as a whole or in different elements in manufacturing contexts.

Yousef et al. (2025) presented a state-of-the-art review on AI- and ontology-based enhancements to FMEA in advanced systems engineering, highlighting how AI, ontologies, and MBSE can support a transition from traditional document-based FMEA to intelligent, model-driven, and semantically enriched risk analysis. Their review concluded that such approaches offer strong potential for explainable, adaptive, and lifecycle-integrated FMEA, although industrial adoption remains constrained by interoperability, standardization, scalability, and validation challenges. Nevertheless, this review has a broader scope than ML/DL applications in manufacturing, as it focuses on advanced systems engineering, ontology-based knowledge representation, and design-phase integration rather than systematically reviewing how ML and DL are applied to support FMEA in manufacturing settings.

Despite these related review studies, the literature still lacks a focused review on the integration of FMEA and ML in manufacturing. In particular, existing reviews do not sufficiently identify integration patterns across multiple analytical dimensions or classify the maturity of this integration in the literature. To address this gap, the present review analyses studies published between 2015 and 2026, with particular attention to decision-support methods, FMEA elements addressed, ML families adopted, and FMEA-ML interaction categories, including operational integration, semi-automated, and automated/closed-loop integration. Accordingly, the following research questions are addressed:

- **RQ1:** How are machine learning and FMEA integrated in manufacturing?
- **RQ2:** What is the maturity level of FMEA-ML implementation in manufacturing?

By addressing these questions, this systematic literature review contributes to the Quality 4.0 literature by showing how FMEA, as a traditional quality and risk management tool, is being expanded through machine learning and data-driven approaches. In doing so, the review highlights where the literature places great emphasis and how this interaction and/or integration is evolving in terms of automation.

The remainder of this paper is organized as follows. Section II presents the methodology adopted to conduct the systematic literature review. The theoretical framework of FMEA and ML in the context of Quality 4.0 is introduced in Section III. Section IV reports the review results. Finally, Section V presents the main conclusions and future research directions.

II. METHODOLOGY

This study adopts a Systematic Literature Review (SLR) and reports it in accordance with the PRISMA 2020 statement, which provides a 27-item checklist to support transparent and complete reporting of systematic reviews (Page et al. 2021). The research conducted considered publications published between 2015 and 2026. Two major academic databases, Scopus and Web of Science (WoS), were used for the literature search. The retrieved records were then organized and screened through a combined procedure: the PRISMA framework guided the study selection process through inclusion and exclusion criteria, while Bibliometrix, an R software package for bibliometric analysis proposed by Aria and Cuccurullo (2017), supported database management and duplicate identification. Based on these processes, a final corpus of papers was applied, which was subsequently analysed in this review.

A. Search strategy

This study was conducted by structuring a search strategy around three thematic blocks, representing core domains of this research:

- Risk assessment – "failure mode and effects analysis" OR "failure mode, effects and criticality analysis".
- Technical – "machine learning" OR "deep learning".
- Manufacturing – manufactur* OR industr* OR production OR "production line*" OR factory OR factories OR "industrial process*" OR "quality control" OR "process monitoring".

These thematic blocks were used in two major academic databases, Scopus and WoS. A preliminary search was conducted to identify publications containing the above keywords, combined with Boolean operators, in the titles, abstracts, and keywords list of the documents. The initial search retrieved 94 papers in Scopus and 62 in WoS.

TABLE I. RESULTS OF PRELIMINARY SEARCH

Database	*Search Strings*	*Number of papers*
Scopus	(TITLE-ABS-KEY (FMEA OR FMECA OR "failure mode and effects analysis" OR "failure mode, effects and criticality analysis")) AND (TITLE-ABS-KEY (“machine learning” OR “Deep learning”)) AND (TITLE-ABS-KEY (manufactur* OR industr* OR production OR "production line*" OR factory OR factories OR "industrial process*" OR "quality control" OR "process monitoring"))	94
Web of Science	(TOPIC (FMEA OR FMECA OR "failure mode and effects analysis" OR "failure mode, effects and criticality analysis")) AND (TOPIC (“machine learning” OR “Deep learning”)) AND (TOPIC (manufactur* OR industr* OR production OR "production line*" OR factory OR factories OR "industrial process*" OR "quality control" OR "process monitoring"))	62

B. Inclusion and exclusion criteria

To screen and refine the retrieved records, inclusion and exclusion criteria were established in accordance with the PRISMA 2020 framework. As shown in Table 2, the review was restricted to articles, conference papers, and early access articles published in English between 2015 and 2026, indexed in selected engineering and manufacturing subject areas, and focused on the integration of ML and FMEA (and its variants) in manufacturing contexts related to process, product, and maintenance. After applying these criteria, 103 papers met the eligibility requirements, of which 63 were retrieved from Scopus and 40 from WoS. The records retrieved from both databases were then merged and processed in Bibliometrix, which supported the identification and removal of 28 duplicates. The remaining records were screened by title and abstract and subsequently assessed for full-text eligibility. During this stage, 54 papers were excluded because, although related to ML or FMEA, they did not demonstrate any integration of both approaches within the manufacturing context defined in this review. Consequently, the final corpus comprised 21 papers from19 different sources and with a period between 2017 and 2026, as shown in Fig. 1.

TABLE II. INCLUSION AND EXCLUSION CRITERIA

Criteria	**Inclusion**	**Exclusion**
Period	2015-2026	<2015
Language	English	Others
Document type	Articles, Proceeding paper, Early Access	Others
Subject area	Engineering, Engineering Multidisciplinary, Computer science, Computer science theory methods Operations Research Management Science, Computer Science Information Systems, Engineering Industrial, Engineering Electrical Electronic, Engineering Manufacturing, Computer Science Artificial Intelligence, Engineering Mechanical	Others
Integration	Integration of ML and FMEA in manufacturing context regarding process, product, and maintenance, or even in manufacturing design stage	Others, such as FMEA applied to ML life cycle, or ML-FMEA integration

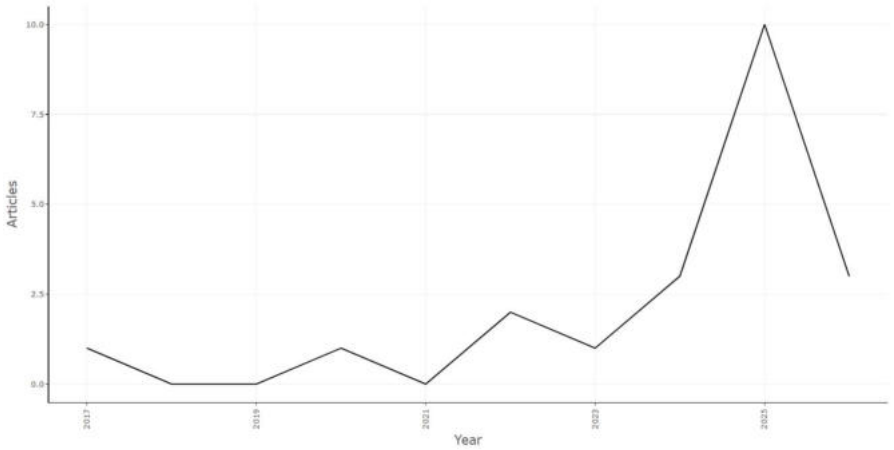


Fig. 1. Annual Scientific Production

III. Failure Mode and Effects Analysis-Machine Learning in the context of quality 4.0 - Concepts

FMEA is a team-oriented, systematic, qualitative, and analytical method used to evaluate potential technical risks of failure in products and processes. Its main objective is to identify product and process functions, the associated potential failure modes, their effects, and their causes. In addition, the method is used to assess whether the planned prevention and detection controls are sufficient and to recommend further actions when necessary. FMEA is commonly divided into DFMEA and PFMEA, with the former associated with product analysis during the design phase and the latter related to process definition during the production phase (AIAG & VDA, 2019). IEC (2018) defines FMEA as a systematic method for identifying failure modes, their effects on the item or process, and their causes. When the analysis also includes a ranking of criticality, involving at least the severity of consequences and often additional measures of importance, it is referred to as failure modes, effects, and criticality analysis (FMECA). Regarding Failure Mode, Effects, and Criticality Analysis (FMECA), U.S. Department of Defense, (1980) explains that this method results from combining FMEA with Criticality Analysis (CA), in which each potential failure mode is ranked according to the combined influence of severity and probability of occurrence.

ML is a subfield of artificial intelligence concerned with the development of computational models that learn patterns from data to support prediction, classification, and decision-making tasks. In practical terms, ML can be viewed as a computationally oriented form of applied statistics aimed at approximating complex functions from data (Goodfellow et al., 2016). In this review, the identified approaches were grouped into two broad ML families: classical ML and neural-network - based methods. Classical ML includes non-neural methods such as DT, NB, SVM, k-NN, linear models, and tree-based ensemble methods. Neural-network-based methods comprise architectures inspired by biological neurons, consisting of interconnected processing units that learn through weighted connections across input, hidden, and output layers (Goz et al., 2019). This family includes both conventional artificial neural networks (ANNs) and DL models, the latter being characterized by multiple layers of nonlinear processing and hierarchical feature learning (Abitami and Chitra, 2020). Common DL architectures include DNN, CNN, GAN, RNN.

Overall, the integration of FMEA and ML may be framed as an element of Quality 4.0, as this approach uses modern technologies to augment the capabilities of quality tools and methods (Dias et al., 2021).

IV. RESULTS

This section presents the results of how the reviewed studies integrate FMEA with ML. It examines FMEA-ML interaction patterns and the categories that characterize this integration. Together, these analyses provide an understanding of how literature keep handling decision support methods in the edge of AI, where the literature places greater emphasis, and how this interaction and/or integration is evolving in terms of automation.

A. Failure Mode and Effects Analysis-Machine Learning interaction patterns

FMEA is one of the most widely used risk assessment methods for supporting decisions aimed at improving products and processes. Although its potential has often remained underexploited, the rapid proliferation of AI technologies has created new opportunities to integrate FMEA with ML to extract more detailed and useful insights for decision-making..

Table III presents analytical dimensions of the FMEA-ML integration dimensions across 21 reviewed studies. These dimensions include the use of decision-support methods, particularly fuzzy and MCDM approaches; the FMEA elements addressed, including risk prioritisation, failure mode and failure effect analysis, and cause analysis; the ML family adopted, distinguishing between classical ML and DL/neural network. Beyond these dimensions, the table also present specific model(s) proposed in each study.

From the perspective of decision-support methods, the review focused specifically on fuzzy and MCDM approaches, as these gave received particular attention in the literature. This tendency is consistent with earlier observations, namely Liu et al. (2019), that prior FMEA improvement research has been largely concentrated on MCDM-based enhancements. The

reviewed studies confirm that this relevance persists, as seven papers still incorporate fuzzy or MCDM techniques in the integration of FMEA and ML.

Regarding FMEA elements, the literature places greater emphasis on failure mode analysis (16 studies) and risk prioritisation (15 studies) than on cause analysis (10 studies). The predominance of risk prioritisation is consistent with previous observations, including Wu et al. (2021), who find that the literature has been strongly centred on improving RPN-based risk assessment, which is still consistent with this reviewed literature. However, different from previous studies, in this reviewed study, recent studies seem increasingly engaging with failure mode and effect-oriented perspectives, moving beyond simple ranking toward failure mechanisms challenges. From the Table, the results also show that seven studies addressed all the FMEA elements represented in Table III. However, the presence of all elements does not necessarily imply more automated or comprehensive integration. In fact, as shown in Table IV, none of the reviewed studies reached a fully automated integration level.

From the perspective of the ML family, the integration of FMEA and ML is still predominantly associated with classical ML, which appears in 16 studies, compared with 8 studies using DL or neural networks. A possible explanation is that many FMEA-related analyses are often grounded in structured and tabular information, which may be more compatible with conventional ML approaches. This pattern may also be reinforced by the prominence of risk prioritisation in the reviewed corpus, as this task is predominantly linked to classical ML and appears in at least 12 studies.

Overall, the findings indicate that the reviewed literature has mainly concentrated on risk prioritisation and failure mode and effects, with classical ML serving as the principal vehicle for adapting FMEA to the emerging Quality 4.0 paradigm.

TABLE III. ANALYTICAL DIMENSIONS OF FMEA-ML INTEGRATION ACROSS THE REVIEWED STUDIES

Authors	Decision support methods	FMEA elements			ML family		Proposed model(s)
		Risk prioritisation	FM/FE analysis	Cause analysis	Classical ML	Neural network	
Ojha et al. (2026)			✓	✓	✓		Data-driven preventive maintenance and fault analysis using multivariate linear regression and Random Forest, with FMEA
Wiangkham and Vongvit (2026)		✓			✓		ML-enhanced FMEA using XGBoost surrogate modelling, SHAP, and clustering (K-Means, Agglomerative, GMM)
Adriel and Harito (2025)	✓	✓	✓	✓	✓		DMAIC + Fishbone Diagram + Fuzzy FMEA + Random Forest-based multi-class defect classification
Bellido-Lopez et al. (2025)		✓	✓		✓	✓	FMEA-informed anomaly detection using NOC models (LR, kNN, DT, SVM, and MLP) and risk curves for real-time failure-mode monitoring and maintenance-effectiveness evaluation.
Bubeník et al. (2025)		✓			✓		AI-enhanced FMEA using Random Forest Regressor, Randon Forest Classifier, and XGBoost for RPN prediction and risk classification
Fei et al. (2025)			✓		✓	✓	ERNIE/BERT-based text vectorization + unsupervised clustering (K-means/DBSCAN/HAC) + LDA keyword extraction + DeepSeek-R1-Qwen-32B
Ferreira et al. (2025)			✓	✓	✓	✓	Autoenconder-based fault detection + XGBoost diagnosis + XAI (LIME, SHAP, SHAP-IQ) to support FMEA/RCA
Grigoryan et al. (2025)		✓	✓		✓		Semi-supervised fault detection and classification using K-means + similarity-based classification + cosine, integrated with data-driven FMEA for RPN
Kucejko and Bugdol (2025)			✓			✓	DFMEA recommendation system using neural network embeddings, Sentence-BERT, TF-IDF, cosine similarity, and hybrid recommendation
Ojeda et al. (2025)	✓	✓	✓	✓	✓		AHP + PFMEA-guided predictive maintenance using Random Forest to predict failure cycles for the critical crack failure mode
Oliveira et al. (2025)			✓			✓	FMEA-based prescriptive DL model using multivariate time-series and MLP/CNN/FCN/Multi-head CNN for failure-mode/repair guidance
Shenoy et al. (2025)			✓	✓		✓	Neurosymbolic FMEA framework using ontologies, knowledge graphs, and neural models (GNN/CNN/LSTM)
Yilmaz and Kose (2025)		✓	✓	✓	✓		Hybrid FMEA and Random Forest regression for RPN prediction and FMEA-derived risk factors
Duan et al. (2024)	✓	✓	✓	✓	✓		Improved FMEA combining knowledge graph, AHP, GRA, and K-means clustering for prioritisation and classification
Liang et al. (2024)	✓	✓	✓		✓	✓	Improved FMEA using TF-IDF, word2vec, component/failure dictionairies, k-means clustering, and multi-view group consensus
Song and Zheng (2024)	✓	✓	✓	✓		✓	BERT + VSM + entropy weighting + TOPSIS for automatic prediction of S/O/D from engineering textual data and FMEA
Sarmast et al. (2023)	✓	✓	✓	✓	✓		Ontology-based social media mining with Naïve Bayes and a DSS to identify defective components, symptoms, and causes, then support FMEA-based prioritisation and repair decisions.
Antomarioni et al. (2022)		✓	✓	✓	✓		Data-driven FMECA framework using Association Rule Mining (FP-growth) and Social Network Analysis
Bhattacharjee et al. (2020)		✓			✓		Industrial equipment FMEA framework using interval numbers, linguistic variables, and Logistic Regression
Bhattacharjee et al. (2022)		✓			✓		Interval-number-based multimonial logistic regression (MLR) for FMEA and PFM

Rezaee et al. (2017)	✓	✓					Multi-stage cognitive map integrated with PFMEA using Extended Delta Rule, Nonlinear Hebbian Learning (NHL), and Differential Evolution (DE)
Total	7	15	16	10	16	8	

B. Categories of Failure Mode and Effects Analysis-Machine Learning interaction

The analysis of the literature indicates that FMEA-ML integration can be categorized into different levels of interaction. These include Descriptive/Informing, Operational Integration, Semi-automated, and Automated/closed-loop, as presented in Table IV. In addition to these integration-oriented categories, the table also includes a Conceptual category. Although this category does not reflect the same level of practical implementation, it was defined to group studies that propose conceptual frameworks for integrating FMEA and ML without demonstrating an operational implementation. Shenoy et al. (2025) is the only paper in the reviewed corpus classified in this category, as it proposes a neurosymbolic framework for semiconductor defect diagnosis by combining a symbolic FMEA-based layer, including ontology and knowledge graph components, with neural models such as GNN, CNN, and Long Short-Term Memory (LSTM), without fully implementing the proposed architecture. If implemented, this framework could potentially automate multiple FMEA-related outputs.

This review identified three papers in the Descriptive/Informing category. These studies were grouped here because FMEA or ML was mainly used to describe, organize, interpret, or report findings for the other side, without clearly demonstrating strong integration. Oliveira et al. (2025) is a representative example of this category, since FMEA-defined failure modes are mainly used as structured class labels for multi-class ML classification, rather than as part of a strongly integrated FMEA-ML workflow.

Regarding the Operation Integration category, the review identified 13 papers, making it the dominant category in the reviewed corpus. This may suggest that the literature on FMEA-ML integration is concentrated at a level where FMEA and ML are used together in practical workflows, but where the outputs of the integration are still not fully automated. In other words, after model training, the integration remains largely analyst-mediated and/or manually interpreted. Representative example of the Operational Integration category includes Yilmaz and Köse (2025), Ojha et al. (2026), and Antomarioni et al. (2022). Yilmaz and Köse (2025), FMEA is operationally combined with Random Forest-based RPN prediction, Pareto analysis, feature importance, and Ishikawa diagrams to support defect prioritization and improvement planning. In Ojha et al. (2026), a data-driven framework integrates IoT-enabled monitoring, machine learning, fault analysis, preventive maintenance, and FMEA within the same workflow. In Antomarioni et al. (2022), FMECA is extended through association rule mining and social network analysis to support the interpretation of failure relationships and maintenance-oriented decision-making. In all three cases, FMEA and ML/data-driven techniques are used together in a practical and meaningful way, but the integration remains largely analyst-mediated and manually interpreted rather than autonomously executed by a developed model.

About the Semi-automated category, the review identified four papers that fit this classification. These studies propose models that automatically or semi-automatically perform, support, or update at least one explicit FMEA-related task or element. This may suggest that, in the relatively few cases where FMEA-ML integration becomes more automated, the automation still tends to focus on specific FMEA-related tasks rather than on the entire FMEA workflow or its core elements a whole. Kucejko and Bugdol (2025) provide a clear example, as they integrate a neural-network-based recommendation system into DFMEA to automate the failure mode-failure effect relationship, in other words, given a failure mode, the system attempts to suggest the most plausible failure effect. Likewise, Song and Zheng (2024) show how textual FMEA-related inputs can be used to automatically estimate S/O/D-related information and support risk ranking.

Finally, the last category identified in this review is Automated/closed loop. This category was defined to capture studies in which trained models autonomously generate, update, or refine multiple FMEA-related outputs when new inputs are provided, thereby enabling dynamic feedback, continuous improvement, or real-time updating across the FMEA workflow with minimal human intervention. No paper in the reviewed corpus fully met this definition. This absence may suggest that fully automated and closed-loop FMEA-ML integration remains an open gap in the literature, indicating that the maturity of FMEA-ML integration is still limited. Although Shenoy et al. (2025) may be regarded as a candidate for this category, the framework is not fully implemented to a degree that justifies its formal inclusion here.

TABLE IV. CATEGORIZATION OF FMEA AND ML INTEGRATION

Category	Description	N° of papers	Papers
Conceptual	The study proposes a conceptual framework, architecture, or theoretical linkage between FMEA and ML, but does not demonstrate an operational workflow showing how outputs from one side are actually used by the other.	1	Shenoy et al. (2025)
Descriptive/Informing	The study uses FMEA or ML mainly to describe, organise, interpret, or report findings from the other side, without clearly showing strong integration	3	Oliveira et al. (2025); Adriel and Harito (2025); Ferreira et al. (2025)
Operation Integration	The study demonstrates a practical use of FMEA and ML within the same workflow, where outputs from one side support activities on the other side. However, the proposed model does not autonomously perform or update an explicit FMEA-related task or element after training. The integration remains operational, but largely analyst-mediated, manually interpreted.	13	Ojha et al. (2026); Yilmaz and Kose (2025); Liang et al. (2024); Antomarioni et al. (2022); Bhattacharjee et al. (2020); Wiangkham and Vongvit (2026); Bhattacharjee et al. (2022); Sarmast et al. (2023); Rezaee et al. (2017); Duan et al. (2024); Ojeda et al. (2025); Bellido-Lopez et al. (2025);
Semi-automated	The study demonstrates that the proposed model automatically or semi-automatically performs, supports, or updates at least one explicit FMEA-related task or element, such as identifying failure modes, recommending failure effects, estimating S/O/D-related	4	Kucejko and Bugdol (2025); Song and Zheng (2024); Grigoryan et al. (2025); Fei et al. (2025);

	information, calculating or predicting risk scores, or prioritizing failure modes. After training, the model should exhibit some degree of autonomy, being capable of generating outputs when the required inputs are provided. However, the automation remains limited to part of the FMEA workflow.		
Automated/closed loop	The study demonstrates automated workflow in which FMEA-related information and operational data are used to train and update ML models, while the trained models autonomously generate, update or refine multiple FMEA-related outputs or their core elements (failure modes, effects and causes) when new inputs are provided. The system therefore supports dynamic feedback, continuous improvement, or real-time updating across the FMEA workflow with minimal human intervention	0	None

V. CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH DIRECTIONS

This study investigated how FMEA has been integrated with ML in the manufacturing context through a systematic literature review of papers retrieved from the Scopus and Web of Science databases. A total of 103 papers were identified after using the initial inclusion and exclusion criteria of PRISMA. After removing 28 duplicates and excluding 54 papers during title, abstract, and full-text eligibility, a final corpus of 21 papers was retained for detailed analysis.

Although the final corpus is relatively focused, this review provides a structured and detailed synthesis of the literature across different analytical dimensions, including decision-support methods, FMEA elements addressed, ML families adopted, and FMEA-ML interaction categories, including operational integration, semi-automated, and automated/closed-loop integration.

The findings suggest that the reviewed literature has mainly concentrated on risk prioritisation and failure mode and effects, with classical ML serving as the principal vehicle for adapting FMEA to the emerging Quality 4.0 paradigm. Nevertheless, the results also show that FMEA-ML research remains largely concentrated in operational integration, where the outputs of integration are still mainly analyst-mediated and/or manually interpreted, rather than semi-automated or fully automated.

Future research should therefore place greater emphasis on semi-automated or, especially, fully automated approaches, in which trained models can generate, update, or refine multiple FMEA-related outputs when new inputs are provided. Such developments could enable more dynamic feedback with minimal human intervention. In addition, future systematic reviews could examine the role of ML in supporting other quality management tools such as SPC, and explore potential interrelationships among FMEA, ML, and SPC in data-driven quality management contexts.

This review has two main limitations. First, the final corpus is relatively small, which constrains the breadth of evidence considered and may limit the representativeness of the identified patterns. However, this also reflects the still-emerging nature of research on FMEA–ML integration in manufacturing. Second, the classification of studies according to interaction categories involves a degree of interpretive judgement, which may lead to different interpretations of the reviewed papers, although the categories were defined as explicitly and exhaustively as possible.

ACKNOWLEDGMENT

This work was financed by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the PhD Studentship [**PRT/BD/154899/2023**], Centro ALGORITMI strategic funding ref. UIDB/00319/2020, DOI 10.54499/UIDB/00319/2020, LASI – Intelligent Systems Associate Laboratory, ref. LA/P/0104/2020, and GEQProd – Grupo de Estudos em Qualidade e Produtividade, Unifei, Brazil.

REFERENCES

Abitami, S., and Chitra, P. (2020), "Energy-efficient edge based real-time healthcare support system". In P. Raj and P. Evangeline (Eds.), The digital twin paradigm for smarter systems and environments: The industry use cases, pp. 339-368.

Adriel, E., and Harito, C. (2025), "Implementation of DMAIC using machine learning to analyze product defects in the welding consumables industry", In 2025 11th International Conference on Control, Decision and Information Technologies (CoDIT), pp. 2922–2927.

AIAG and VDA (2019), AIAG & VDA FMEA Handbook: Design FMEA, Process FMEA, Supplemental FMEA for Monitoring & System Response, 1st ed., Automotive Industry Action Group and Verband der Automobilindustrie, Southfield, MI/Berlin.

Antomarioni, S., Ciarapica, F. E., and Bevilacqua, M. (2022), "Association rules and social network analysis for supporting failure mode effects and criticality analysis: Framework development and insights from an onshore platform", Safety Science, Vol. 50, p. 105711.

Aria, M., Cuccurullo, C. (2017), "bibliometrix: An R-tool for comprehensive science mapping analysis", Journal of Informetrics, Vol. 11, No. 4, pp. 959-975.

Bellido-Lopez, F. J., Sanz-Bobi, M. A., Muñoz, A., Gonzalez-Calvo, D., and Alvarez-Tejedor, T. (2025), "A novel method for evaluation of the maintenance impact in the health of industrial components", Results in Engineering, Vol. 27, p. 105809.

Bhattacharjee, P., Dey, V., Mandal, U. K., and Paul, S. (2022), "Quantitative risk assessment of submersible pump components using interval number-based multinomial logistic regression (MLR) model", Reliability Engineering & System Safety, Vol. 226, p. 108703.

Bhattacharjee, P., Dey, V., and Mandal, U. K. (2020), "Risk assessment by failure mode and effects analysis (FMEA) using an interval number based logistic regression model", Safety Science, Vol. 132, p. 104967.

Bubeník, P., Rakyta, M., Buzalka, M., Biňasová, V., and Kovaríková, Z. (2025), "Optimization of business processes using artificial intelligence", Electronics, Vol. 14, p. 2105.

Carvalho, A.M., Sampaio, P., Rebentisch, E. and Oehmen, J. (2020), "Technology and quality management: a review of concepts and opportunities in the digital transformation", Proceedings of the International Conference on Quality Engineering and Management (ICQEM 2020), Braga, Portugal, pp. 698-714.

Dias, A.M., Carvalho, A.M. and Sampaio, P. (2021), "Quality 4.0: literature review analysis, definition and impacts of the digital transformation process on quality", International Journal of Quality & Reliability Management, Vol. 39 No. 6, pp. 1312-1335.

Duan, C., Zhu, M., and Wang, K. (2024), "Reliability analysis of intelligent manufacturing systems based on improved FMEA combined with machine learning", Journal of Intelligent & Fuzzy Systems, Vol. 46, pp. 10375–10392.

Fei, Q., Mei, W., Qin, J., Zhou, Y., Gong, W., Zhao, K., Zhou, H., and Zou, G. (2025), "Extracting failure modes from Chinese short texts using machine learning and large language model", In 2025 3rd International Conference on Mathematics and Machine Learning (ICMML 2025), pp. 200–206.

Ferreira, C., Otto, K. N., Li, W., and Eisenbart, B. (2025), "Comparative study of LIME, SHAP and SHAP interaction quantification for improved FMEA development in energy-based maintenance", In 2025 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 385–389.

Goodfellow, I., Bengio, Y., and Courville, A. (2016), Deep learning. MIT Press, Cambridge, MA.

Goz E., Yuceer M., and Karadurmus E. (2019), "Total Organic Carbon Prediction with Artificial Intelligence Techniques", Computer Aided Chemical Engineering, Vol. 46, pp. 889-894.

Grigoryan, H., Quadrini, W., Polenghi, A., and Formentin, S. (2025), "Data-driven fault detection and risk assessment for machine tools using sensor data", In 2025 IEEE 64th Conference on Decision and Control (CDC), pp. 2584–2589.

International Electrotechnical Commission (2018), IEC 60812:2018 Failure Modes and Effects Analysis (FMEA and FMECA), 3rd ed., International Electrotechnical Commission, Geneva.

Janiesch, C., Zschech, P., and Heinrich, K. (2021), "Machine learning and deep learning", Electronic Markets, Vol. 31, pp. 685-695.

Kucejko, M., and Bugdol, M. (2025), "Recommendation system for DFMEA analysis using neural network embeddings", Management and Production Engineering Review, Vol. 16, No. 3, pp. 1–13.

Liang, D., Li, F., and Chen, X. (2024), "Failure mode and effect analysis by exploiting text mining and multi-view group consensus for the defect detection of electric vehicles in social media data", Annals of Operations Research, Vol. 340, pp. 289–324.

Liu, H.-C., Chen, X.-Q., Duan, C.-Y. and Wang, Y.-M. (2019), "Failure mode and effect analysis using multi-criteria decision-making methods: A systematic literature review", Computers & Industrial Engineering, Vol. 135, pp. 881-897.

Mokhtarzadeh, M., Rodríguez-Echeverría, J., Semanjski, I. and Gautama, S. (2025), "Hybrid intelligence failure analysis for Industry 4.0: a literature review and future prospective", Journal of Intelligent Manufacturing, Vol. 36, pp. 2309-2334.

Ojeda, J. C. O., de Moraes, J. G. B., Filho, C. V. d. S., Pereira, M. d. S., Pereira, J. V. d. Q., Dias, I. C. P., da Silva, E. C. M., Peixoto, M. G. M., and Gonçalves, M. C. (2025), "Application of a predictive model to reduce unplanned downtime in automotive industry production processes: A sustainability perspective", Sustainability, Vol. 17, pp. 3926.

Ojha, V. K., Goyal, S., Chand, M., and Kumar, A. (2026), "A novel data driven decision-making framework for preventive maintenance and fault analysis in modern production systems", Concurrent Engineering: Research and Applications, Vol 0, No. 0, pp. 1-15.

Oliveira, D. F., Brito, M. A., and Brandão, D. J. (2025), "FMEA based prescriptive model for equipment repair guidance", Frontiers in Artificial Intelligence, Vol. 8, p. 1630907.

Page, M.J., Moher, D., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P. and McKenzie, J.E. (2021), "PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews", BMJ, Vol. 372, n160.

Pereira, A. C., and Romero, F. (2017), "A review of the meanings and the implications of the industry 4.0 concept", Procedia Manufacturing, Vol. 13, pp. 1206-1214.

Rezaee, M. J., Yousefi, S., and Babaei, M. (2017), "Multi-stage cognitive map for failures assessment of production processes: An extension in structure and algorithm", Neurocomputing, Vol. 232, pp. 69–82.

Sader, S., Husti, I., Daroczi, M. (2022), "A review of quality 4.0: definitions, features, technologies, applications, and challenges", Total Quality Management & Business Excellence, Vol. 33, No 9-10, pp. 1164-1182.

Sarmast, Z., Shokouhyar, S., Ghanadpour, S. H., and Shokoohyar, S. (2023), "Unravelling the potential of social media data analysis to improve the warranty service operation", Industrial Management & Data Systems, Vol. 123, No. 5, pp. 1281–1309.

Shenoy, M., Sahasrabudhe, S., and Ameri, F. (2025), "Neurosymbolic AI-Driven Zero-Defect Manufacturing in Semiconductor Assembly: A Hybrid Framework", 36th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC).

Song, W., and Zheng, J. (2024), "A new approach to risk assessment in failure mode and effect analysis based on engineering textual data", Quality Engineering, Vol. 36 No. 4, pp. 805–823.

U.S. Department of Defense (1980), MIL-STD-1629A: Procedures for Performing a Failure Mode, Effects and Criticality Analysis, Department of Defense, Washington, DC.

Wiangkham, A., and Vongvit, R. (2026), "Integrating machine learning with FMEA for failure prioritization and process risk analysis in manufacturing", Results in Engineering, Vol., 29, p. 109128.

Wu, Z., Liu, W. and Nie, W. (2021), "Literature review and prospect of the development and application of FMEA in manufacturing industry", The International Journal of Advanced Manufacturing Technology, Vol. 112 No. 5-6, pp. 1409-1436.

Yılmaz, A. F., and Kose, O. (2025), "Advancing sustainable shipbuilding through a hybrid risk prioritization model: Integrating FMEA and machine learning", Engineering Science and Technology, an International Journal, Vol. 70, p. 102158.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Jishu-ken: The Autonomous Self-Study Group as the Structural Driver of Lasting Organizational Change

Eduardo L. García-Garcia

Factory Management Institute, NPO (FMI) | EMPWR
Corresponding author: eduardo.garcia@factorymanagementinstitute.com edu.lgg@gmail.com

***Abstract* — Jishu-ken (自主研) is one of the most consequential and most consistently misunderstood concepts in quality and operations management. Despite its unambiguous etymology — autonomous self-study — it has been systematically reduced in Western practice to an externally directed, instrumentally oriented learning activity. Even the most systematic academic treatments of the concept remain grounded in a model of directed learning that falls short of the autonomous self-directed inquiry that the concept's etymology and original practice encode. This paper is a conceptual study grounded in three complementary analytical methods: systematic literature analysis, primary source analysis of first-person testimonies from senior Toyota practitioners, and qualitative corpus analysis of the open-access TQM lecture series of Sensei Koichi Kimura (2022–2024). Building on this analysis, the recurring difficulties reported in sustaining Western Jishu-ken imitations are rooted in a fundamental epistemological error: treating a capability-building mechanism as a problem-solving tool. By drawing on Nonaka and Takeuchi's (1995) SECI model, Argyris and Schön's (1978) theory of organisational learning, and the pedagogical progression of Shu-Ha-Ri, the research demonstrates that Jishu-ken functions as the structural driver of the most enduring management innovations in industrial history — the Toyota Production System, Total Quality Management, and Hoshin Kanri. The paper concludes that lasting organisational change cannot be transferred from the outside; it must be grown from within through the specific structural mechanism of Jishu-ken.**

Keywords* — *Jishu-ken, autonomous study group, organisational knowledge creation, TQM, Toyota Production System, self-directed learning, continuous improvement, SECI model, Koichi Kimura, conceptual paper

I. INTRODUCTION

In the global diffusion of Japanese management methods over the past five decades, few concepts have been simultaneously so frequently cited and so poorly understood as Jishu-ken. The operations management and Lean manufacturing literature tends to present it primarily as a management-directed kaizen activity — a structured, time-bounded problem-solving exercise conducted by a cross-functional team of managers to achieve a defined operational improvement target. Even the most systematic academic treatment of Jishu-ken in English (Marksberry, Badurdeen and Gregor 2010) — which explicitly warns against this single-purpose misreading and defends a dual function of operational improvement and TPS understanding development — remains grounded in a model of externally directed, instrumentally oriented learning. The richer tradition of autonomous self-directed inquiry documented in Toyota practitioner testimonies and in Kimura's TQM corpus has received no systematic academic treatment.

However, a systematic engagement with the etymology of the concept, with primary testimonies from senior Toyota practitioners, and with the practitioner literature originating in the Japanese TQM tradition reveals a substantially different and richer understanding: one in which autonomous self-directed learning — not problem resolution — is the primary purpose, and in which the development of persons and organisational culture are the primary outcomes. This gap between the dominant operational interpretation and the original conceptual depth of Jishu-ken constitutes the research problem this paper addresses.

The practical stakes of this gap are significant. Organisations that implement Jishu-ken as a time-boxed improvement workshop achieve results for as long as the workshop runs. Organisations that understand and practice Jishu-ken in its original depth build something that cannot be easily dismantled: a culture of continuous self-directed learning, a shared language of improvement, and the organisational capability to navigate change from within. This distinction is, this paper argues, the key to understanding why the most successful and enduring corporate management systems in industrial history were not implemented by external consultants but generated internally over time, by the people who would inhabit them.

This paper addresses three explicit research questions:

RQ1: What is Jishu-ken — etymologically, historically, and in its most complete operational form?

RQ2: How does a Jishu-ken group operate in practice — its composition, structure, session dynamics, and temporal rhythm?

RQ3: Why is Jishu-ken, properly understood, a structural driver of lasting organisational change rather than a short-term improvement instrument?

II. THEORETICAL FRAMEWORK

This paper is grounded at the intersection of three theoretical traditions: organisational knowledge creation, organisational learning theory, and the Japanese pedagogical philosophy of mastery transmission. Each contributes a distinct analytical lens for interpreting Jishu-ken and its organisational function.

A. 2.1 Organisational Knowledge Creation: The SECI Model

The cornerstone of the paper's theoretical framework is Nonaka and Takeuchi's (1995) model of organisational knowledge creation. Their central argument — that sustained competitive advantage derives not from the management of explicit knowledge (procedures, manuals, technical specifications) but from an organisation's capacity to convert tacit knowledge into shared organisational understanding — provides the epistemological foundation within which Jishu-ken can be most accurately situated. Nonaka and Takeuchi developed this argument by studying precisely the kinds of Japanese organisational practices of which Jishu-ken is a primary example.

Their SECI model describes four modes of knowledge conversion: Socialisation (tacit to tacit, through shared experience and proximity), Externalisation (tacit to explicit, through dialogue and structured reflection), Combination (explicit to explicit, through synthesis and systematic organisation), and Internalisation (explicit to tacit, through embodied learning-by-doing). These modes form a continuous spiral operating across individual, group, and organisational levels. The complementary concept of 'ba' (場) — a shared context of creation, physical, virtual, or mental, in which knowledge conversion takes place — is introduced by Nonaka and Konno (1998) and is directly applicable to Jishu-ken as an intentionally constructed organisational ba.

As the analysis of this paper demonstrates, a functioning Jishu-ken group engages all four SECI modes in each cycle of collective inquiry, and over successive cycles the spiral deepens: each round of inquiry increases members' individual capacity to generate new understanding. This cumulative dynamic is the mechanism through which Jishu-ken produces organisational capability that is genuinely durable.

Empirical corroboration of this dynamic at the inter-organisational level is provided by Dyer and Nobeoka (2000), whose study of Toyota's supplier network demonstrates that knowledge diffuses more rapidly within Toyota's network than in competing automaker networks, and identifies Jishu-ken as one of six institutionalised knowledge-sharing routines through which this advantage is produced. Their finding that Toyota's network-level learning capability constitutes a source of competitive advantage that extends beyond firm boundaries directly supports the argument that Jishu-ken's function cannot be reduced to within-firm problem-solving: it is a mechanism of collective knowledge creation with network-wide organisational consequences.

B. 2.2 Organisational Learning: Single-Loop and Double-Loop

Argyris and Schön's (1978) distinction between single-loop and double-loop learning provides the second theoretical lens. Single-loop learning corrects errors within existing frameworks of assumption; double-loop learning questions the frameworks themselves. This distinction maps precisely onto the difference between the two Jishu-ken traditions identified in this paper: a problem-solving Jishu-ken group typically engages in single-loop learning — it corrects a process anomaly within the existing management system. A self-development Jishu-ken group, operating over time on self-selected themes that include the organisation's own management assumptions, systematically engages in double-loop learning. Senge's (1990) concept of the learning organisation — built on shared vision, team learning, mental model examination, systems thinking, and personal mastery — describes the organisational culture that a properly functioning Jishu-ken group is specifically designed to produce.

C. 2.3 Shu-Ha-Ri: The Transmission of Tacit Knowledge

The Japanese progression of Shu-Ha-Ri (守破離) describes three stages of mastery in traditional disciplines. Shu (守, follow the form) is the stage of disciplined imitation, in which the learner reproduces the master's method precisely, without deviation, until it is fully internalised. Ha (破, break from the form) is the stage of creative departure grounded in deep understanding: the practitioner begins to adapt and question the form in light of direct experience. Ri (離, transcend the form) is the stage at which the practitioner's understanding is fully independent: they can generate new forms from first principles. Kimura's TQM lecture series explicitly employs this progression in depicting the development of the Jishu-ken group across the fourteen lectures. Shu-Ha-Ri is therefore treated in this paper not as a cultural analogy but as the pedagogical architecture of Jishu-ken

D. 2.4 Dynamic Capabilities: Jishu-ken as Microfoundation

The fourth theoretical lens situates Jishu-ken within the dynamic capabilities framework introduced by Teece, Pisano and Shuen (1997) and elaborated by Teece (2007). Dynamic capabilities are defined by Teece et al. (1997, p. 516) as 'the firm's ability to integrate, build, and reconfigure internal and external competences to address rapidly changing environments.' Their original framework specifies three dimensions through which dynamic capabilities are built and exercised: processes (the routines through which the firm coordinates and combines resources), positions (the firm-specific assets, including difficult-to-imitate knowledge, that the firm has accumulated), and paths (the evolutionary trajectory and path dependencies that constrain and enable the firm's future options).

A functioning Jishu-ken group operates at the intersection of all three dimensions. As a process, it constitutes a recurrent organisational routine for collective knowledge integration through self-directed inquiry. As a position, it

accumulates tacit knowledge assets — shared interpretive frames, rigorous habits of inquiry, mutual analytical trust — that are constitutively non-transferable and therefore sources of imperfectly imitable competitive advantage. As a path, it produces a cumulative developmental trajectory in which each cycle of inquiry expands the group's collective capacity for the next, generating the kind of irreversible organisational learning that Teece et al. (1997) identify as the source of sustained competitive advantage in environments of rapid change.

The connection becomes more explicit in Teece's (2007) elaboration of the microfoundations of dynamic capabilities, in which the high-order capacities of sensing opportunities and threats, seizing through resource mobilisation, and transforming through ongoing organisational renewal are decomposed into the underlying skills, processes, procedures, organisational structures, decision rules, and disciplines that produce them. Jishu-ken is, on this account, an observable instance of all three. The disciplined gemba observation that opens each cycle of inquiry is a sensing microfoundation: a routine through which the organisation continuously detects anomalies, weak signals, and emerging opportunities at the operational front. The collective deliberation through which the group selects its theme and designs its countermeasures is a seizing microfoundation: a routine through which the organisation mobilises distributed knowledge to capture value from sensed opportunities. The self-generated commitments through which each member integrates session learning back into individual practice are a transforming microfoundation: a routine through which the organisation continuously renews its operational and managerial competences.

This framing yields three implications relevant to the paper's argument. First, it situates Jishu-ken at the core of the strategy literature on competitive advantage, not at its periphery as a Japanese cultural artefact. Second, it identifies what is missing from the dynamic capabilities literature itself: while Teece (2007) names the microfoundations, the literature has paid relatively little attention to the specific organisational routines through which microfoundations are built over time. Jishu-ken is one such routine, with documented forty-five-year longitudinal evidence of its capability-building trajectory. Third, it strengthens the argument that capability transfer through external consultancy is structurally limited: the dynamic capabilities literature has long established that the most valuable capabilities are precisely those that cannot be purchased on the market because they emerge from firm-specific processes, positions, and paths. The Toyota Supplier Support Center evidence discussed in Section 8.1 is, on this reading, an empirical confirmation of a theoretical prediction Teece et al. (1997) made two decades earlier: that explicit knowledge transfer cannot substitute for the path-dependent, firm-specific microfoundations through which durable capabilities are constructed.

III. METHODOLOGY

A. 3.1 Research Design: The Conceptual Paper

This paper is a conceptual study in the tradition established by Gilson and Goldberg (2015): a form of scholarly contribution that addresses gaps in the literature that empirical studies cannot fill — specifically, gaps of conceptual clarity, theoretical integration, and definitional precision. The need for this form of contribution is directly evidenced in the Jishu-ken literature: Marksberry et al. (2010, p. 684) conclude their empirical study by noting that 'most attempts to imitate Toyota fail because techniques are adopted piecemeal with little understanding of why they exist or what kind of organisational culture is needed to keep them alive.' This paper responds to that diagnostic directly. Its purpose is not to generate new empirical data but to reconstruct, from multiple independent bodies of evidence, the full conceptual depth of Jishu-ken and to explain its organisational function theoretically.

The research design employs three complementary analytical methods — systematic literature analysis, primary source analysis, and qualitative corpus analysis — whose findings are integrated through cross-source triangulation (Denzin 1978). The logic of triangulation is central to the paper's methodological rigour: where independent sources converge in their description of Jishu-ken, the finding is treated as robustly supported; where they diverge, the divergence is treated as analytically significant.

B. 3.2 Method 1: Systematic Literature Analysis

A systematic search of the academic and practitioner literature on Jishu-ken was conducted across databases including Emerald Insight, ScienceDirect, and Google Scholar, supplemented by targeted review of practitioner publications from established quality and Lean organisations (Planet Lean, Shinka Management, Gemba Academy). Search terms included 'Jishu-ken,' 'Jishuken,' '自主研,' 'Toyota autonomous study group,' and 'Toyota manager kaizen.' Sources were included if they met at least one of three criteria: (a) they provided a direct definitional or operational description of Jishu-ken; (b) they provided first-person accounts of Jishu-ken participation by practitioners within Toyota or its affiliated organisations; or (c) they provided theoretical frameworks directly applicable to the analysis of Jishu-ken as an organisational learning mechanism.

C. 3.3 Method 2: Primary Source Analysis — Toyota Practitioner Testimonies

A distinctive feature of the Jishu-ken literature is the availability of direct first-person testimonies from senior Toyota practitioners who participated in Jishu-ken over extended periods. Two such testimonies are used as primary sources in this paper. The first is that of Hideshi Yokoi, former Plant Director at Toyota Motor Kyushu and former President of the Toyota Production System Support Center (TSSC) in North America, recorded in interview format by Planet Lean (2019). The second is that of Akinori Hyodo, a practitioner who spent over a decade in Toyota's internal Jishu-ken programme and documented his experience for Shinka Management (2024). Both sources were selected for their insider perspective, their seniority within the Toyota organisational structure, and the specificity and internal consistency of their accounts.

D. 3.4 Method 3: Qualitative Corpus Analysis — Kimura's TQM Lecture Series

The primary analytical source for this paper is the TQM lecture series of Sensei Koichi Kimura, comprising fourteen lectures (TQM-1 through TQM-14) published as open-access documents by the Factory Management Institute on Internet Archive under a Creative Commons CC-BY-NC-ND 4.0 licence. Sensei Kimura accumulated forty-five years of production gemba experience at Sumitomo Electric Industries, working alongside the architects of TPS, TQM, and TPM at Toyota, Honda, and Denso. The corpus was analysed using inductive qualitative thematic analysis (Braun and Clarke 2006) across three analytical dimensions: (1) Kimura's explicit articulation of Jishu-ken's purpose and philosophy; (2) the structural and operational features of the Jishu-ken group; and (3) the longitudinal development of the group and its individual members. Particular analytical attention was given to TQM-8 and subsequent lectures, in which the study group is named explicitly as Jishu-ken.

E. 3.5 Reflexivity and Triangulation

The author's role as editor of the Kimura TQM lecture series — responsible for structural editing and English-language refinement while preserving the author's authentic voice and pedagogical intent — provided privileged analytical access to the corpus. This proximity is acknowledged as both an epistemological asset and a potential source of confirmation bias. To mitigate this risk, the analysis follows an explicit triangulation protocol with three operating rules. First, no claim about Jishu-ken's purpose, structure, or organisational function is advanced on the authority of the Kimura corpus alone: each substantive proposition presented in Sections 5 to 8 is required to find independent corroboration in at least one of the two external Toyota practitioner testimonies (Yokoi 2019) (Hyodo 2024) in the peer-reviewed academic literature (Marksberry, Badurdeen and Gregor 2010) (Liker 2004) (Dyer and Nobeoka 2000). Second where three sources converge, the finding is treated as robustly supported and presented declaratively. Third, where the sources diverge — as they do, for instance, on the question of compulsory versus voluntary participation, on the role of external versus self-selected themes, and on the intensity of review culture (Imaizumi in Liker, 2004 and INEAK, 2009 versus Yokoi, 2019 and Hyodo, 2024) — the divergence is preserved as analytically significant rather than resolved by appeal to any single source. It is precisely from these documented divergences that the central distinction between the operational and self-development traditions emerges (Section 5), and from which the scaffolded-autonomy reading of the historical record follows (Section 4.1). The triangulation protocol does not eliminate interpretive proximity but renders the interpretation accountable to independent evidence at every substantive claim.

IV. ETYMOLOGY AND ORIGINS: THE NAME AS PROGRAMME

Jishu-ken derives from the phrase kanban houshiki bukachou jishu kenkyuukai (かんばん方式部課長自主研究会): 'kanban system department and section manager autonomous study groups' (Miller, 2021; Baudin, 2012). Abbreviated to Jishu-ken, its three constitutive kanji are unambiguous: ji (自, myself), shu (主, self-motivated and autonomous), and ken (研, to study and research). The most accurate English rendering is therefore 'autonomous self-study.' Autonomy, self-direction, and the commitment to learning are not adjectives applied to Jishu-ken from outside — they are its substance. The purpose is encoded in the name.

The concept emerged in the early development of the Toyota Production System. Taiichi Ohno required department and section managers from the Motomachi and Kamigo plants to gather on the factory floor, conduct hands-on improvement activity, select their own study themes, and work collectively on process problems (Miller 2021). The formal designation 'Jishu-ken' appeared in 1977, when Toyota's Operations Management Consulting Division (OMCD) — the elite corps of TPS experts founded by Ohno in the mid-1960s — structured these activities into a supplier network development programme. Toyota's fifty-five to sixty key suppliers were organised into geographically proximate groups of four to seven companies, rotating through three- to four-month improvement projects at each company in turn (Liker 2004) (INEAK 2009).

Baudin (2012) documents a critical evolutionary dynamic within the first decade: Jishu-ken groups progressed from pure study, to proposing changes, to implementing them — becoming, in Baudin's description, 'an umbrella for improvement activities at all levels in the hierarchy, using a variety of formats.' The persistence of the original name — and the spirit of autonomous self-directed inquiry it carries — across this functional evolution is not merely lexical. It signals that those who developed Jishu-ken understood its learning orientation as constitutive, not incidental.

A. 4.1 Scaffolded Autonomy: The Apparent Paradox of Ohno's Method

A reader of the historical record encounters what initially appears as a fundamental contradiction. The accounts of Ohno's management style (Liker 2004) (Ohno 1988) describe an exacting, demanding, and at times confrontational figure who imposed on his managers the obligation to stand at the gemba, to observe until they could see, and to ask why five times before claiming to understand a problem. Imaizumi's account of the early Japanese supplier Jishu-ken (Liker 2004) (INEAK 2009) similarly describes a practice that was compulsory rather than voluntary, with themes assigned by Toyota's experts rather than chosen by participants, and review sessions whose intensity was psychologically demanding. How can a practice whose name encodes autonomous self-direction emerge from a tradition characterised, at its origin, by such pronounced external demand?

The contradiction is only apparent. What Ohno imposed was not the content of inquiry but the obligation to inquire. His exigence operated at the level of disciplined attention: he required his managers to study, to observe at the gemba, to refuse premature conclusions, and to remain with a problem until they had genuinely understood it. He did not, in the documented record, prescribe their hypotheses, dictate their countermeasures, or instruct them on what they should conclude. The method he transmitted was the discipline of inquiry itself; the substance of inquiry was theirs to develop.

This pattern maps with precision onto the Shu stage of the Shu-Ha-Ri progression introduced in Section 2.3. The Shu stage is defined by disciplined adherence to an external form — a form transmitted by a sensei who has earned the right to demand it. Autonomy is not abandoned in Shu; it is being constructed. The learner who has not yet been disciplined by the form has no foundation from which genuine autonomous inquiry could later emerge. What appears externally as imposition is, internally, the construction of the cognitive and behavioural infrastructure within which later autonomy becomes possible.

The Jishu-ken tradition therefore rests on what may be termed scaffolded autonomy: a developmental architecture in which the discipline imposed at the Shu stage builds the capacity that the Ha and Ri stages express as autonomous inquiry. Ohno's demanding presence is not in tension with the autonomy of his managers; it is the precondition through which their autonomy is grown. The same logic applies to the early supplier Jishu-ken's compulsory and assigned character: the practice was operating at the Shu stage, scaffolding a capacity that, in its mature self-development form (documented in Yokoi (2019), Hyodo (2024), and Kimura (2022–2024)), expresses itself as genuine autonomous self-direction. The progressive moderation of the supplier programme over time — toward voluntary participation, autonomous theme selection, and supportive review — is not a softening of the original method but its developmental progression from Shu toward Ha.

Recognising this scaffolded structure resolves the apparent paradox and yields a substantive implication for practice: organisations cannot leap directly into autonomous Jishu-ken without traversing the disciplined Shu stage that earns the autonomy. The mature self-development tradition is not the starting point but the destination of a developmental trajectory in which discipline precedes autonomy and external form precedes internal generation.

V. TWO TRADITIONS UNDER ONE NAME

A. 5.1 The Operational Tradition

Marksberry et al. (2010) offer the most systematic academic treatment of Jishu-ken in the English-language literature, and their position is more nuanced than a purely operational reading would suggest. They explicitly warn that Jishu-ken is frequently misunderstood as a single-purpose plant improvement activity, and argue that its correct function is dual: resolving operational problems at the gemba while simultaneously deepening managers' understanding of the Toyota Production System. Their study documents an eight-step Jishu-ken process and analyses its function as a management development mechanism within the TPS framework, characterising it as a 'management-directed kaizen activity' (Marksberry, Badurdeen and Gregor 2010) structured around theme selection, gemba analysis, countermeasure implementation, standardisation, and lateral knowledge transfer through Yokoten.

This acknowledgement of a developmental dimension is significant. However, the development Marksberry describes remains externally structured and instrumentally oriented: managers deepen their TPS understanding in order to execute it more effectively, under the direction of OMCD's TPS experts. The self-development tradition documented by Yokoi (2019), Hyodo (2024), and Kimura (2022–2024) goes further in a qualitatively different direction: the group selects its own themes, sets its own pace, and treats the development of persons and organisational culture as an end in itself — not as a means to more effective TPS implementation. The distinction is not between problem-solving and learning, but between directed learning and autonomous inquiry.

A further nuance within the operational tradition is documented by Imaizumi, executive at Araco Corporation, one of Toyota's most sophisticated Japanese suppliers. Imaizumi describes the Japanese supplier Jishu-ken as fundamentally different in character from its American counterpart: participation was compulsory, themes were assigned by Toyota's TPS experts rather than chosen by the group, and the review culture was described as severe to the point of causing psychological distress among participants (Liker 2004) (INEAK 2009). This testimony is consistent with a substantial body of evidence on the human costs of lean production more broadly. Studies have documented that lean implementations can intensify work pace, narrow decision latitude, and elevate psychological strain (Landsbergis, Cahill and Schnall 1999) (Conti, et al. 2006), with effects mediated by configurational choices in job design rather than by lean per se (de Treville and Antonakis 2006). Within the Jishu-ken practice itself, intensity is structural: the demand to study, observe, and remain with unresolved complexity exacts cognitive and emotional cost. The self-development tradition does not eliminate this intensity but reframes it: pressure becomes the disciplined demand of a learning practice the participant has chosen, rather than the punitive review of a deliverable assigned from above. Toyota has gradually moderated the supplier programme over time, particularly in its North American activities, in a direction consistent with this reframing. The progression is significant: it suggests that even within the operational tradition, the original Japanese Jishu-ken operated in tension with the principle of autonomy encoded in its name — a tension that the self-development tradition, as documented by Yokoi (2019) and Kimura (2022–2024), explicitly resolves by placing autonomous voluntary inquiry at the centre.

B. 5.2 The Self-Development Tradition

The systematic literature analysis and primary source review reveal a second tradition — less visible in the English-language academic literature but consistently present in the accounts of senior Toyota practitioners and in Kimura's TQM corpus. This tradition understands Jishu-ken not as a problem-solving mechanism but as a vehicle of personal and organisational development, in which the operational problem is the medium of learning, not the destination.

Yokoi (2019) draws the distinction precisely: in kaizen events, the target is assigned by management and the objective is its achievement; in Jishu-ken, the team selects its own target and the objective is 'reflection and learning.' Yokoi emphasises the group's ambition to produce 'important and sustainable changes to the production system' — changes that last because they were built by people who understand them, not installed by people who were told to

deliver them. Hyodo (2024) is more direct: the formally correct name within Toyota is 'Toyota Production System Jishu-Kenkyu-Kai (literally: Self-directed TPS Study Group)' and its main objective is 'personnel development.'

In Kimura's teaching, this orientation is most fully articulated. The growth of the individual manager — their capacity for rigorous observation, disciplined reasoning, honest self-assessment, and open dialogue — is treated as equivalent in importance to any specific operational outcome. The group's progress is measured not by its problem-solving outputs alone but by the quality of understanding its members develop and carry permanently into their practice.

C. 5.3 Why the Distinction Has Organisational Consequences

The difference between these two traditions is not definitional; it is consequential. A Jishu-ken group operating under the operational tradition produces improvements contingent on the group's existence: when the project closes, the improvement must be sustained through standardisation and audit. A Jishu-ken group operating under the self-development tradition produces a categorically different output: managers who have genuinely learned, who carry that learning as tacit knowledge, and who can generate new understanding independently of any external framework or facilitator. The first output is a solution. The second is a capability. Solutions can be lost — as every organisation that has survived a consultant's departure and watched their system erode knows. Genuine capability, once built, is persistent.

VI. HOW JISHU-KEN WORKS: COMPOSITION, STRUCTURE, AND SESSION ANATOMY

A. 6.1 Group Composition

A Jishu-ken group is composed of managers and supervisors — neither frontline workers, who participate in Quality Control Circles at the operational level, nor senior executives, who set direction and protect the group's space from operational interference. Academic literature describes typical groups of four to six members (Marksberry, Badurdeen and Gregor 2010); in the Kimura tradition, cross-functional composition across production, quality, engineering, logistics, and administrative management is constitutive rather than merely recommended, since the challenges of the gemba are rarely single-function problems. Membership is stable across time, and continuity is the mechanism through which trust, shared language, and collective reasoning capacity accumulate cycle by cycle.

B. 6.2 Theme Selection: The Self-Directed Imperative

The autonomous selection of the study theme is one of the most critical — and most consistently misunderstood — features of Jishu-ken. Yokoi (2019) is explicit: the team selects the target, and managers challenge them on it. Theme selection rests on the group's direct observation of the gemba, its assessment of where organisational capability is weakest or development potential greatest, and its collective judgment about what most needs to be understood.

C. 6.3 Anatomy of a Session

The operating logic of a Jishu-ken session differs substantially from both a conventional management meeting and a structured kaizen event. Between sessions, one or more members take individual responsibility for studying a specific topic in depth — a commitment that is self-generated from the group's collective deliberation, not assigned by a facilitator. The presenting member arrives having genuinely studied the topic and is expected to bring not only what they have understood but also what they have not yet resolved.

The session itself is structured around open, non-hierarchical dialogue. The presentation is not a report but an invitation to collective inquiry: members question interpretations, challenge assumptions, propose alternative framings, and reason toward understanding together. In Kimura's depiction, the quality of a session is measured not by the elegance of its conclusions but by the honesty of its questions and the depth to which the group is willing to remain with unresolved complexity.

Sessions close not with a summary presentation but with a self-generated distribution of individual commitments. Each member identifies, from their own understanding of the session's dialogue, what they personally need to study, observe, test, or develop before the next session. These commitments are not assigned from above; they emerge from the collective reasoning of the group and are held by each member as a personal accountability to shared learning.

D. 6.4 Frequency and Temporal Rhythm

Jishu-ken does not operate on a fixed timetable. Its frequency is responsive to the urgency and complexity of the group's current inquiry. Under conditions of steady study, sessions are typically weekly or fortnightly: frequent enough to maintain continuity and momentum, without overwhelming members' operational responsibilities. When the group is confronting an urgent challenge, the rhythm accelerates and sessions may occur several times a week.

In the most critical cases — when a problem of genuine operational significance is on the table — the session does not have a fixed closing time: it continues until a satisfactory understanding of the problem has been reached and a solution pathway designed. This absence of an artificial time boundary distinguishes Jishu-ken most sharply from the time-boxed kaizen events with which it is most frequently confused.

E. 6.5 The Role of the Sensei

The Jishu-ken group is not leaderless, but its leadership operates differently from a conventional project manager or workshop facilitator. The sensei — whether internal or external to the organisation — does not direct the group's inquiry or supply its answers. The sensei creates the conditions in which genuine inquiry becomes possible: by posing questions that raise the quality of the group's thinking, by challenging assumptions without resolving them prematurely, and by modelling the intellectual humility that self-directed learning requires. Liker (2004) observes that this mode of

teaching — asking rather than telling — is one of the most consistently misunderstood features of Toyota's management development approach.

VII. JISHU-KEN IN CONTEXT: A COMPARATIVE ANALYSIS

Table 1 systematises the key distinctions between Jishu-ken in its fullest form and the related improvement practices with which it is most frequently conflated in Western operations management literature.

Fig. 1. **Table 1.** *Comparative analysis of Jishu-ken and related improvement practices.*

Dimension	Quality Circle / Kaizen Event	Jishu-ken (self-development tradition)
Primary objective	Solve a defined problem	Develop people through self-directed inquiry
Theme selection	Assigned by management	Chosen autonomously by the group
Duration	Time-boxed (hours to weeks)	Open-ended; recurrent over months or years
Session end	Fixed by agenda or time slot	When solution designed; no fixed end if problem is live
Participants	Frontline workers or technical specialists	Managers and supervisors, cross-functional
Post-session actions	Assigned by facilitator	Self-generated by each member from group dialogue
Group lifecycle	Disbanded on completion	Continuous; group persists and grows
Primary value	Process improvement; short-term ROI	Cultural change; cumulative organisational capability

[a] *Source: Author's synthesis from Yokoi (2019), Hyodo (2024), Marksberry et al. (2010), and Kimura (2022–2024).*

The systematic degradation of Jishu-ken into a management-directed, time-bound problem-solving event represents a fundamental misunderstanding of the Toyota Production System's pedagogical core. While the Operational Tradition provides short-term gains in efficiency, it fails to generate the internal capacity required for a true Learning Organisation. The 'Self-Development Tradition' is not an optional cultural nuance of Japanese management; it is the *conditio sine qua non* for organisational resilience. Western organisations must confront what may be termed the Paradox of Control: the more management attempts to direct the Jishu-ken process toward pre-defined KPIs, the more they stifle the very socialisation and internalisation processes that build deep capability.

VIII. WHY JISHU-KEN DRIVES LASTING ORGANISATIONAL CHANGE

A. 8.1 Generated, Not Installed: The Historical Argument

The central proposition of this paper is that Jishu-ken, in its original form, is the generative mechanism through which Japan's most enduring management innovations were produced. The Toyota Production System was not designed by an engineering team and implemented across the factory. It was developed iteratively, over decades, through sustained collective inquiry on the production floor by the managers who inhabited it. Ohno's method was to pose difficult problems to his managers, require them to observe at the gemba, and demand they study and report back — repeatedly, across years (Liker 2004). The system that emerged expressed their understanding of their organisation's operational reality; they were capable of maintaining, developing, and teaching it independently because they had developed it themselves.

The same generative dynamic is evident in the development of TQM at Sumitomo Electric and in the emergence of Hoshin Kanri in Japan's leading corporations. Kimura's lecture series documents how the quality management practices that defined Sumitomo's operational excellence were developed through sustained collective study within the organisation — not imported from a consulting firm's proprietary framework (K. Kimura 2026). In each case, the management system and the learning process that generated it were inseparable. Transferring one without the other — a pattern documented repeatedly in the literature on Western Lean adoptions (Marksberry, Badurdeen and Gregor 2010) (Liker 2004) — produces, at best, a system without the organisational immune system that sustains it.

The most direct empirical evidence for this proposition comes from Hajime Ohba, former OMCD member and director of the Toyota Supplier Support Center (TSSC) in North America. Ohba's results were, by any conventional metric, exceptional: across 31 completed projects, TSSC reduced inventory by an average of 75% and improved productivity by an average of 124% (Liker 2004) (INEAK 2009). Yet companies that achieved dramatic improvements on TSSC-supported lines consistently fell back to lower TPS levels once TSSC visits reduced. Ohba's own explanation was direct: while TPS experts could force installation of TPS principles with extraordinary results on selected lines, they could not inoculate the suppliers with Toyota Way genes (Liker 2004) (INEAK 2009). The failure was not shop-floor resistance — it was the absence of committed top management. What TSSC transferred was explicit knowledge without the tacit knowledge that keeps those procedures alive when the expert is no longer in the room.

B. 8.2 The SECI Explanation: Why Tacit Knowledge Is the Durable Kind

Nonaka and Takeuchi's (1995) framework provides the theoretical explanation for why Jishu-ken produces durable capability. Their core argument is that organisational knowledge that is genuinely difficult to imitate — and therefore a source of sustained competitive advantage — is tacit: embedded in the judgements, intuitions, and practices of individuals and teams, invisible to competitors and resistant to reverse engineering. Tacit knowledge cannot be transferred through training programmes or implementation guides. It can only be grown through sustained experiential learning in a structured shared context — a ba.

A functioning Jishu-ken group is precisely such a context. Each cycle of the group's inquiry engages all four SECI modes in sequence: members share tacit knowledge through their gemba observations (Socialisation); they articulate their understanding through presentation and challenge (Externalisation); they synthesise their emerging insights with existing organisational knowledge (Combination); and they carry the group's conclusions back into their individual tacit knowledge through practice at the gemba (Internalisation). Dyer and Nobeoka (2000) provide network-level empirical evidence consistent with this proposition: suppliers that participate actively in Toyota's knowledge-sharing routines demonstrate measurably faster learning rates than those in competing networks.

C. 8.3 Longitudinal Evidence from Kimura's Corpus

The most direct evidence for Jishu-ken as a driver of cumulative organisational capability is the longitudinal trajectory documented in Kimura's TQM series. The group depicted across fourteen lectures does not solve a problem and dissolve. It encounters successive domains of increasing complexity — waste analysis, quality tools, the KJ method, creative problem-solving, the Taguchi method, the conceptual architecture of TPM and TQM or TQC — and with each encounter, its members develop. Their questions become more precise. Their capacity to challenge each other becomes more respectful and more analytically rigorous.

> *The group that engages with waste analysis in TQM-1 is not the same group that confronts the Taguchi method in TQM-13. It is not the same group in terms of shared language, mutual trust, analytical rigour, or capacity to navigate genuine novelty. This cumulative development — this growing capability — is what a genuinely functioning Jishu-ken produces. And it is precisely what no time-boxed improvement event, however well-facilitated, can replicate.*

A further analytical dimension is provided by the naming evolution within the series. The group is initially presented under functional labels — 'cost reduction team,' 'cross-functional study group' — and is named explicitly as Jishu-ken only from TQM-8 onwards. This progression is itself a pedagogical statement: the form of the group — its spirit, discipline, and commitment to autonomous learning — precedes and earns the name.

IX. EVIDENCE-BASED IMPLEMENTATION PRINCIPLES

A. 8.1 Transitioning From Management-Driven to Autonomous Inquiry

A practitioner reading the preceding sections may legitimately ask how an organisation that currently operates under a traditional management-driven improvement system — kaizen events called by senior leadership, themes assigned from above, time-boxed deliverables, audit-based sustainment — can move toward the autonomous self-development tradition the paper has described. Treating this transition as a single act of cultural conversion misunderstands what is being constructed. The transition follows the same Shu-Ha-Ri progression that operates within a Jishu-ken group itself, only now operating at the level of the organisation rather than the individual.

In the Shu phase of organisational transition, the practice begins under conditions that resemble the early Japanese supplier Jishu-ken documented by Imaizumi (Liker 2004) (INEAK 2009): senior leadership convenes a fixed group of managers, the discipline of preparation and rigorous inquiry is externally enforced, and themes may be initially scoped by leadership to ensure operational relevance. What distinguishes this from a conventional kaizen event is what is being demanded: not the achievement of a target, but the practice of disciplined collective inquiry. Participants are expected to study, to challenge each other's reasoning, to remain with unresolved complexity, and to generate their own commitments. The form is externally imposed; the substance is internally generated. Leadership signals — through its presence, its questions, and its tolerance of unresolved sessions — that what is being built is the capacity to inquire, not the capacity to deliver.

In the Ha phase, the group begins to depart from the imposed form. Themes increasingly emerge from the group's own gemba observation rather than from leadership scoping. Session rhythm adapts to the urgency and complexity of the inquiry rather than to a corporate calendar. The group develops the analytical and interpersonal habits — productive disagreement, comfort with ambiguity, mutual challenge — that the operational tradition's deliverable orientation tends to suppress. At this stage the role of leadership shifts from convener to protector: the group's autonomy must be defended against operational pressure to convert the practice back into a problem-solving event with a deliverable.

In the Ri phase, the group has internalised the discipline of autonomous inquiry to the point where the external scaffolding can be progressively withdrawn. Theme selection, session design, and post-session commitments are entirely self-generated. Leadership's function becomes ensuring institutional protection — that the group's time, membership, and authority to act on its findings remain intact under operational pressure — rather than directing or shaping the inquiry itself.

Three implications follow for organisations attempting this transition. First, the transition is sequential, not simultaneous: organisations cannot leap from management-driven kaizen directly to autonomous self-development without traversing the disciplined Shu phase in which the form of inquiry is externally enforced before its substance is internally generated. Second, the transition requires sustained leadership commitment over years, not quarters: the developmental progression from Shu through Ha to Ri operates on the time horizon of capability building, not project delivery. Third, the most common failure mode is premature closure: organisations that interpret the early Shu-phase sessions as inefficient kaizen events and reimpose deliverables-orientation collapse the developmental trajectory and produce neither operational improvement nor capability accumulation. The autonomy that defines mature Jishu-ken is not announced; it is earned through a developmental sequence the organisation must commit to traverse.

B. 8.2 Five Principles for Mature Practice

Nonaka and Takeuchi's (1995) framework provides the theoretical explanation for why Jishu-ken produces durable

The following five implementation principles are derived directly from the three analytical methods applied in this paper and are offered as evidence-based guidance for practitioners seeking to implement Jishu-ken in its fullest form.

The first principle is the necessity of a baseline. Hyodo (2024) is explicit: Jishu-ken requires a gemba where members already understand their current situation and can observe its challenges clearly. Launching Jishu-ken without this foundation is, in Hyodo's words, a formidable undertaking — not because the group cannot study, but because it has no operational reality against which to anchor its inquiry.

The second principle is genuine autonomy in theme selection. Senior leadership must create the space for the group to choose its own study theme and resist the impulse to direct that choice toward current operational priorities. This autonomy in theme selection is inseparable from a complementary form of organisational empowerment: the group must also hold the authority to act on what it discovers — to implement, test, and standardise without requiring permission at each step.

The third principle is stability of membership and protection of the group's time. The collective intelligence that Jishu-ken builds is an interpersonal achievement: it lives between specific people who have studied together over time. Frequent membership turnover, or the routine cancellation of sessions for operational reasons, prevents accumulation and collapses the group into the operational tradition's problem-solving format.

The fourth principle is a sensei capable of distinguishing between directing and guiding. The sensei's function is to raise the quality of the group's questions, not to supply their answers. This requires both deep operational knowledge of the gemba and specific personal humility before the complexity of organisational reality.

The fifth principle — the most important and the most frequently neglected — is patience. The results of Jishu-ken in its full form are not visible in early sessions or early months. What is being built is not a process improvement but a learning culture. That culture reveals itself not when the group solves a problem — any group can solve a problem — but when it confronts a genuinely novel challenge and responds with confident, grounded, collective inquiry rather than dependency on external expertise.

X. LIMITATIONS AND DISCUSSION

A. 10.1 Limitations

Three limitations of this study require explicit acknowledgement. First, as a conceptual paper, this study does not provide new empirical data on the outcomes of Jishu-ken implementation in specific organisational contexts. The propositions advanced await systematic empirical testing — specifically, longitudinal comparative case studies measuring the development of organisational capability in groups implementing Jishu-ken in its self-development form against groups implementing it in its operational form.

Second, the primary corpus — Kimura's TQM lecture series — is a practitioner source rather than a peer-reviewed research publication. The author's dual role as editor and analyst of this corpus creates interpretive proximity that may have shaped findings. This risk is addressed through triangulation with the independent Toyota practitioner testimonies of Yokoi (2019) and Hyodo (2024). Convergence across all three independent bodies of evidence strengthens confidence in the findings.

Third, the analysis is grounded in the Japanese organisational context in which Jishu-ken developed, and two transferability questions remain open. The first is cultural: how the practice transfers to organisations outside Japan, and what adaptations are required for effective implementation. The literature on the cross-cultural transfer of Japanese management practices documents that techniques travel more readily than the cultural and institutional substrate that sustains them (Recht and Wilderom 1998), with implications for which elements of Jishu-ken — its inquiry structure, theme selection, sensei role — are likely to transfer with relatively little adaptation, and which — its rhythm, intensity, tolerance for unresolved sessions — may require substantive reconfiguration. The second is organisational: the multinational matrix structures characteristic of contemporary global enterprises (Galbraith 2009)(Galbraith, 2009) place managers in functional reporting lines that converge only at corporate headquarters, not at the local site. Whereas the Toyota plants in which Jishu-ken developed shared a unified local hierarchy that made cross-functional study natural, today's site-level managers belong to separate global towers — operations, quality, engineering, finance — each with its own policies, KPIs, and authority structure. The cross-functional Jishu-ken composition described in Section 6.1 is, in such organisations, structurally illegitimate at the local level: managers cannot autonomously authorise themselves to study together across reporting lines. This implies that, beyond cultural adaptation, Jishu-ken in modern multinationals requires explicit corporate-level authorisation as a global programme that legitimises local cross-functional autonomous study. Both transferability questions — cultural and organisational-structural — represent priority directions for empirical research.

B. 10.2 Discussion

The analysis has implications at three levels. Theoretically, the paper contributes to the organisational learning and knowledge management literature by proposing Jishu-ken as a concrete, historically evidenced instance of a sustained organisational ba in which the full SECI spiral of knowledge creation operates across multiple cycles, producing cumulative rather than episodic organisational capability.

Methodologically, the paper demonstrates the research value of open-access practitioner corpora — specifically the Kimura TQM series on Internet Archive — as a source for qualitative analysis of management practices. The corpus provides longitudinal observational depth that is rarely available to academic researchers studying Japanese management, and its open-access format makes it fully available for replication and extension by other scholars.

Practically, the paper offers a clear and evidence-based answer to a challenge of enduring relevance: why do management system implementations so often fail to produce lasting change? The answer is that explicit knowledge can be transferred, but tacit knowledge must be grown. Jishu-ken is the structural mechanism through which organisations grow the tacit knowledge that makes any management system genuinely their own — and therefore genuinely sustainable. This argument applies with particular force to the governance challenges posed by AI-assisted and autonomous industrial operations, which require precisely the kind of deep, self-directed collective inquiry that Jishu-ken cultivates. That application is addressed in a companion paper currently under preparation.

XI. CONCLUSION

This paper has addressed three research questions. RQ1 establishes that Jishu-ken, in its complete form, is an autonomous self-study group of managers and supervisors whose primary purpose is the development of persons and organisational culture, not the resolution of defined operational problems. RQ2 documents its anatomy: stable cross-functional composition, autonomous theme selection, an adaptive rhythm responsive to urgency, organisational empowerment to act on its own findings, and a sensei whose function is to guide rather than direct. RQ3 advances the central proposition, grounded in the SECI model and corroborated by three independent bodies of evidence, that Jishu-ken is the structural driver of lasting organisational change because it is the mechanism through which tacit organisational knowledge is grown, not transferred. The historical record of TPS, TQM, and Hoshin Kanri is consistent with this proposition: these systems endured because they were generated from within, through the kind of autonomous collective inquiry that Jishu-ken represents.

The change that lasts is the change an organisation learns its way into, together. Jishu-ken is the structural form through which that learning takes place, and its availability to any organisation willing to practice it in its original depth — with patience, genuine autonomy, and the guidance of someone who has earned the role of sensei — constitutes its most important and most under-appreciated feature.

XII. . ACKNOWLEDGEMENTS

The author expresses deep gratitude to Sensei Koichi Kimura, whose TQM lecture series constitutes the primary intellectual source of this work. The generosity of publishing forty-five years of gemba learning as open-access documents reflects the spirit of autonomous collective inquiry that animates the Jishu-ken tradition itself.

XIII. . REFERENCES

Akinori Hyodo. *Toyota Jishuken.* 2024. https://shinkamanagement.com/toyota-jishuken/.

Argyris, Ch., and D.A. Schön. “Organizational Learning: A Theory of Action Perspective.” *Revista Española de Investigaciones Sociológicas* 77/78 (1997): 345-348.

Baudin, M. “value-stream mapping, kaizen blitzes, and jishuken. lean manufacturing blog.” *LinkedIn Pulse.* 15 11 2012. https://www.linkedin.com/pulse/20121115174924-337087-value-stream-mapping-kaizen-blitzes-and-jishuken/.

Braun, V., and V. Clarke. “Using thematic analysis in psychology.” *Qualitative Research in Psychology* 3, no. 2 (2006): 77-101.

Conti, R., J. Angelis, C. Cooper, B. Faragher, and C. Gill. “The Effects of Lean Production on Worker Stress.” *International Journal of Operations & Production Management* 26, no. 9 (2006): 1013-1038.

de Treville, S.,, and J. Antonakis. “Could lean production job design be intrinsically motivating? Contextual, configurational, and levels-of-analysis issues.” *Journal of Operations Management* 24, no. 2 (2006): 99-123.

Denzin, Norman K. *The research act : a theoretical introduction to sociological methods (3rd ed).* Englewood Cliffs, N.J. : Prentice Hall, 1978.

Dyer, Jeffrey H., and Kentaro. Nobeoka. “Creating and managing a high-performance knowledge-sharing network.” *Stretegic Management Journal - Special Issue: Strategic Networks* 21, no. 3 (2000): 345-367.

Gilson, L. L., and C. B. Goldberg. “Editors' comment: So, what is a conceptual paper?” *Group & Organization Management* 40, no. 2 (2015): 127-130.

INEAK. *Working with Suppliers for Mutual Learning of TPS.* 14 April 2009. https://www.ineak.com/working-with-suppliers-for-mutual-learning-of-tps/.

Kimura, K. “TQM-14 TQC and TQM.” *Internet Archive.* 01 01 2026. https://archive.org/details/tqm-14-tqc-and-tqm/mode/2up.

Kimura, Koichi. “TQM 13 Taguchi Method.” *Internet Archive.* 15 10 2025. https://archive.org/details/tqm-13-taguchi-method.

Kimura, Koichi. “TQM 1 - Prologue, And What is TQM.” *Internet Archive.* 01 08 2022. https://archive.org/details/tqm-1-prologue-and-what-is-tqm.

kimura, Koichi. “TQM 5 The KJ Method.” *Internet Archive.* 01 09 2023. https://archive.org/details/tqm-5-the-kj-method.

Kimura, Koichi. “TQM-8, The Arrow Diagram.” *Internet Archive.* 01 08 2024. https://archive.org/details/tqm-8-the-arrow-diagram.

Landsbergis, P. A., J. Cahill, and P. Schnall. “The Impact of Lean Production and Related New Systems of Work Organization on Worker Health.” *Journal of Occupational Health Psychology* 4, no. 2 (1999): 108-130.
Liker, J. K. *The Toyota Way: 14 Management Principles from the World's Greatest Manufactuer.* McGraw-Hill, 2004.
Marksberry, Phillip, Fazleena Badurdeen, and Bob Gregor. “Management directed kaizen: Toyota’s Jishuken process for management development.” *https://cdn.chools.in/.* 2010. https://cdn.chools.in/LEAN_PDF/Kaizen%20Toolkit/Management_directed_kaizen_Toyotas_Jishu.pdf?
Miller, Jon. “What is Jishuken?” *Gemba Academy.* 2 April 2021. https://blog.gembaacademy.com/2006/08/27/what_is_jishuken/.
Nonaka, Ikujiro, and Noboru Konno. “The Concept of “Ba”: Building a Foundation for Knowledge Creation.” *California Management Review* 40, no. 3 (1998): 40-54.
Nonaka, Kujiro, and Hirotaka Takeuchi. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation.* Oxford University Press, 1995.
Ohno, T. *Toyota Production System: Beyond Large Scale Production.* Productivity Press, 1988.
Ram Kumar, S., V Nimesh Nathan, S.I. Mohammed Ashique, V Rajkumar, and P. Arun Karthick. “Productivity enhancement and cycle time reduction in toyota production system through jishuken activity – Case study.” *Science Direct* 37, no. 2 (2021): 964-966.
Schön, Donald A. *The Reflective Practitioner.* Bacis Books, 1983.
Senge, Peter. *The fifth discipline: The art and practice of the learning organization.* Doubleday/Currency, 1990.
Teece, D. J. , G. Pisano, and A. Shuen. “Dynamic Capabilities and Strategic Management.” *Strategic Management Journal* 18, no. 7 (1997): 509-533.
Teece, D. J. “Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance.” *Strategic Management Journal*, Ago 2007: 1319-1350.
Toyota Matereial Handling. *Jishuken project.* 26 02 2026. https://toyota-forklifts.eu/toyota-lean-academy/jishuken-project/.
Yokoi, Hideshi. *Jishuken at Toyota Motor Kyushu.* 2 March 2019. https://www.planet-lean.com/articles/jishuken-toyota-hideshi-yokoi.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

A Process-Based Measurement Framework for Teaching Quality in Higher Education: Development of a Composite Index (TQI) and Illustrative Application

Aneta Kucińska - Landwójtowicz [1], Izabela Czabak – Górska [1]

[1] Opole University of Technology, Faculty of Production Engineering and Logistics; a.kucinska-landwojtowicz@po.edu.pl; i.czabak-gorska@po.edu.pl

Abstract – **The aim of the paper is to develop a process-based framework to be used for measuring teaching quality in higher education and to operationalize it by constructing a composite Teaching Quality Index (TQI) at the program level. The study adopts a process-oriented perspective, structuring teaching quality around the management, the core as well as supportive processes. A set of program-level indicators is defined, then normalized, weighted, and aggregated into a composite index. The proposed framework is illustrated through an empirical application that demonstrates its usefulness for evaluating and comparing study programs within a higher educational institution. The findings highlight that teaching quality is inherently multidimensional and shaped by the interaction of interrelated organizational processes. The TQI provides a coherent and operational tool that supports internal benchmarking, identifies strengths and weaknesses across programs, and facilitates evidence-based decision-making. The originality of the study lies in its process-based approach to teaching quality assessment, the development of a composite index tailored for program-level evaluation, and the integration of students' dropout as a structurally embedded component of quality assessment rather than an isolated outcome metric. The proposed model offers both theoretical and practical implications for quality management in higher education.**

Keywords – Teaching quality evaluation, Teaching Quality Index (TQI), Composite index, Higher education

Conceptual paper

I. INTRODUCTION

Higher education institutions (HEIs) increasingly operate in a complex and rapidly changing environment shaped by technological advancement, the growing role of artificial intelligence and ever evolving labor-market demands. These developments challenge traditional models of academic governance and expose the limitations of existing approaches to managing and evaluating educational processes. At the same time, HEIs operate in a highly competitive global landscape in both education and research, while facing growing internal pressure to demonstrate the effectiveness of their activities. In this context, university rankings have become a dominant force, shaping the higher education sector, reinforcing the need for more structured approaches to performance management. As noted by (Kallio *et al.*, 2017), measurement of performance in universities is often reduced to a limited set of indicators, primarily focused on a research output, teaching activities, and graduates' numbers. Moreover, increasing attention is being given to the argument that strongly research-oriented evaluation systems may negatively affect creativity and innovation in teaching (Cadez *et al.*, 2017). As a result, the ability to monitor, assess, and continuously improve teaching quality has become a key organizational capability rather than a purely supporting function. Despite the growing importance of quality assurance, existing evaluation frameworks remain conceptually and methodologically fragmented. This fragmentation reflects both differences in national quality assurance systems and variations in institutional practices (Cadez *et al.*, 2017; Chui *et al.*, 2016; Fajčíková and Fejfarová, 2019; Hernandez-Diaz *et al.*, 2020; Kallio *et al.*, 2017; Kohoutek *et al.*, 2025; Moreno-Gómez *et al.*, 2020; Santiago *et al.*, 2015).

Defining and operationalizing the quality of teaching remains particularly challenging. In the literature, quality is often associated with students' and employers' satisfaction concerning educational services (Gulden *et al.*, 2020). While this perspective highlights the importance of stakeholders, it also reveals the limitations of relying solely on perceptual measures. Satisfaction is inherently subjective, context-dependent, and influenced by factors beyond the direct control of educational processes. Student's evaluations, therefore, provide only a partial view of teaching quality and need to be complemented by other indicators to achieve a more comprehensive and objective assessment (Mao *et al.*, 2025). Therefore, its assessment should be complemented by structured internal evaluation mechanisms that support a more comprehensive, process-oriented understanding of teaching quality. At the same time, there is a need for a continuous improvement of the quality of educational activities within educational processes (e.g., scientific circles, internationalization, and access to internships and quality of them) simultaneously enhancing the educational value and relevance for students.

Existing approaches tend to emphasize either outcome-based metrics, such as the academic reputation, graduates' employability, faculty–student ratios, and completion rates (Abubakar *et al.*, 2018; Bogt and Scapens, 2012), or perceptual measures, such as: student's experience and satisfaction (Arnold, 2008; Bedggood and Donovan, 2012). At the same time, they often overlook the systemic and process-oriented nature of educational quality. Performance measurement systems typically support macro-level evaluation at the institutional level (Angiola *et al.*,

2018; Hernandez-Diaz *et al.*, 2020; Yuan and Li, 2021), whereas less attention is paid to the dynamics of individual educational processes. From a managerial perspective, however, the quality of teaching within specific study programs plays a critical role in shaping overall institutional outcomes.

In this context, composite indices offer a promising approach for integrating diverse performance dimensions into a unified framework, thereby enabling a more comprehensive internal evaluation of the educational quality. Composite indices have been widely used to aggregate multidimensional information into a single measure (OECD, 2008). In higher education, they are most often applied in rankings and performance evaluation systems based on selected indicators or specific quality dimensions. A substantial body of research focuses on the institutional-level efficiency and the quality assessment, using methods such as Data Envelopment Analysis (DEA), Benefit-of-the-Doubt (BoD) models, Principal Component Analysis (PCA), and composite rankings (El Gibari *et al.*, 2018, 2022; Murias *et al.*, 2008; Râlea, 2024; Szuwarzyński and Julkowski, 2014). These studies demonstrate the usefulness of composite indicators for comparing universities, identifying efficient units and supporting benchmarking. However, they typically focus on the university as a whole rather than on individual study programs. Related work also includes system-oriented and process-based evaluation frameworks covering areas such as governance, student's services, faculty support and resource allocation (Zhai and Zheng, 2025) as well as KPI-based approaches towards performance management in higher education (Asif and Searcy, 2014). While these studies provide valuable insights into the institutional performance and the quality management, they are generally oriented towards a high-level evaluation and governance rather than a program-level, process-based assessment of the teaching quality.

The literature confirms that composite indices are a valid and useful tool for the performance measurement in higher education. At the same time, their construction requires a careful consideration of methodological choices. Studies such as (OECD, 2008), (Benito and Romera, 2011; Chakrabartty, 2017; Gnaldi and Ranalli, 2016), and (Williams and de Rassenfosse, 2016) emphasize the importance of decisions, related to normalization, weighting, and aggregation. These choices can significantly influence the resulting rankings and interpretations, highlighting the need for transparency and robustness in a composite index design.

Contrary to this background, this paper aims to develop a process-based framework for an educational process evaluation in higher education and operationalize it through the construction of a composite index (TQI), demonstrating its applicability using an illustrative case. The proposed framework offers a structured, process-based approach to the measuring of teaching quality, integrating multiple dimensions of educational processes into a coherent composite index while accounting for students' dropout within the broader assessment framework.

The remainder of the paper is structured as follows. Section II outlines the problem statement and research methodology. Section III presents the proposed process-based measurement framework for teaching quality in higher education. Section IV demonstrates its application and concludes the paper.

II. Problem statement and research methodology

As evidenced in the literature, performance measurement in higher education remains an important area of research from both practical and academic perspectives (Zhai and Zheng, 2025). While prior studies provide valuable methodological and institutional foundations for quality and performance assessment, relatively little attention has been given to developing a process-based composite index focused specifically on teaching quality at the program-level, particularly for internal evaluation and benchmarking in technical universities. Although numerous system-oriented approaches to evaluating teaching quality have been proposed (Feng, 2023; Li, 2026; Li *et al.*, 2022; Li and Hu, 2022; Lu and Hao, 2022; Mao *et al.*, 2025), most operate at the system or institutional level and focus on selected performance dimensions. As a result, they offer only a partial view of the teaching quality and fail to fully capture the interdependencies among processes that shape the educational outcomes. In particular, relatively few studies adopt a perspective that integrates management, core, and support activities within a unified analytical framework (Kucińska-Landwójtowicz *et al.*, 2023). Moreover, while students' dropout is widely recognized as an important outcome variable, it is rarely analysed within a broader process-based view of educational quality (Asif and Searcy, 2014; Dwaikat, 2020; Latif *et al.*, 2019). To the best of the authors' knowledge, existing studies haven't developed a process-based composite index of the teaching quality at the program level that explicitly integrates students' dropout within a unified analytical framework.

Against this background, there is a need for a structured, operational composite measure that enables a more comprehensive and comparable assessment of teaching quality at the program level and can serve as a decision-supportive tool for the internal evaluation, benchmarking, and management, particularly in technical universities. To address this need, this study proposes the Teaching Quality Index (TQI) as a process-based composite measure that captures the multidimensional nature of the teaching quality. The index integrates indicators obtained from management, core, and support processes within a coherent analytical framework, enabling a more systemic, analytically consistent evaluation of the educational performance.

The contribution of this study is threefold. First, it advances the conceptualization of teaching quality by framing it as an emergent property of interacting educational processes. Second, it operationalizes this perspective by developing a composite index that integrates diverse indicators into a consistent, process-oriented measurement framework. Third, it incorporates students' dropout as a structurally embedded component of the teaching quality assessment, rather than treating it as an isolated outcome variable.

Based on the identified research gap, the following research questions are formulated:

RQ1: How can a process-based decomposition of the teaching quality into management, core, and support processes be used to develop a coherent measurement framework?

RQ2: How can a composite index (TQI) be constructed to operationalize a process-based approach to teaching quality assessment?

RQ3: How should indicators be structured and integrated within the TQI to ensure consistency, comparability, and interpretability of results?

RQ4: How can students' dropout be conceptually and operationally integrated as a component of teaching quality within the proposed framework?

This paper contributes to the literature in three ways. It proposes a process-based measurement framework for the teaching quality in higher education, operationalizes it through the development of the TQI, and demonstrates its applicability with an illustrative case, highlighting its potential to support decision-making and quality management in higher education institutions.

To implement this approach, the study follows a structured multi-stage procedure, including: (1) the formalization of a process-based indicator system aligned with the functional architecture of higher education institutions, (2) the normalization of heterogeneous indicators to ensure comparability, (3) the application of weighting schemes reflecting both structural and strategic priorities, and (4) the aggregation of indicators into a composite index enabling comparative assessment across study programs. In this procedure, students' dropout is included as an indicator within the core process dimension.

The literature review was conducted using the Scopus database. The search was conducted using the advanced search function, with the following keywords: '*quality evaluation*' and '*quality control*' and '*high educations*'. In the initial stage, 267 publications were identified. After excluding 24 retracted papers, a total of 243 records remained for further analysis. Subsequently, the selection process was refined through an abstract review, which enabled the elimination of publications not directly related to the subject of study. As a result, 31 articles were selected for detailed analysis. The final set of publications served as the basis for developing the proposed framework and its methodological assumptions. Fig. 1 presents the research scheme.

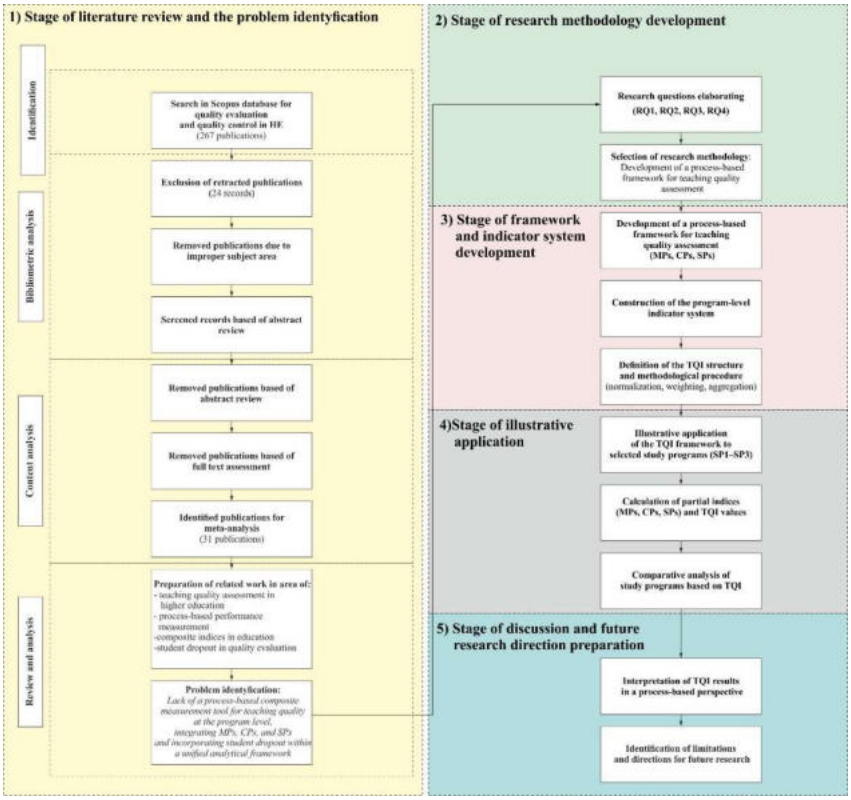


Fig. 1. Scheme of the conducted research.

III. A Process-Based Measurement Framework for Teaching Quality in Higher Education

The framework builds on a process-based view of higher education institutions, distinguishing between management (MP), core (CP), and support (SP) processes. The Teaching Quality Index (TQI) is a composite measure of teaching quality at the program level. Within this perspective, teaching quality is understood as an emergent property of interactions among these process groups.

A. Processes stucture

The process structure adopted in this study follows on a previously developed performance measurement model for higher education institutions (Kucińska-Landwójtowicz *et al.*, 2023), which identified detailed sets of MPs, CPs, and SPs. In that model, management processes included activities such as: development of the quality policy, teaching process management, human resource management, and continuous improvement. Core processes covered program design, students' recruitment and the education process itself, while support processes encompassed: infrastructure, information management, the program evaluation, IT systems, and stakeholders' cooperation. In the present study, this detailed architecture is preserved conceptually but aggregated into three higher-level process dimensions (MPs, CPs, SPs) to operationalize teaching quality within a composite measurement framework. This simplification maintains the process-based logic of the original model while ensuring analytical clarity and tractability. Rather than replicating the full granularity of the initial framework, it serves as a conceptual foundation for defining the Teaching Quality Index (TQI) structure.

Within this framework, each process group fulfills a distinct but complementary role. Management processes define the strategic and organizational context of teaching by setting priorities, shaping quality policies as well as providing governance mechanisms. Core processes represent the central domain of teaching quality, encompassing teaching, learning and student's progression. Support processes, although not directly involved in teaching delivery, provide the infrastructural and organizational conditions necessary for the effective functioning of core activities. A key feature of the framework is the inclusion of students' dropout within the core process dimension. While dropout is traditionally treated as an outcome variable, here it is interpreted as a process-sensitive indicator reflecting disruptions in student's progression and the continuity of the educational pathway. From this perspective, the dropout is closely linked to the effectiveness of teaching, curriculum design, assessment practices, and student's support mechanisms operating within the core processes. This interpretation allows the dropout to be used not only as a measure of outcomes but also as a diagnostic signal of underlying process inefficiencies. At the same time, it retains its role as an outcome variable at the system level, supporting the interpretation and validation of TQI results. Put together, this dual perspective helps to bridge the process-based measurement and the outcome-oriented evaluation.

The following designations for processes and subprocesses are adopted for the program-level perspective:

1) ***M – management processes****:*
- **M1** – management of the teaching process within the study program,
- **M2** – management of teaching staff assigned to the program,
- **M3** – provision of material and teaching resources for the program,
- **M4** – management of information and communication processes supporting the program.

2) ***C – core processes****:*
- **C1** – design and development of study programs,
- **C2** – recruitment to the study program,
- **C3** – implementation of the education process and student's progression within the program.

3) ***S – support processes****:*
- **S1** – maintenance of teaching infrastructure used within the program,
- **S2** – assessment and assurance of the teaching quality within the program,
- **S3** – provision of information flow and the teaching documentation for the program,
- **S4** – evaluation of the program outcomes and learning results,
- **S5** – cooperation with external stakeholders in the context of the program.

Fig. 2 presents the process-based framework for the teaching quality measurement. The model integrates MP, CP, and SP processes within a composite structure, with the teaching quality represented by the Teaching Quality Index (TQI). It highlights the interactions among process groups and their role in shaping program-level teaching quality.

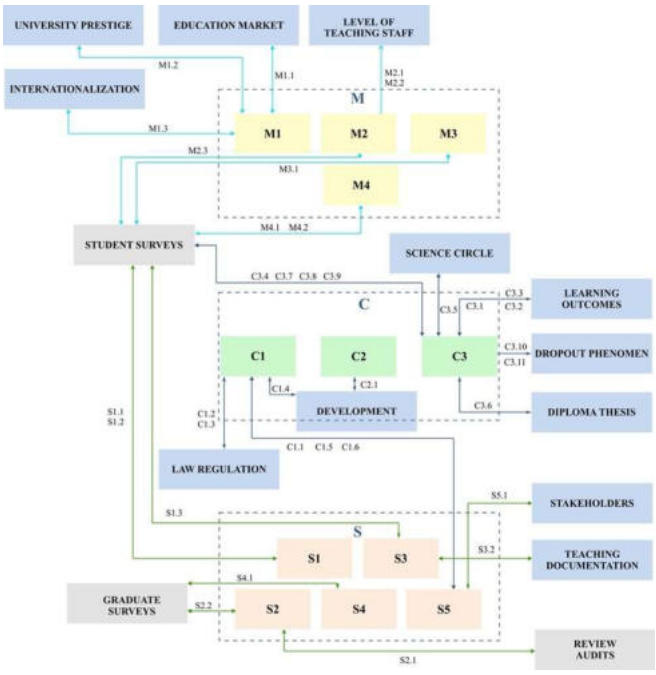


Fig. 2. Process-based architecture of the Teaching Quality Index (TQI) within a performance measurement system for higher education.

B. Indicator system

The indicator system adopted in this study builds on the process-based performance measurement model developed in prior research (Kucińska-Landwójtowicz et al., 2023), which assigns a comprehensive set of indicators to MPs, CPs, and SPs. In the present framework, this approach is retained at the conceptual level but refined to support the construction of a composite index focused on the teaching quality at the program level. Rather than replicating the full set of institutional indicators, the model selects and restructures those that can be directly linked to the design, delivery and outcomes of the individual study programs. In particular, only indicators that could be clearly attributed to the program-level were retained. Indicators reflecting purely institutional characteristics or those uniformly defined across all programs (e.g., centrally imposed conditions and standards) were excluded, as they do not differentiate between the programs and therefore do not support the comparative assessment. The final selection of indicators was guided by three criteria: relevance, non-redundancy and comprehensiveness. First, only indicators directly related to the design, delivery or outcomes of individual study programs were included. Second, indicators were selected in order to capture distinctive aspects of the teaching quality, thereby limiting redundancy within the framework. Although some indicators address related areas, it is that they provide complementary perspectives rather than duplicate information. Third, the indicator system was designed to reflect the multidimensional nature of the teaching quality by integrating: the management, core, and support processes. To achieve this, the framework combines structural indicators describing program characteristics and resources, performance indicators reflecting educational outcomes as well as stakeholder-based measures derived from students' and graduates' evaluations. As a result, the proposed indicator system provides a balanced representation of the teaching quality at the program level while maintaining consistency with the framework's process-based architecture.

Each process group is operationalized through a set of program-level indicators reflecting its role in shaping the teaching quality. Management indicators capture aspects of the program governance, the teaching organization and the allocation of the academic resources. The core indicators represent the central teaching and learning activities, including the curriculum design, the educational delivery, and student's progression. The support indicators reflect infrastructural and organizational conditions which enable effective implementation of the educational processes. A distinctive feature of the proposed system is the explicit incorporation of students' dropout within the core process dimension. Unlike traditional approaches, which treat the dropout primarily as an outcome variable, this framework treats it as a process indicator reflecting disruptions in student's progression and the continuity of the educational pathway. Both the dropout rate of students in the first year of studies and in subsequent years were taken into account. This allows the dropout to be interpreted not only as a final result but also as a signal of underlying inefficiencies in the core educational processes.

Table 1 presents the set of program-level indicators used in the TQI framework. Their selection is based on relevance to program-level evaluation and on the consistency with the model's process-based structure (figure 3). The table also details the methodological basis for calculating each indicator and outlines its relevance within the program-level evaluation framework. This structured approach provides a coherent, operational representation of the

teaching quality, serving as the basis for a subsequent normalization, weighting, and aggregation within the Teaching Quality Index (TQI). Together, these indicators provide an operational representation of teaching quality and form the basis for the subsequent stages of the TQI construction procedure.

TABLE I. PROGRAM-LEVEL INDICATORS USED IN THE TQI FRAMEWORK

Process		*Subject of analysis*	*Indicator*	*Description*
M – management processes	**M1**	Flexibility of the education process	**M1.1** % of teaching hours conducted in terms of the form of education	Assessment of the ability to adapt to changes in the socioeconomic environment by the flexibility of the course (on-site/remote/hybrid)
			M1.2 number of points obtained in the indicated rankings of study fields	External evaluation of the position of universities on the educational services market
			M1.3 % students participating in international exchange programs	Assessment of the university's activity in international exchange projects
	M2	Teaching staff quality	**M2.1** Student–teacher ratio	Assessment of the quality level of the didactic process through indicators related to the organization of work
			M2.2 Teaching staff evaluation (classroom auditing)	The average assessment of classroom auditing for lectures/projects/laboratories/seminars (rating scale in surveys: 1–10)
			M2.3 Teaching staff evaluation (student surveys)	The average grade of an academic teacher based on student surveys (rating scale in surveys: 1–10)
	M3	Teaching infrastructure (program-related use)	**M3.1** Student evaluation of infrastructure	The average grade for teaching infrastructure based on student questionnaires (rating scale in surveys: 1–10)
	M4	Communication and digitalization	**M4.1** Website evaluation (students)	The average grade of the website evaluation based on student surveys (rating scale in surveys: 1–10)
			M4.2 Student evaluation of communication	The average grade of internal communication understood as direct contact with teaching staff based on student surveys (rating scale in surveys: 1–10)
C – core processes	**C1**	Program design and adaptation	**C1.1** % of classes conducted by industry representatives	Assessment of adapting education programmes to the needs of the industry
			C1.2 % of laboratory hours	Evaluation of the study program in terms of the number of laboratory classes and practical skills acquired
			C1.3 % of practical classes	Evaluation of the study program in terms of the number of practical classes (projects, exercises, seminars) and practical skills acquired
			C1.4 % of new specializations in the fields of study	Assessment of development of new fields of education adapted to the socioeconomic environment
			C1.5 internships in the study program evaluation	Assessment of whether the study program includes professional internships (0/1)
			C1.6 Average duration of internships	Number of internship hours in the study program
	C2	Recruitment	**C2.1** Number of students enrolled	Confirmed number of students accepted for the field of study
	C3	Educational process performance and learning outcomes	**C3.1** % graduating on time	Assessment of the process of preparing and defending the thesis within the timeframe required by the legal regulations of the study program
			C3.2 % of graduates in employment	Assessment of level of achievement of learning outcomes
			C3.3 % of students passing exams on time	Assessment of level of achievement of learning outcomes
			C3.4 Student survey results (learning outcomes)	The average grade of learning outcomes based on student surveys (rating scale in surveys: 1–10)
			C3.5 % students active in scientific circles	Assessment of the quality of the offer of science circles
			C3.6 Average grade of diploma theses	Assessment of quality of/ability to implement diploma theses
			C3.7 Student evaluation of classes	The average grade of classes based on student surveys (rating scale in surveys: 1–10)
			C3.8 Evaluation of remote classes	The average grade of remote classes based on student surveys (rating scale in surveys: 1–10)
			C3.9 Internship student evaluation	The average grade of internship based on student surveys (rating scale in surveys: 1–5)
			C3.10 % of students dropping out of studies in the first year of study	Dropout rate in the first year
			C3.11 % of students dropping out of studies in the following years of study	Dropout rate in the following years
S – support processes	**S1**	Teaching infrastructure	**S1.1** Quality of teaching rooms evaluation	The average grade of teaching rooms based on student surveys (rating scale in surveys: 1–5)
			S1.2 Quality of e-learning platform evaluation	The average grade of e-learning based on student surveys (rating scale in surveys: 1–5)
	S2	Teaching quality assurance	**S2.1** Internal audit results	Assessment through class inspections The results of internal audits (audit grading scale: 0–5)
			S2.2 Graduate survey results	The average grade of teaching activities based on graduate surveys (rating scale in surveys: 1–5)
	S3	Student services	**S3.1** Student service evaluation results	The average grade of student service based on graduate surveys (rating scale in surveys: 1–5)
			S3.2 Average ratings for audited didactic documentation	The assessment based on a review of the didactic documentation conducted by the supervisor (rating scale: 1–5)

	S4	Program outcomes and labor market alignment	**S4.1** % employed graduates after 1 years	The assessment of compliance of education programmes with the requirements of the labour market
	S5	Cooperation with the buisness environment	**S5.1** % of completed student internships	% of successfully completed internships

The indicators presented in Table 1 were selected to reflect the multidimensional nature of the teaching quality from a program-level perspective. Their selection was guided by three criteria: relevance to the educational process, complementarity and the ability to differentiate between study programs. Consequently, only indicators directly linked to the design, delivery, support or outcomes of individual programs were retained. The indicator system combines several types of measures. Structural indicators describe program characteristics and educational resources, such as laboratory classes, internships, or students–teachers ratios. Process indicators capture the organization and implementation of educational activities, including the teaching quality assurance, communication, and student's support. Outcome indicators reflect educational results through variables such as graduation, employment and dropout rates. In addition, stakeholder-based indicators derived from stuents' and graduates' evaluations provide insight into how educational processes are experienced by their participants. Redundancy was limited by assigning indicators to distinct process dimensions and retaining only those, that captured unique aspects of a program performance. Although some indicators relate to similar areas, they represent different perspectives on teaching quality and therefore provide complementary rather than duplicate information. For example, students' evaluations, graduates' surveys and internal audit results assess educational quality from different viewpoints. Put together, the indicators provide a balanced representation of the management, core, and support processes while preserving the process-based logic of the framework.

C. Teaching Quality Index (TQI) – composite index

The teaching quality at the program level arises from the interaction of MPs, CPs, and SPs and is assessed through a unified analytical framework. The construction of the index follows a structured procedure consistent with established guidelines for composite indicator development (OECD, 2008). A two-level weighting structure is adopted. At the first level, weights are assigned to individual indicators within each process group (MPs, CPs, SPs), reflecting their relative contribution to the corresponding partial indices. At the second level, weights $\alpha, \beta,$and $\gamma$are assigned to the three process dimensions, capturing their overall importance in the construction of the TQI. Figure 3 presents the framework's logic.

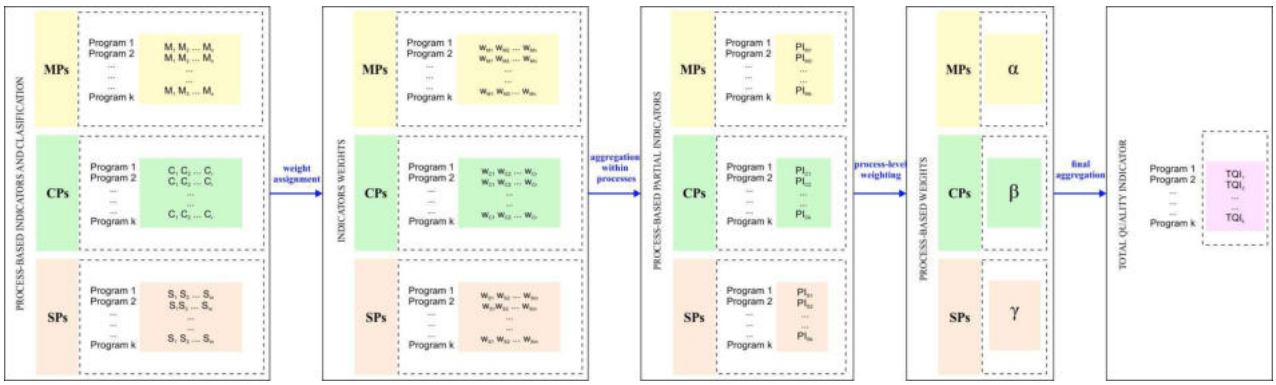


Fig. 3. The logical structure of the Teaching Quality Index (TQI).

The TQI is developed through a five-stage procedure that progressively transforms raw indicators into a single composite measure of teaching quality. First, indicators are classified according to their direction of preference (stimulants or destimulants). This distinction reflects the fact that not all indicators influence the teaching quality in the same direction. For stimulants, higher values indicate better performance and contribute positively to the teaching quality (e.g., graduate employability, student satisfaction, or participation in international exchange programs). In contrast, for destimulants, lower values are preferred because higher values indicate lower process effectiveness or educational quality (e.g., higher students' dropout rates or a higher students–staff ratio. This classification ensures that all indicators are interpreted consistently before normalization, regardless of whether higher or lower values are considered desirable. Second, all indicators are normalized to a common scale to eliminate differences in measurement units and value ranges, thereby enabling direct comparison and subsequent aggregation. Without normalization, indicators expressed in different units (e.g., percentages, survey scores or absolute counts) could not be meaningfully combined within a single framework. The outcome is a set of dimensionless values expressed on a common scale, preserving differences between the study programs while preventing any single indicator from dominating the analysis due to its numerical range. These normalized values serve as the basis for the weighting and aggregation procedures applied in subsequent stages. Third, weights are assigned to indicators within each process group in order to reflect their relative importance in the assessment framework. This stage determines the contribution of individual indicators to the corresponding process dimension and allows the framework to account for differences in their perceived relevance. Weighting, therefore, introduces a mechanism for translating strategic priorities or expert judgments into the measurement model. The outcome of this stage is a weighted set of normalized indicators, which forms the basis for constructing synthetic measures of management, core, and support process performance. Fourth,

the normalized and weighted indicators are aggregated into partial indices representing the management, core, and support processes. This stage simplifies the indicator system by combining multiple variables into three synthetic measures that correspond to the framework's main process dimensions. As a result, multidimensional information contained in individual indicators is transformed into concise measures of process performance, facilitating interpretation and comparison across the study programs. These partial indices provide a structured representation of strengths and weaknesses within individual process groups and serve as direct inputs for the construction of the overall TQI. Finally, the partial indices are combined into the overall Teaching Quality Index (TQI), providing a single composite measure of teaching quality at the program level. This final aggregation stage integrates information from the management, core, and support processes into a unified assessment framework while preserving the model's process-based structure. The resulting TQI value enables comparison between the study programs, supports benchmarking activities and provides a quantitative basis for monitoring the educational performance and identifying areas, which require improvement.

Stage 1. Indicator classification.

The indicators are first classified by direction of preference, distinguishing between stimulants (higher values indicate better performance) and *destimulants* (lower values indicate better performance). This distinction ensures consistent interpretation in subsequent stages of the analysis. As shown in the Tab. Most indicators are stimulants, reflecting improvements in teaching quality, effectiveness of educational processes, resource availability, and student's outcomes. These include, in particular, measures related to the teaching staff's evaluation, infrastructure quality, student's satisfaction, employability, and engagement in educational and research activities. Only a limited number of indicators are treated as *destimulants*. These include the students' dropout rate, the student–staff ratio, and the time required to revise the study program, as higher values of these variables indicate lower teaching quality and process efficiency. The classification of indicators into stimulants and destimulants constitutes a prerequisite for the normalization stage. Without this distinction, indicators that affect teaching quality in opposite directions could not be meaningfully aggregated into a common composite framework.

Stage 2: Normalization.

To ensure comparability, all indicators are transformed to a common [0,1] scale using min–max normalization. This approach is widely applied in composite index construction due to its transparency and a straightforward interpretation. While alternative normalization methods exist, including point-based and distance-based approaches (e.g., (El Gibari *et al.*, 2018)), they are primarily designed for benchmarking and efficiency analysis. In contrast, the present framework focuses on program-level assessment and integration of heterogeneous indicators, for which min–max normalization is particularly well suited. For *stimulants*, the normalized value is calculated as:

$$x_{ij}^{*} = \frac{x_{ij} - \mathrm{mi}\ (x_j)}{\max(x_j) - \min(x_j)}, \quad (1)$$

whereas for *destimulants*:

$$x_{ij}^{*} = \frac{\max(x_j) - x_{ij}}{\max(x_j) - \min(x_j)}, \quad (2)$$

where: x_{ij} denotes the value of indicator j for program i, and x_{ij}^{*} is its normalized value. In cases where no variation is observed for a given indicator across the analyzed programs (i.e., when $max(x_j) = min(x_j)$), the normalized value is set to 1 for all units. This reflects the fact that such indicators do not differentiate between programs but still represent conditions uniformly fulfilled across the analyzed units. The normalization procedure preserves the relative differences between programs while ensuring that all indicators are expressed on a comparable scale.

Stage 3: Weighting.

Each indicator is assigned a weight w_j, reflecting its relative importance within the assessment framework. The weights are defined at the level of individual indicators and express their contribution to the corresponding process dimension. Weighting is performed within each process group (MPs, CPs, and SPs). This means that the weights are normalized separately for each group, such that:

$$\sum_{j \in k} w_j = 1, \mathrm{k} \in \{\mathrm{MP}, \mathrm{CP}, \mathrm{SP}\}. \quad (3)$$

Thus, w_j represents the relative importance of indicator j in relation to other indicators belonging to the same process group, rather than across the entire set of indicators. In the baseline specification adopted in this study, equal weights are assigned to indicators within each process group. Consequently, for a process group containing n_k indicators, the weight of each indicator is defined as $w_j = 1/n_k$. Alternative weighting schemes, based on expert judgment or data-driven approaches, may also be applied, depending on the objectives of the analysis. Thus, w_j represents the relative importance of indicator j within a given process group rather than across the entire set of indicators.

In the illustrative application presented in this study, equal weights are assigned to indicators within each process group and to the three process dimensions. This assumption was adopted to ensure transparency of the calculation procedure and to avoid the introduction of additional subjective assumptions into the illustrative example. However, the proposed framework is not restricted to equal weighting. Depending on the objectives of the analysis and

institutional priorities, alternative weighting approaches may be applied. These include expert-based methods, such as Delphi panels or the Analytic Hierarchy Process (AHP), as well as data-driven approaches based on statistical properties of the indicators. Different weighting schemes may lead to variations in partial indices and final TQI values, particularly when substantial differences exist between the study programs. Consequently, the choice of weights should be aligned with the intended purpose of the assessment and may be examined through sensitivity and robustness analyses in future empirical applications of the framework.

Stage 4: Aggregation at process level.

For each process group (MP, CP, SP), partial indices are calculated as weighted sums of normalized indicators:

$$PI_k, i = \sum_{j \in k} w_j \cdot x_{ij}^*, \tag{4}$$

where: $k \in \{MP, CP, SP\}$, w_j denotes the weight of indicator j within process group k, normalized such that $\sum_{j \in k} w_j = 1$.

Stage 5: Construction of the TQI.

The overall Teaching Quality Index is calculated as an aggregate of the partial indices:

$$TQI_i = \alpha \cdot PI_{MP,i} + \beta \cdot PI_{CP,i} + \gamma \cdot PI_{SP,i}, \tag{5}$$

where: α, β, γ denote weights assigned to the three process dimensions, such that $\alpha + \beta + \gamma = 1$. The weights α, β and γ reflect the relative importance assigned to MPs, CPs and SPs, respectively. These weights reflect the relative importance of the three process dimensions. While equal weights may be adopted in the baseline specification, the framework allows for flexible adjustment depending on institutional priorities. In particular, greater emphasis may be placed on core processes, given their central role in delivering teaching quality. The resulting TQI provides a synthetic measure of teaching quality at the program-level, enabling comparative analysis, benchmarking, and performance monitoring across study programs.

IV. ILUSTRATIVE APPLICATION

To illustrate the proposed framework, the Teaching Quality Index (TQI) was calculated for three sample study programs conducted at a technical university: study program 1 (SP1), study program 2 (SP2), and study program 3 (SP3). The example is based on a hypothetical dataset constructed solely for methodological illustration and does not constitute an empirical validation of the proposed framework. Its purpose is to demonstrate the sequence of calculations and the interpretation of obtained results, generated by the TQI methodology. The analysis is based on the set of indicators defined in Section III. A (Processes structure) and Section III.B (Indicator system), and follows the procedure described therein. In the baseline specification, equal weights were assigned to indicators within each process group. This means that for MPs, each indicator was assigned a weight of $1/9 \approx 0.111$, for CPs a weight of $1/18 \approx 0.056$, and for SPs a weight of $1/8 \approx 0.125$. Similarly, equal weights were assumed for the three process dimensions at the aggregation stage, with $\alpha = \beta = \gamma = 1/3$. The calculation procedure included normalization of indicators, indicator-level weighting, aggregation into partial indices for MPs, CPs, and SPs, and, finally, the construction of the composite index.

For illustrative purposes, two indicators may be used to demonstrate the calculation logic. Following the classification introduced in Stage 1, C3.1 (% graduating on time) is considered a stimulant, whereas the students' dropout rate is considered a *destimulant*. Accordingly, the normalization procedure is applied using eq. (1) for *stimulants* and eq. (2) for *destimulants*. For example, in the case of C3.1, the observed values for the three analyzed programs were 80%, 65%, and 50%. For the SP1 program, the normalized value is therefore calculated as $(80 - 50)/(80 - 50) = 1.000$. In turn, for the SP2 program the result is $(65 - 50)/(80 - 50) = 0.500$, while for SP3 it equals $(50 - 50)/(80 - 50) = 0.000$. By contrast, C3.10 (the dropout rate in the first year) in the first year is treated as a *destimulant*. The observed values for the analyzed programs were 23%, 25%, and 30%. The normalized value for SP1 is $(30 - 23)/(30 - 23) = 1.000$, for SP2 $(30 - 25)/(30 - 23) = 0.714$, and for SP3 $(30 - 30)/(30 - 23) = 0.000$. The results of the analysis are presented in Tab. 2.

TABLE II. RESULTS OF THE TQI CALCULATION FOR THE ANALYZED STUDY PROGRAMS, INCLUDING PARTIAL INDICES FOR SP1, SP2, AND SP3.

Study program	*Indicator*				*Rank*
	PI_{MPs}	PI_{CPs}	PI_{SPs}	*TQI*	
SP1	0.722	0.536	0.339	0.533	2
SP2	0.631	0.555	0.771	0.652	1
SP3	0.136	0.442	0.646	0.408	3

The illustrative application shows how the Teaching Quality Index (TQI) can be used to compare study programs using a set of process-oriented indicators. The results for the three analyzed programs point to clear differences in their performance across management, core, and support processes. The highest TQI value was observed for the SP2, at 0.652. This result reflects relatively consistent performance across all process groups ($I_{MPs} = 0.631, I_{CPs} = 0.555, I_{SPs} =$

0.771, respectively), with particularly strong outcomes in support processes. The SP1 program achieved a TQI of 0.533. In this case, strong results in MPs ($I_{MPs} = 0.722$) offset by weaker performance in SPs ($I_{SPs} = 0.339$). The SP3 program had the lowest TQI (0.408), mainly due to its relatively low scores in MPs ($I_{MPs} = 0.140$), despite its comparatively better performance in core and support-related indicators ($I_{CPs} = 0.442, I_{SPs} = 0.646$, respectively).

A closer look at the partial indices suggests that the observed differences don't stem from any single process group, but rather from their combined effect. For example, although SP1 performs best in MPs, its lower scores in SPs reduce its overall position. In contrast, SP2 maintains consistently strong results across all dimensions, resulting in the highest overall score. The example is consistent with the assumption that teaching quality may be interpreted as the outcome of interactions between different process groups rather than isolated performance factors. The illustrative results suggest the potential diagnostic value of the TQI for identifying differences between the study programs. The decomposition of the index into partial indices illustrates how the framework could support the identification of program-level areas for improvement when applied to real institutional data. For instance, lower results in SPs may indicate a need to strengthen the infrastructure, digital tools or students' services. At the same time, weaker core process outcomes may point to issues in the curriculum design or in the organization of teaching.

It should be noted that the results are based on an illustrative dataset and are intended primarily to demonstrate the calculation procedure and the framework's logic. Nevertheless, the example demonstrates the computational feasibility of the proposed approach and illustrates how the TQI may be used to support comparison, benchmarking and decision-making at the level of the individual study programs.

V. Conclusions

The proposed TQI contributes to the literature in three main ways. First, it extends the conceptualization of teaching quality by embedding it within a process-based, systems-oriented perspective, demonstrating how teaching quality can be decomposed into management, core, and support processes to form a coherent measurement framework. Second, it operationalizes this perspective by developing a composite index (OECD, 2008) that integrates diverse indicators into a structured and comparable measurement system. Third, it incorporates students' dropout as a structurally embedded component of teaching quality assessment, rather than treating it as an isolated outcome metric. Put together, these contributions provide a direct and comprehensive response to the research questions ($\boldsymbol{RQ1}$-$\boldsymbol{RQ4}$) formulated in this study.

This contribution emerges in response to dominant approaches in the literature, which have primarily concentrated on macro-level performance evaluation of higher education institutions (Kallio *et al.*, 2017). Existing studies have largely focused on developing indicators to assess universities as a whole, typically across research, teaching, and external engagement. Such indicators are widely used for inter-institutional comparisons, forming the basis for rankings and evaluations of universities as public organizations and often influencing national funding allocation mechanisms (Kivipõld *et al.*, 2021). However, as highlighted in the introduction, these approaches provide limited support for internal, program-level decision-making. From a managerial perspective, effective and dynamic governance of teaching requires access to more granular and process-oriented data that reflect the actual functioning of educational processes within specific study programs. This includes not only program management practices, but also the core processes shaping teaching quality and the supporting mechanisms that enable their effective implementation.

The conceptual framework developed in this study reflects the multidimensional nature of teaching quality and should be understood as the outcome of interrelated processes that respond to diverse stakeholders' expectations, particularly those of students. In this context, effective education requires flexibility in adapting to labor market needs, high-quality teaching practices, and well-functioning institutional support systems. Backed by this background, the proposed TQI directly addresses the gap identified in the literature by providing a process-based, program-level composite index that integrates management, core, and support processes within a unified analytical framework, while explicitly incorporating students' dropout as a structurally embedded component of teaching quality assessment.

Several limitations should be considered when interpreting the results of this study. First, the paper focuses on the development and operationalization of a process-based measurement framework rather than on its empirical validation. Consequently, the present study should be regarded as a methodological and conceptual contribution rather than an empirical validation of the proposed framework. The illustrative example demonstrates the computational feasibility of the approach but does not provide evidence regarding its explanatory power, predictive capability or practical effectiveness in real institutional settings. This reflects the study's conceptual and methodological scope. Secondly, the illustrative application of the TQI is based on a hypothetical dataset designed to demonstrate the framework's computational logic. Accordingly, the presented results should not be interpreted as an evidence of the empirical validity of the framework, but rather as an illustration of its methodological structure and computational procedure.

While this approach allows for a clear presentation of normalization, weighting, and aggregation procedures, it does not capture the full complexity and variability of real higher education environments, including potential data quality issues. Consequently, the resulting rankings and index values should be interpreted as illustrative rather than empirical. Finally, the construction of the composite index involves several methodological choices, including the indicator selection, normalization methods, and weighting schemes. Although these decisions are theoretically grounded, they introduce a degree of subjectivity and may affect the resulting TQI values and program rankings. This

points to the need of systematic robustness and sensitivity analysis, which lies beyond the scope of the present study. Despite these limitations, the framework provides a structured and operational basis for further research, particularly for empirical validation, methodological refinement, and practical implementation in higher education institutions.

The proposed framework opens several directions for future research. A priority is the empirical validation of the TQI using real-world institutional data. In particular, it would be important to test how TQI scores relate to the key educational outcomes such as dropout, completion, and graduates' employability. At the same time, applying the framework in smaller institutions may require methodological adjustments, for example, by using longitudinal data or multi-institutional datasets. A second line of research concerns the robustness and sensitivity of the TQI. Given the composite nature of the index, its results may depend on the normalization methods and weighting schemes used. Systematic comparisons of alternative normalization methods and weighting schemes would therefore help assess the stability, robustness, and reliability of the resulting TQI values and program rankings. Finally, the framework can be further extended by developing the underlying indicator system. This may include incorporating new indicators that capture emerging trends in higher education, such as the integration of artificial intelligence in teaching, the expansion of digital learning environments, and the growing role of hybrid education models. Future studies should also examine the adaptability of the framework across different institutional and national contexts, ensuring its scalability and broader applicability.

ACKNOWLEDGMENT

In preparing this manuscript, the authors used ChatGPT (OpenAI) and Grammarly to improve linguistic clarity, grammatical correctness, and overall readability. The authors reviewed and edited all content generated with these tools and take full responsibility for the final version of the publication.

References

Abubakar, A., Hilman, H. and Kaliappen, N. (2018), "New Tools for Measuring Global Academic Performance", *SAGE Open*, SAGE Publications Inc., Vol. 8 No. 3, p. 215824401879078, doi: 10.1177/2158244018790787.

Angiola, N., Bianchi, P. and Damato, L. (2018), "Performance management in public universities: overcoming bureaucracy", *International Journal of Productivity and Performance Management*, Emerald Group Publishing Ltd., Vol. 67 No. 4, pp. 736–753, doi: 10.1108/IJPPM-01-2017-0018.

Asif, M. and Searcy, C. (2014), "A composite index for measuring performance in higher education institutions", *International Journal of Quality and Reliability Management*, Vol. 31 No. 9, pp. 983–1001, doi: 10.1108/IJQRM-02-2013-0023.

Benito, M. and Romera, R. (2011), "Improving quality assessment of composite indicators in university rankings: A case study of French and German universities of excellence", *Scientometrics*, Vol. 89 No. 1, pp. 153–176, doi: 10.1007/s11192-011-0419-5.

Bogt, H.J. and Scapens, R.W. (2012), "Performance Management in Universities: Effects of the Transition to More Quantitative Measurement Systems", *European Accounting Review*, Routledge, Vol. 21 No. 3, pp. 1–47, doi: 10.1080/09638180.2012.668323.

Cadez, S., Dimovski, V. and Zaman Groff, M. (2017), "Research, teaching and performance evaluation in academia: the salience of quality", *Studies in Higher Education*, Routledge, Vol. 42 No. 8, pp. 1455–1473, doi: 10.1080/03075079.2015.1104659.

Chakrabartty, S.N. (2017), "Composite Index: Methods and Properties Composite Index: Methods and Properties Composite Index: Methods and Properties", *Journal of Applied Quantitative Methods*, Vol. 12 No. 2, pp. 31–41.

Chui, T.B., Ahmad, M.S., Bassim, F.A. and Zaimi, N.A. (2016), "Evaluation of Service Quality of Private Higher Education Using Service Improvement Matrix", *Procedia - Social and Behavioral Sciences*, Vol. 224, pp. 132–140.

Dwaikat, N.Y. (2020), "A comprehensive model for assessing the quality in higher education institutions", *TQM Journal*, Vol. 33 No. 4, pp. 841–855, doi: 10.1108/TQM-06-2020-0133.

Fajčíková, A. and Fejfarová, M. (2019), "Evaluation of the quality of teaching from the perspective of university students", *Journal on Efficiency and Responsibility in Education and Science*, Vol. 12 No. 2, pp. 34–40, doi: 10.7160/eriesj.2019.120201.

Feng, X. (2023), "The comprehensive evaluation of innovative education quality from the perspective of balanced and stable development", *International Journal of Innovation and Sustainable Development*, Vol. 1 No. 1, pp. 1-thisLastPage, doi: 10.1504/ijisd.2023.10056677.

El Gibari, S., Gómez, T. and Ruiz, F. (2018), "Evaluating university performance using reference point based composite indicators", *Journal of Informetrics*, Vol. 12 No. 4, pp. 1235–1250, doi: 10.1016/j.joi.2018.10.003.

El Gibari, S., Gómez, T. and Ruiz, F. (2022), "Combining reference point based composite indicators with data envelopment analysis: application to the assessment of universities", *Scientometrics*, Vol. 127 No. 8, pp. 4363–4395, doi: 10.1007/s11192-022-04436-0.

Gnaldi, M. and Ranalli, M.G. (2016), "Measuring University Performance by Means of Composite Indicators: A Robustness Analysis of the Composite Measure Used for the Benchmark of Italian Universities", *Social Indicators Research*, Vol. 129 No. 2, pp. 659–675, doi: 10.1007/s11205-015-1116-1.

Gulden, M., Saltanat, K., Raigul, D., Dauren, T. and Assel, A. (2020), "Quality management of higher education: Innovation approach from perspectives of institutionalism. An exploratory literature review", *Cogent Business and Management*, Cogent OA, Vol. 7 No. 1, pp. 1–21, doi: 10.1080/23311975.2020.1749217.

Hernandez-Diaz, P.M., Polanco, J.A. and Escobar-Sierra, M. (2020), "Building a measurement system of higher education performance: evidence from a Latin-American country", *International Journal of Quality and Reliability Management*, Emerald Group Holdings Ltd., Vol. 38 No. 6, pp. 1278–1300, doi: 10.1108/IJQRM-04-2020-0133.

Kallio, K.M., Kallio, T.J. and Grossi, G. (2017), "Performance measurement in universities: ambiguities in the use of quality versus quantity in performance indicators", *Public Money and Management*, Vol. 37 No. 4, pp. 293–300, doi: 10.1080/09540962.2017.1295735.

Kivipõld, K., Türk, K. and Kivipõld, L. (2021), "Performance appraisal, justice and organizational effectiveness: a comparison between two universities", *International Journal of Productivity and Performance Management*, Vol. 70 No. 1, pp. 87–108, doi: 10.1108/IJPPM-05-2019-0229.

Kohoutek, J., Antonowicz, D. and Kováts, G. (2025), "Higher education quality assurance in Central Europe–Beyond accreditation towards divergence", *Assessment and Evaluation in Higher Education*, Vol. 50 No. 5, pp. 747–759, doi: 10.1080/02602938.2025.2468327.

Kucińska-Landwójtowicz, A., Czabak-Górska, I.D., Lorenc, M., Domingues, P. and Sampaio, P. (2023), "Performance measurement model for technical universities – case study", *International Journal of Quality & Reliability Management*, Vol. 40 No. 10, pp. 2632–2663, doi: 10.1108/IJQRM-04-2021-0107.

Latif, K.F., Latif, I., Farooq Sahibzada, U. and Ullah, M. (2019), "In search of quality: measuring Higher Education Service Quality (HiEduQual)", *Total Quality Management and Business Excellence*, Vol. 30 No. 7–8, pp. 768–791, doi: 10.1080/14783363.2017.1338133.

Li, C. (2026), "Fuzzy comprehensive evaluation of higher education quality based on multi-layer BP neural network", *Progress in Artificial Intelligence*, doi: 10.1007/s13748-026-00433-z.

Li, Y. and Hu, C. (2022), "The Evaluation Index System of Teaching Quality in Colleges and Universities: Based on the CIPP Model", *Mathematical Problems in Engineering*, Vol. 2022, doi: 10.1155/2022/3694744.

Li, Y., Zhang, L., Tian, Y. and Qi, W. (2022), "Research on Teaching Practice of Blended Higher Education Based on Deep Learning Route", edited by Ding, B.*Computational Intelligence and Neuroscience*, Vol. 2022, pp. 1–8, doi: 10.1155/2022/5906335.

Lu, J. and Hao, Z. (2022), "Analysis on Quality Evaluation of Higher Engineering Education from the Perspective of Outcome-based Education", *Journal of Engineering Science and Technology Review*, Vol. 15 No. 4, pp. 96–103, doi: 10.25103/jestr.154.14.

Mao, Y., Wang, F., Zhang, Y. and Chen, Q. (2025), "Teaching Quality Evaluation Model Based on CNN-Framework with Entropy Weight Method", *Computer-Aided Design and Applications*, Vol. 22, pp. 297–312, doi: 10.14733/cadaps.2025.S1.297-312.

Moreno-Gómez, J., Calleja-Blanco, J. and Moreno-Gómez, G. (2020), "Measuring the efficiency of the Colombian higher education system: a two-stage approach", *International Journal of Educational Management*, Vol. 34 No. 4, pp. 794–804, doi: 10.1108/IJEM-07-2019-0236.

Murias, P., de Miguel, J.C. and Rodríguez, D. (2008), "A composite indicator for university quality assesment: The case of Spanish higher education system", *Social Indicators Research*, Springer, Vol. 89 No. 1, pp. 129–146, doi: 10.1007/s11205-007-9226-z.

OECD. (2008), *Handbook on Constructing Composite Indicators: Methodology and User Guide*, *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD Publishing, doi: 10.1787/9789264043466-en.

Râlea, I.-A. (2024), "Quality Assessment in Higher Education Using a Composite Index. Empirical Evidence for Central and Eastern Eu Countries", *Journal of Public Administration, Finance and Law*, Vol. 30, pp. 197–211, doi: 10.47743/jopafl-2023-30-18.

Santiago, R., Carvalho, T. and Cardoso, S. (2015), "Portuguese academics' perceptions of higher education institutions' governance and management: a generational perspective", *Studies in Higher Education*, Vol. 40 No. 8, pp. 1471–1484, doi: 10.1080/03075079.2015.1060709.

Szuwarzyński, A. and Julkowski, B. (2014), "Wykorzystanie wskaźników złożonych i metod nieparametrycznych do oceny i poprawy efektywności funkcjonowania wyższych uczelni technicznych", *Edukacja*, Instytut Badań Edukacyjnych, Vol. 3 No. 128, pp. 54–74.

Williams, R. and de Rassenfosse, G. (2016), "Pitfalls in aggregating performance measures in higher education", *Studies in Higher Education*, Vol. 41 No. 1, pp. 51–62, doi: 10.1080/03075079.2014.914912.

Yuan, M. and Li, C. (2021), "Research on Global Higher Education Quality Based on BP Neural Network and Analytic Hierarchy Process", *Journal of Computer and Communications*, Vol. 09 No. 06, pp. 158–173, doi: 10.4236/jcc.2021.96009.

Zhai, W. and Zheng, H. (2025), "Advanced CoCoSo Technique for Interval Neutrosophic MAGDM and Applications to Quality Evaluation in Higher Education Management", *International Journal of Decision Support System Technology*, Vol. 17 No. 1, pp. 1–23, doi: 10.4018/IJDSST.378431.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Improving the Quality of Patient Post-Discharge Care using a Machine Learning–Based Decision Support System

Matilde M Santos[1], Mariana Peyroteo[1] and Luís Velez Lapão[1]

[1] UNIDEMI, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516, Caparica, Portugal; l.lapao@fct.unl.pt

***Abstract*— Hospital readmissions represent a persistent challenge for healthcare systems, often stemming from inadequate post-discharge monitoring. This study presents a Machine Learning-based Clinical Decision Support System (CDSS) designed to enhance nursing risk assessment in post-discharge care. Developed using a Design Science Research Methodology, the artefact integrates a digital questionnaire, a Machine Learning (ML)-based risk stratification model and a real-time dashboard to optimise follow-up processes. The system was developed at a medical-surgical inpatient unit of a private hospital in Lisbon. A retrospective dataset of 10,134 structured telephone follow-up records, classified by nurses into three risk levels—stable (A), requiring reassessment (B) and clinically concerning (C)—was used to train and evaluate the ML model. Among the evaluated classifiers, Logistic Regression was selected for deployment based on its high specificity (0.9993), precision (0.9879) and absence of critical false negatives. The CDSS enables real-time risk classification based on patient-reported outcomes, supporting timely identification and prioritisation of patients requiring clinical attention. Simulation results indicate a potential reduction of up to 79% in nurse follow-up workload, while preserving care quality by focusing resources on moderate- and high-risk patients. This study demonstrates the feasibility and utility of ML-based dynamic risk analysis for post-discharge monitoring, offering a scalable and interpretable solution that enhances clinical decision-making and resource allocation in healthcare opening the way to digital transformation.**

Keywords – Machine Learning, Post-discharge, Monitoring, Follow-up, Process Improvement, Clinical Decision Support System

I. Introduction

Hospital readmissions continue to represent a significant clinical and economic challenge for healthcare systems worldwide [1,2,3]. Defined as patient re-hospitalisations occurring within 30 days post-discharge [4], they are associated with increased morbidity, lower patient satisfaction and substantial financial burden to health institutions [3, 5]. In Europe, readmission rates range from 5% to 29% [2,3,5,6,7], with similar figures observed in Portugal at 8.8% [1]. The principal causes of hospital readmissions include the progression of chronic conditions, inadequate discharge planning and insufficient post-discharge follow-up [6]. Notably, deficiencies in discharge procedures, such as poor inter-professional communication and absence of structured post-discharge support, are directly linked to increased readmission rates [6]. Evidence supports that identifying high-risk patients before discharge and ensuring coordinated follow-up care can reduce the likelihood of readmissions [7]. Post-discharge monitoring detects early signs of patient deterioration, promotes treatment adherence and provide reassurance during the recovery period [7]. Standard approaches include scheduled follow-up appointments or telephonic follow-up (TFU), typically conducted by nurses [7,9]. TFU has demonstrated positive outcomes in reinforcing discharge instructions, enhancing medication compliance and improving patient satisfaction [9,10,11, 12]. Despite these benefits, TFU still faces multiple operational challenges like invalid contact information, low response rates, delayed discharge notifications and high staff workload, making the process time-consuming and resource-intensive [6,10,11,14]. Such limitations highlight the need for innovative, scalable and automated solutions to support effective post-discharge care.

Risk assessment is a fundamental process in healthcare, encompassing both risk prediction estimating the likelihood of future adverse events [15,16,17], and risk stratification categorising patients based on urgency or case severity to prioritise care [18,19,20,21]. In post-discharge settings, risk prediction is commonly operationalised through tools such as the LACE Index and the HOSPITAL Score, which evaluate clinical parameters to estimate the probability of 30-day readmission [15,16,21]. However, these tools have shown limited performance across diverse populations, with AUROC values ranging from 0.58 to 0.75 [16,17]. Risk stratification is often applied in pre-admission scenarios such as emergency triage [23,24], where tools like the Manchester Triage System support timely decision-making based on presenting symptoms. These methods allow for targeted care delivery but are rarely extended to post-discharge follow-up [22,24].

Machine Learning (ML) have facilitated the development of data-driven models capable of improving both risk prediction and stratification [19,20,25]. ML algorithms such as Random Forest (RF), Logistic Regression (LR) and eXtreme Gardient Boosting (XGBoost) have demonstrated superior performance over traditional models in predicting adverse outcomes including readmissions, particularly when incorporating Electronic Health Record (EHR) data [26,27]. Yet ML adoption in post-discharge settings remains limited. A recent review revealed that most studies rely on static in-hospital data, with minimal integration of post-discharge patient-reported outcomes [27]. ML-based risk stratification has been explored in pre-admission care [21], but its application to post-discharge follow-up intensity remains underdeveloped.

This study aimed at designing, developing and evaluating an ML-driven Clinical Decision Support System (CDSS) to enhance nursing risk assessment in post-discharge care. Addressing limitations of static, discharge-based models, the proposed system incorporates patient-reported outcomes (PROs) collected after discharge to dynamically stratify risk

and guide follow-up necessities. The risk stratification is performed by a supervised ML-model trained to replicate nurse-assigned classifications following telephone follow-up (TFU).

II. Methodology

A. Helathcare Setting

The study was conducted at the Medical-Surgical Inpatient Unit of a private hospital in the Lisbon region. This unit admits adult patients undergoing surgical procedures requiring overnight hospitalisation, as well as those admitted under internal medicine for acute conditions or diagnostic evaluation. It comprises 65 beds with an average occupancy rate of 70%, and an average patient age of 62 years old. The team includes internal medicine physicians, surgeons, and 20 nurses working in shifts, five of whom are responsible for conducting post-discharge TFU calls. The unit follows a structured nurse-led TFU programme, as part of its post-discharge care policy. This setting provided a relevant context for assessing existing limitations and developing a data-driven solution to support clinical decision-making.

B. Database

A retrospective, anonymised clinical dataset was used, derived from the nurse-led TFU programme. Due to institutional restrictions, the dataset is not publicly available. Data were collected by nurses during structured TFU calls and initially comprised 53 variables: eight related to administrative and management processes, 42 reflecting patient-reported health status during the call, and three concerning patient satisfaction. The study cohort included all patients admitted between 1 Jan. 2021 and 31 Dec. 2023 who received at least one documented TFU call. Of the 12,879 records retrieved, those lacking a nurse-assigned classification were excluded, as they represented call attempts rather than completed assessments. This yielded a final dataset of 10,134 valid records, with each entry corresponding to a single TFU call. For ML-model development, only the 42 variables concerning health status were retained. Satisfaction-related variables were excluded due to their collection after the nurse's classification, to avoid introducing post-outcome bias. Administrative and management variables were also removed, resulting in a final modelling dataset with 42 variables per record. The model was developed as a supervised multi-class classifier trained to replicate the nurse-assigned classification. The target variable was a nurse-assigned classification, categorised as A (clinically stable), B (requiring reassessment) and C (clinical risk), with a highly imbalanced distribution: 9987 (A), 59 (B) and 88 (C).

C. Research Team

The study was developed by a multidisciplinary team of five members: three researchers from NOVA University with expertise in Industrial and biomedical engineering and digital health and two clinicians from the hospital, including the unit coordinator and the nurse TFU programme coordinator. This collaboration ensured alignment between technical development and clinical practice in accordance with design science co-design perspective.

D. . Study Design

The research used Design Science Research Methodology (DSRM), a structured approach suited for developing artefacts that address real-world problems, particularly in complex healthcare settings [8]. DSRM comprises six iterative stages: (1) Problem Identification, (2) Definition of Objectives, (3) Design & Development, (4) Demonstration, (5) Evaluation and (6) Communication [8]. Each stage was executed as a research activity & validated by the clinical team.

E. Problem Identification

The objective was to identify the limitations of the current post-discharge follow-up process. The process mapping was developed with BPMN using three complementary methods: observations, multidisciplinary meetings and Exploratory Data Analysis (EDA) of the clinical dataset. A shadowing approach was adopted to observe 21 TFU calls conducted by three different nurses, enabling real-time monitoring of care delivery and workload management with minimal disruption to clinical routines. Additionally, a structured set of meetings with the clinical team mapped the steps involved in the TFU process and to identify operational challenges. EDA of the TFU dataset complemented these findings, allowing identification of operational bottlenecks and inconsistencies in classification and symptom reporting patterns.

F. Objectives for a Solution

Based on the process mapping and insights gained during the previous phase, this activity aimed to define the objectives of the proposed artefact and establish a revised follow-up process. To achieve this, insights from the previous phase (e.g., BPMN) were analysed collaboratively with the clinical team.

G. Design and Development

This phase involved the creation of the three main components of the CDSS artefact: the ML-model, the data acquisition mechanism and the graphical visualisation interface. This stage was developed in close collaboration with the clinical team, with often validations of the CDSS artefact. The ML-Model - A supervised ML model was developed to replicate nurse-assigned classifications during follow-up calls. The complete pipeline includes data preparation, modelling and evaluation.

Two new variables (1)"Had a recent medical appointment or one coming up?" and (2)"Weighted Total of Symptoms" were derived based on insights from the EDA. Before any processing, the dataset was split using a stratified sampling strategy into a training set (80%) and a test set (20%) to preserve class proportions.

During the data processing stage, variables with more than 85% missing data were excluded. For those with clinically meaningful missingness, values were imputed using the SimpleImputer class from sklearn with a constant strategy [26]; the remaining missing values were imputed using the most frequent category, as all variables were categorical.

Encoding of categorical variables was performed using the class OrdinalEncoder for ordinal features and the class OneHotEncoder for nominal ones. All features were standardised using the class StandardScaler.

Feature selection was performed using Recursive Feature Elimination with Cross-validation with LR as the base estimator and the class StratifiedKFold for cross-validation. Data imbalance was addressed by combining the class RandomUnderSampler with Synthetic Minority Over-Sampling Technique. Dimensionality reduction was performed using Principal Components Analysis, retaining components accounting for 95% of cumulative variance. Three models were trained (RF,LR & XGBoost) and tuned using the class GridSearchCV with 5-fold stratified cross validation. Evaluation metrics included accuracy, precision, sensitivity, specificity, F1 score, AUROC and normalised confusion matrices.

Data Acquisition–A digital survey was developed using Google Forms, to replace the initial TFU call and enable automated risk assessment. The questionnaire was co-designed with the clinical team to ensure consistency with the existing TFU protocol. Responses were stored in a Google Sheets database, allowing real-time access by the ML-model.

Graphical Visualisation–A web-based dashboard was developed using Streamlit application to support effective nursing decision-making. The interface enables nursing staff to view risk classifications, prioritise follow-up actions and access patient-reported data in a structured, timely and efficient format.

H. Demonstration

The three system components (ML-model, digital survey and dashboard) were integrated into a functional artefact. Google Sheets was used as a central data repository, enabling real-time synchronization between the survey, model and dashboard. The system leverages the job lib library for model deployment and the Google Sheets API for real-time data retrieval and update.

I. Evaluation

The evaluation assessed whether the developed artefact met the defined objectives. The system's impact was Estimated based on projected reductions in follow-up workload and improved prioritisation of high-risk patients. Feedback from the clinical team during a structured workshop was used to validate alignment with clinical workflow needs.

J. Communication

Communication activities were conducted throughout the research to ensure alignment with clinical needs and promote the dissemination of the findings. Regular meetings with the clinical team supported iterative development and validation. Results were also shared through a multidisciplinary workshop and the submission of publications.

K. Ethical Consideration

Ethical approval was granted by the hospital's Administration and Clinical Board and its Ethics Committee for Health. All data were anonymised before analysis and managed in accordance with the GDPR.

III. RESULTS

A. Problem Identification

The initial phase of the study focused on identifying inefficiencies within the post-discharge follow-up process currently implemented in the Medical-Surgical Inpatient Unit. A combination of direct observation, multidisciplinary meetings and EDA was used to map the workflow (using BPMN) and assess its effectiveness. The mapped process revealed a uniform follow-up approach, where all discharged patients received a telephone call irrespective of their clinical status or perceived risk. Observational data highlighted variability in practice and workload across the nursing team, while meetings with clinical stakeholders confirmed recurrent operational constraints. These included excessive time demands, duplicated tasks and lack of prioritisation criteria. The EDA of the clinical dataset supported these findings, showing that most patients (98.6%) were classified as low risk, suggesting that a large portion of follow-up calls may be of limited clinical value (i.e., contributing to significant inefficiency). Additionally, inconsistencies were identified in the documentation of symptoms and risk classifications, further emphasising the need for standardisation and decision support. Overall, the process was deemed inefficient due to its disproportionate resource consumption and ineffective as it failed to identify and prioritise higher-risk patients. Furthermore, the absence of continuous monitoring mechanisms limited the system's responsiveness to evolving patient needs following discharge and therefore quality.

B. Objectives of the Solution

In response to the identified limitations, two primary objectives were defined by the research and clinical team for the proposed solution:

1. To improve the efficiency of the follow-up process by reducing the time nurses dedicate to follow-up activities.

2. To enhance effectiveness by enabling the timely identification of patients at risk of readmission through targeted, risk-based interventions.

A revised workflow was proposed, integrating a CDSS that uses an ML-driven risk stratification model. In the new process, only patients classified as moderate or high risk (yellow or red) would receive telephone follow-up call, allowing nursing resources to be focused where most clinically needed. Low-risk patients (green) would be monitored digitally, thus reducing manual workload while maintaining continuity and quality of care.

C. Design and Development

ML Risk Stratification Model–Table 1 presents the performance metrics for three models: RF, LR and XGBoost, with the highest values in each category highlighted. XGBoost achieved the best results in accuracy (0.9206), sensitivity (0.9206) and F1 score (0.9490), indicating strong overall performance and effective identification of high-risk patients. In contrast, LR outperformed the other models in precision (0.9879), specificity (0.9994) and AUROC (0.9879), suggesting superior capability in minimizing false positives and enhancing overall discriminative power. While RF showed competitive results, it did not lead in any metric.

TABLE I. PERFORMANCE METRICS FOR EACH MODEL, WITH BEST VALAUES HIGHLITHED IN BLUE.

Model	Accuracy	Precision	Sensitivity	F1 score	Specificity	AUROC
RF	0.9038	0.9873	0.9038	0.9399	0.9665	0.9842
LR	0.9068	0.9879	0.9068	0.9416	0.9994	0.9879
XGBoost	0.9206	0.9863	0.9206	0.9490	0.9010	0.9841

Further insights are provided by the normalized confusion matrices for LR and XGBoost. LR showed higher classification consistency for class B (0.92) and lower misclassification rates overall, particularly for class C, where 83% of cases were correctly classified and 17% misclassified as B. XGBoost demonstrated slightly higher performance for class A (0.92) and class C (0.83), but a lower classification rate for class B (0.75), with 17% of B instances misclassified as class C. These findings highlight the complementary strengths of XGBoost and LR, each excelling in different performance dimensions. XGBoost is particularly suitable for scenarios prioritizing sensitivity, whereas LR may be preferred when specificity and interpretability are critical. As no single model consistently outperforms across all metrics, model selection should be guided by clinical priorities and the specific trade-off required in the healthcare setting.

For the CDSS, the survey was structured into five sections and designed to be user-friendly. It includes mandatory questions identified through feature selection, complemented by free-text fields to provide additional context to the nursing team. The form features brief, plain-language introductions to each section, guiding patients through the process. The survey content aligns directly with the model's selected predictive variables and supports real-time risk classification. The dashboard, co-designed with the clinical team, presents real-time risk predictions and facilitates structured follow-up. Upon login, users access three main interface views. The Home Screen displays each nurse's agenda, including scheduled and overdue calls, with summaries of high- and moderate-risk patients and a progress bar tracking completed calls. The Call Management Screen lists patients requiring follow-up, prioritised by urgency. Nurses can view responses, register calls and assign classifications. Green patients are removed; yellow and red remain until follow-up is complete. After three unanswered attempts, patients can be excluded. The Complete Patient List provides a searchable, filterable history of all follow-up records.

D. Demonstration

The fully functional CDSS was demonstrated through the successful integration of its components: the post-discharge survey, the risk stratification model, the shared database and the clinical dashboard. The digital artefact for post-discharge care elements was connected to enable real-time classification and follow-up decision support.

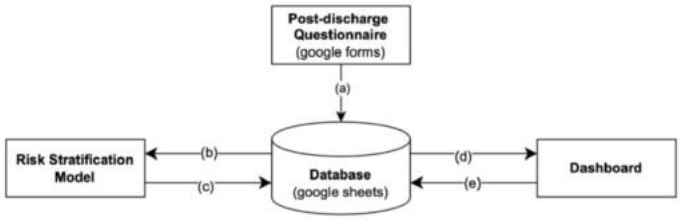


Figure 1: Information Flow in the iFollow@Care artefact.

Figure 1 shows the information flow within the system. Patients initiate the process by completing the post-discharge questionnaire via Google Forms. Their responses are automatically recorded in a centralised database implemented using Google Sheets (step a). The ML-based risk stratification model then retrieves this input (step b), processes the data, and updates the database with a risk classification (step c). The dashboard, accessed by the nursing team, reads the updated classifications from the same database (step d), enabling immediate visualisation of patient status and follow-up priorities. Any actions performed by nurses, such as logging call outcomes, assigning classifications or updating patient information, are written back to the database (step e), ensuring real-time synchronisation.

E. Evaluation

The redesigned follow-up process (in BPMN) introduces a risk-based approach that prioritises only moderate- and high-risk patients (classified as yellow and red) for telephone contact. Patients identified as low risk (green) are automatically excluded from manual follow-up, enabling a substantial reduction in nursing workload. Data collected indicated that 98.6% of patients were classified as green. However, considering evidence from the literature suggesting that 16% of patients may over-report symptoms [28], a conservative estimate was applied, projecting a maximum achievable reduction of approximately 79% in follow-up volume. In operational terms, nurses perform an average of 65 follow-up calls per week, each lasting around seven minutes. This corresponds to a total weekly follow-up time of

approximately 455 minutes (7 hours & 35 minutes). Under the revised process, the expected reduction in workload would lower this to approximately 96 minutes per week (1 hour & 36 minutes). The second objective relates to the potential impact of the artefact on reducing 30-day hospital readmissions. The implementation of a dynamic, risk-based follow-up process, combined with the possibility of extending monitoring to 30 days, aligns with best practices in transitional care and offers a strong foundation for improving long-term outcomes.

IV. Discussion

A. Selection of the best model according to the clinical healthcare setting

The selection of the optimal ML-model was guided by clinical priorities rather than by overall performance metrics alone. In this setting, the primary objective is to avoid misclassifying patients who require follow-up (those assessed as moderate- or high-risk (B or C), as clinically stable (A). These false negatives are the most serious type of classification error, as they result in patients being excluded from the follow-up process, potentially leaving them without the necessary medical care. The confusion matrices show that LR had the strongest performance, avoiding all false negateves involving moderate- and high-risk patients misclassified as low-risk. This guarantees that all patients requiring follow-up would be flagged for clinical review. In contrast, XGBoost misclassified 8% of moderate-risk and 11% of high-risk patients as stable, introducing a significant risk of missed interventions. A lesser concern involves the misclassification of high-risk patients as moderate-risk. Although these patients would still receive follow-up, the delay in intervention could compromise clinical outcomes. In this error category, XGBoost showed stronger performance, misclassifying only 6% of high-risk patients as moderate-risk, compared to 17% under LR. False positives, where stable patients are mistakenly classified as moderate- or high-risk, pose less clinical risk but increase nursing workload due to unnecessary follow-ups. LR generated fewer of these cases, making it more efficient in terms of resource allocation. These findings indicate that while both models have strengths, LR aligns more closely with the clinical requirements. Its ability to avoid the most critical classification errors, combined with its efficiency in minimising unnecessary interventions, makes it the most appropriate model for deployment within the developed CDSS.

B. Comparative Analysis with Existing Models

The selected LR-model was evaluated against ML approaches reported in recent literature, with reference to a recent review on post-discharge solutions with AI [27]. This review summarised a broad range of model performances in post-discharge care, reporting AUROC values between 0.695 and 0.994 across diverse settings. The LR-model developed achieved an AUROC of 0.9879, placing it near the upper bound of reported values. Its recall of 0.9068 also compares favourably, falling within the range of 0.650 to 0.947 observed in the literature. In addition, the model demonstrated exceptional specificity (0.9993) and precision (0.9879), both exceeding the highest values reported in the review (0.913 and 0.951, respectively). These results suggest a low rate of false positives and a high degree of reliability in identifying patients requiring follow-up. Other studies have predominantly focused on models trained with retrospective clinical data extracted from EHR, often limited to information available at the time of discharge [27, 20]. These models, while effective in certain contexts, do not account for changes in a patient's condition during the post-discharge period and rarely incorporate patient-reported outcomes. As a result, their applicability to dynamic follow-up processes is limited. In contrast, the LR model was designed to operate with structured, patient-reported data and developed in collaboration with clinical staff. Its combination of high predictive performance, low complexity and strong clinical interpretability makes it particularly suitable for integration into nurse-led follow-up processes. Compared to models reported in the literature, it offers a competitive and practical solution for post-discharge risk stratification.

C. Comparative Analysis with traditional risk score models

Traditional risk prediction tools such as the LACE Index and HOSPITAL Score have shown limited discriminative performance, with AUROC values typically ranging between 0.58 and 0.75 [15,16]. These models rely on static discharge data and do not account for the patient's evolving condition after hospitalisation [15,16,21]. In contrast, the LR-model developed achieved an AUROC of 0.9879, demonstrating stronger predictive accuracy. Unlike traditional tools, the model incorporates patient-reported outcomes collected post-discharge, enabling dynamic risk assessment. This approach improves performance and aligns more closely with clinical needs by supporting timely, risk-based follow-up.

D. Assessment of the artefact using the Österle's Design Principles

The developed artefact was evaluated through the lens of the four Österle principles — abstraction, originality, justification, and benefits [25], which serve as a theoretical foundation for assessing contributions in design-oriented information systems research. Two hospital physicians, two nurses, an industrial engineer specialised in health information systems, and a digital health expert participated in the assessment. This framework supports a critical discussion of the artefact's broader value and relevance.

E. Limitations of the study

This study presents some limitations that should be acknowledged when interpreting the results. First, the dataset used did not include demographic variables such as age, sex, or socioeconomic status. Although these factors are known to influence post-discharge outcomes, they were not available from the healthcare provider and could not be incorporated into the risk model. Second, the study lacked access to hospital readmission data, which limited the ability to directly assess the system's impact on reducing 30-day readmissions. While the model's performance suggests strong potential for effective follow-up risk stratification, its influence on actual readmission rates remains to be validated. Third, the model evaluation was restricted to three machine learning algorithms: LR, RF and XGBoost. Although these models performed well, future advances in artificial intelligence may offer alternative techniques better suited to this application. Lastly, the system was validated using retrospective real-world data but not yet with real-time clinical data

as on-site implementation is still under development. As such, its practical impact on workflow integration, user experience and clinical outcomes will require further investigation through the next on-site deployment.

V. Conclusions

This study presented the design, development and evaluation of a machine learning-driven CDSS to support nursing risk assessment in post-discharge care. Developed using a DSRM approach, the artefact integrates patient-reported outcomes, an ML-based risk stratification model, and a real-time dashboard to support follow-up risk stratification and resource optimisation. The selected LR-model demonstrated strong predictive performance, outperforming both traditional scores (LACE and HOSPITAL) and several ML-based approaches reported in the literature. Its use of post-discharge data and high specificity and precision render it particularly suitable for real-world deployment in nurse-led follow-up programmes. The system achieved its intended objectives by significantly reducing unnecessary follow-up workload, with an expected optimisation of up to 79%, while supporting the timely identification of high-risk patients. It also promotes digital engagement among healthcare professionals and patients, aligning with broader goals in digital health innovation. While limitations exist, particularly regarding the absence of demographic and readmission data, this work establishes a robust foundation for future use of artificial intelligence in clinical integration and improving workforce efficiency as a digital transformation approach. The proposed system offers a scalable, interpretable and clinically aligned solution for enhancing post-discharge monitoring through data-driven decision support.

Acknowledgment

The authors acknowledge Fundação para a Ciência e a Tecnologia I.P. for its financial support via the project UID/00667: UNIDEMI. MP was supported by FCT (2023.00994.BD).

References

[[1] Aguiar, Rute and Jacinto, Margarida and Tribolet de Abreu, Tiago, 30-Day Readmission in Internal Medicine Wards: Rates and Risk Factors, Internal Medicine vol. 22 no. 2 pp.63-66 2015. doi:10.24950/rspmi.888

[2] Fatima, Samar and Shamim, Sara and Raffat, Simra and Tariq, Muhammad, Hospital readmissions in Internal Medicine Specialty: Frequency, associated factors and outcomes., Pak J Med Sci vol. 37 no. 7 pp. 2008– 2013 2021. doi: 10.12669/pjms.37.7.3575.

[3] Sousa-Pinto, Bernardo and Gomes, Ana Rita and Oliveira, Andreia and Ivo, Carlos and Costa, Gustavo and Ramos, Jo˜ao and Silva, Joel and Carneiro, Maria Carolina and Domingues, Maria Joao and Cunha, Maria Joao and Costa-Pereira, Altamiro Da and Freitas, Alberto, Hospital Readmissions in Portugal over the Last Decade, Acta Med Port vol. 26 no. 6 pp. 711–720 2013. doi: 10.20344/amp.2178.

[4] Allaudeen, Nazima and Vidyarthi, Arpana and Maselli, Judith and Auerbach, Andrew, Redefining readmission risk factors for general medicine patients., J Hosp Med vol. 6 no. 2 pp. 54–60 2011. doi: 10.1002/jhm.805.

[5] Benbassat, J and Taragin, M, Hospital readmissions as a measure of quality of health care: advantages and limitations., Arch Intern Med vol. 160 no. 8 pp. 1074–1081 2000. doi: 10.1001/archinte.160.8.1074.

[6] Harrison, Patricia L. and Hara, Pamela A. and Pope, James E. and Young, Michelle C. and Rula, Elizabeth Y., The Impact of Post-discharge Telephonic Follow-Up on Hospital Readmissions, Population Health Management vol. 14 no. 1 pp. 27-32 2011. doi: 10.1089/pop.2009.0076.

[7] Coppa, Kevin and Kim, Eun Ji and Oppenheim, Michael I and Bock, Kevin R and Conigliaro, Joseph and Hirsch, Jamie S, Examination of Post-discharge Follow-up Appointment Status and 30-Day Readmission., J Gen Intern Med vol. 36 no. 5 pp. 1214–1221 2021. doi: 10.1007/s11606- 020-06569-5.

[8] Lapão, Luís Velez et al., Implementation of Digital Monitoring Services During the COVID-19 Pandemic for Patients With Chronic Diseases: Design Science Approach, J Med Internet Res vol. 23 no. 8 pp. e24181 2021. doi: 10.2196/24181.

[9] Harrison, James D and Auerbach, Andrew D and Quinn, Kathryn and Kynoch, Ellen and Mourad, Michelle, Assessing the impact of nurse post-discharge telephone calls on 30-day hospital readmission rates., J Gen Intern Med vol. 29 no. 11 pp. 1519–1525 2014. doi: 10.1007/s11606-014-2954-2.

[10] Tang, Ning and Fujimoto, Jeffrey and Karliner, Leah, Evaluation of a Primary Care-Based Post-Discharge Phone Call Program: Keeping the Primary Care Practice at the Center of Post-hospitalization Care Transition, Journal of General Internal Medicine vol. 29 no. 11 pp. 1513–1518 2014. doi: 10.1007/s11606-014-2942-6.

[11] Woods, Cindy E and Jones, Rikki and O'Shea, Eilish and Grist, Elizabeth and Wiggers, John and Usher, Kim, Nurse-led postdischarge telephone follow-up calls: A mixed study systematic review., J Clin Nurs vol. 28 no. 19-20 pp. 3386–3399 2019. doi: 10.1111/jocn.14951.

[12] Dudas, Vicky and Bookwalter, Thomas and Kerr, Kathleen M and Pantilat, Steven Z, The impact of follow-up telephone calls to patients after hospitalization, The American Journal of Medicine vol. 111 no. 9 pp. 26–30 2001. doi: 10.1016/S0002-9343(01)00966-4.

[13] M Clari, S Frigerio, F Ricceri, A Pici, R Alvaro, and V Dimonte, Follow-up telephone calls to patients discharged after undergoing orthopaedic surgery: double-blind, randomised controlled trial of efficacy., J Clin Nurs vol. 24 no. 19-20 pp. 2736–2744 2015. doi: 10.1111/jocn.12795.

[14] Anne Dewar and Jan Scott and Janice Muir, Telephone follow-up for day surgery patients: Patient perceptions and nurses' experiences, Journal of PeriAnesthesia Nursing vol. 19 no. 4 pp. 234-241 2004. doi: https://doi.org/10.1016/j.jopan.2004.04.004.

[15] Robinson, Robert and Hudali, Tamer, The HOSPITAL score and LACE index as predictors of 30 day readmission in a retrospective study at a university-affiliated community hospital., PeerJ vol. 5 pp. e3137 2017. doi: 10.7717/peerj.3137.

[16] Miller, William Dwight and Nguyen, Kimngan and Vangala, Sitaramand Dowling, Erin, Clinicians can independently predict 30-day hospital readmissions as well as the LACE index, BMC Health Services Research vol. 18 no. 1 pp. 32 2018. doi: 10.1186/s12913-018-2833-3.

[17] SN Mohsin, A Gapizov, C Ekhator, N Ain, S Ahmad, M Khan, C Barker, M Hussain, J Malineni, A Ramadhan, N Halappa, The Role of Artificial Intelligence in Prediction, Risk Stratification and Personalized Treatment Planning for Congenital Heart Diseases. Cureus vol. 15(8) pp. e44374 2023.

[18] Janes, Holly and Pepe, Margaret S and Gu, Wen, Assessing the value of risk predictions by using risk stratification tables., Ann Intern Med vol. 149 no. 10 pp. 751–760 2008. doi: 10.7326/0003-4819-149-10-200811180-00009.

[19] Simukayi Mutasa and Paul H. Yi, Clinical Artificial Intelligence Applications: Musculoskeletal, Radiologic Clinics of North America vol. 59 no. 6 pp. 1013-1026 2021. doi: https://doi.org/10.1016/j.rcl.2021.07.011.

[20] Huang, Yinan and Talwar, Ashna and Chatterjee, Satabdi and Aparasu, Rajender R., Application of machine learning in predicting hospital readmissions: a scoping review of the literature, BMC Medical Research Methodology vol. 21 no. 1 pp. 96 2021. doi: 10.1186/s12874-021-01284-z.

[21] Y Ning, S Li, M Ong, F Xie, B Chakraborty, D Ting, and N Liu, A novel interpretable machine learning system to generate clinical risk scores: An application for predicting early mortality or unplanned readmission in a retrospective cohort study. PLOS Digit Health vol. 1(6) pp. e0000062 2022.

[22] G Kipourgos, A Tzenalis, C Koutsojannis, and I Hatzilygeroudis, An artificial intelligence based application for triage nurses in emergency department, using the emergency severity index protocol, International Journal of Caring Sciences vol. 15 no. 3 pp. 1764–1772 2022

[23] F Loots, M Smits, K Jenniskens, A Leeuwenberg, Paul Giesen, L Ramerman, R Verheij, A van Zanten, R Venekamp, Predicting sepsis-related mortality and ICU admissions from telephone triage information of patients presenting to out-of-hours GP cooperatives with acute infections: A cohort study of linked routine care databases, Plos one vol. 18 no. 12 pp. e0294557 2023. doi: 10.1371/journal.pone.0294557.

[24] Inokuchi, Ryota and Iwagami, Masao and Sun, Yu and Sakamoto, Ayaka and Tamiya, Nanako, Machine learning models predicting undertriage in telephone triage., Ann Med vol. 54 no. 1 pp. 2990–2997 2022. doi: 10.1080/07853890.2022.2136402.

[25] H Osterle, J Becker, U Frank, T Hess, D Karagiannis, H Krcmar, P Loos, P Mertens, A Oberweis, and EJ Sinz, Memorandum on design-oriented information systems research, European Journal of Information Systems vol. 20 pp. 7–10 2011. doi: 10.1057/ejis.2010.55.

[26] Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V., Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research vol. 12 no. 85 pp. 2825–2830 2011.
[27] M Santos, M Peyroteo, and LV Lapão, AI Powered Post-Discharge Monitoring to Prevent Hospital Readmissions: A Global Scoping Review in Hospital Care, 2024 IEEE International Conference on Engineering, Technology, and Innovation (ICE/ITMC) pp. 1-7 2024. doi: 10.1109/ICE/ITMC61926.2024.10794414.
[28] B Dandachi-FitzGerald, B van Twillert, P van de Sande, Y van Os and R. W.H.M. Ponds, Poor symptom and performance validity in regularly referred Hospital outpatients: Link with standard clinical measures and incentives, Psychiatry Research vol.239 pp.47-53 2016.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Enhancing Value Stream Mapping through Low-Cost Digital Data Collection: An Industrial Case Study in Cork Stopper Production

Miguel Moreira [1], Hermilio Vilarinho [2]

[1] Faculdade de Engenharia, Universidade do Porto; up202308639@edu.fe.up.pt
[2] INEGI, Faculdade de Engenharia, Universidade do Porto; hvilarinho@inegi.up.pt, hcfernandes@fe.up.pt

Abstract*—Value Stream Mapping (VSM) is a widely used Lean tool for identifying waste and improving production flow, yet its practical effectiveness strongly depends on the reliability of operational data. In many traditional manufacturing environments, particularly those with low digital maturity, data collection practices limit process visibility and constrain improvement efforts. This paper presents an industrial case study conducted at a technical cork stopper manufacturing unit, aimed at assessing how investments in data governance, particularly improvements in data reliability, can enhance VSM-based continuous improvement initiatives. An As-Is VSM analysis revealed excessive lead times, high inventory levels, and a predominance of non-value-added activities, with major inefficiencies concentrated in disc production and inventory management. Based on these findings, a set of improvement actions was defined, including inventory optimization, adjustments to production planning, and the introduction of structured digital data collection mechanisms designed to standardize and validate production information. A To-Be VSM scenario was subsequently developed to evaluate the impact of these interventions. The results show a reduction in total lead time of more than 78% and a decrease of approximately 23% in non-value-added activities, primarily driven by reductions in waiting times and inventory-related delays. Improved data reliability supported more accurate process analysis, enhanced traceability, and more robust decision-making throughout the improvement process. The study provides empirical evidence that investments in data governance, through improved data reliability and structured data collection practices, can act as a critical enabler of effective VSM and continuous improvement initiatives. It demonstrates that meaningful operational improvements can be achieved through pragmatic, low-cost solutions without reliance on complex digital infrastructures, particularly in traditional manufacturing contexts.

Keywords – Value Stream Mapping, Lean, continuous improvement, digitalization

I. Introduction

Manufacturing organizations increasingly face the challenge of improving operational performance while maintaining high standards of quality, traceability, and process control. In this context, Quality Engineering and Management provide structured approaches to reduce variability, monitor processes, and support systematic decision-making. A fundamental requirement for the effective application of these approaches is the availability of reliable and consistent operational data that accurately reflects process behaviour across the production system.

Lean Manufacturing has been widely adopted as a comprehensive philosophy for improving efficiency and quality through the systematic elimination of waste. Rooted in the Toyota Production System, Lean emphasizes value creation from the customer's perspective and the continuous pursuit of process improvement (Ohno, 1988; Womack and Jones, 2003). Over the past decades, Lean principles have been successfully applied across a wide range of industrial contexts, contributing to reductions in lead time, inventory levels, and operational costs, while simultaneously supporting quality improvement and process stability.

Among Lean tools, Value Stream Mapping (VSM) is recognized as one of the most effective methods for analyzing and redesigning production systems. By providing a holistic visualization of material and information flows, VSM enables the identification of customer value-added, business value-added, and non-value-added activities throughout the value stream (Rother and Shook, 1999). This structured representation supports the identification of bottlenecks, excessive inventories, and information gaps, forming a robust basis for defining targeted improvement actions aimed at improving flow and overall system performance (Hines and Rich, 1997).

Despite its widespread use, the practical effectiveness of VSM is strongly influenced by the quality of the data used to describe process times, inventories, and information flows. In many manufacturing environments, particularly in small and medium-sized enterprises and traditional industries, production data is still collected through manual or semi-structured methods. These practices often result in inconsistencies, limited traceability, and reduced confidence in performance indicators, constraining the accuracy of process analysis and the sustainability of improvement initiatives (Arnheiter and Maleyeff, 2005).

From a Quality Engineering perspective, such limitations have important implications. Reliable data is essential not only for diagnosing process inefficiencies but also for monitoring variability, supporting standardized work, and evaluating the impact of improvement actions over time. When data collection practices are weak, organizations may struggle to distinguish between actual process behaviour and measurement artefacts, leading to suboptimal decisions and limited learning from improvement efforts.

Recent discussions in manufacturing systems research have highlighted the potential role of digital tools in supporting data collection, traceability, and process monitoring (Ilyer, Sangwan and Dhiraj, 2023; Horsthofer-Rauch, Schumann, Milde, Vernim

and Reinhart, 2022). In this context, improving the reliability and accessibility of operational data has become increasingly relevant for the effective application of Lean and continuous improvement initiatives.

This gap is especially relevant in industries characterized by complex production flows, material variability, and strict quality requirements. In such contexts, the challenge is not necessarily the absence of improvement methodologies, but rather the limited availability of reliable data to support their effective application. Addressing this challenge requires empirical investigation into how improvements in data reliability, frequently linked to early data governance initiatives, can enhance established Lean tools, such as Value Stream Mapping, while also representing a pragmatic first step toward digital transformation without relying on complex or costly technological infrastructures.

In this context, the present paper reports an industrial case study conducted in a technical cork stopper manufacturing unit. The study applies Value Stream Mapping to analyse the production system and identify major sources of waste and inefficiency, while simultaneously improving data collection practices to support more accurate process analysis. An As-Is VSM analysis revealed excessive lead times, high inventory levels, and a predominance of non-value-added activities. Based on these findings, a set of improvement actions was defined and evaluated through the development of a To-Be VSM scenario.

Unlike many recent studies on digitalized Value Stream Mapping that emphasize advanced Industry 4.0 infrastructures and data-rich environments, this study focuses on pragmatic and low-cost improvements in data reliability and traceability adapted to a traditional manufacturing context with low digital maturity. The paper contributes primarily through an empirical and practice-oriented perspective, demonstrating how structured and accessible digital data collection practices can support more robust VSM-based continuous improvement initiatives without requiring complex technological architectures.

The remainder of the paper is organized as follows. Section II reviews the relevant literature. Section III describes the research methodology. Section IV presents the results, followed by a discussion in Section V. Finally, Section VI concludes the paper.

II. Literature Review

Lean Manufacturing, rooted in the Toyota Production System, is widely recognized as a systematic approach to improve operational performance through waste elimination and continuous improvement (Ohno, 1988; Womack and Jones, 2003). From a Quality Engineering and Management perspective, Lean contributes to quality outcomes by promoting process stability, standardized work, and disciplined problem-solving routines, conditions that reduce variability and support consistent performance over time (Womack and Jones, 2003). However, the practical deployment of Lean depends on the organization's ability to observe and measure processes reliably, since improvement cycles require trustworthy operational feedback.

Value Stream Mapping (VSM) is a well-established Lean method for representing end-to-end material and information flows and for distinguishing value-added and non-value-added activities across the value stream (Rother and Shook, 1999). VSM supports a structured diagnosis of current performance ("As-Is") and the design of a future improved state ("To-Be"), typically targeting lead-time reduction, inventory rationalization, and flow improvements (Hines and Rich, 1997). In quality and process improvement contexts, VSM is valued for enabling a shared representation of system performance and for making delays, inventories, and disconnects in information flow visible.

Nevertheless, traditional VSM practices commonly rely on manual data collection and tend to deliver a static "snapshot" of operations. This limitation becomes more pronounced as manufacturing systems become more complex, variant-rich, and dynamic. The dependence on accurate inputs (cycle times, waiting times, WIP levels, batch sizes, information triggers) means that weak or inconsistent data can directly undermine the diagnostic accuracy and credibility of VSM-based decisions. This measurement dependency has long been recognized in the broader improvement literature (e.g., the Lean–Six Sigma integration discussion) as a practical constraint on the robustness of improvement programs (Arnheiter and Maleyeff, 2005). Recent research has re-emphasized this limitation by explicitly studying digitalized VSM and highlighting how manual data acquisition restricts both accuracy and scalability of mapping in contemporary production contexts (Horsthofer-Rauch, Schumann, Milde, Vernim and Reinhart, 2022).

The literature on manufacturing digital transformation increasingly treats Lean maturity and process discipline as important contextual factors shaping how firms digitalize. Rossini, Cifone, Kassem, Costa and Portioli-Staudacher (2021), based on empirical evidence, argue that being lean influences how manufacturing companies shape digital transformation trajectories and identify distinct patterns of transformation. This perspective suggests that Lean and digitalization can be complementary: Lean routines can help clarify which information is needed, where variability is generated, and which decisions require timely and reliable data.

In parallel, systematic reviews map the intersection between Lean and Industry 4.0, highlighting how digital technologies can strengthen process visibility and measurement foundations. Gallo, Cagnetti, Silvestri and Ruggieri (2021) review Industry 4.0 tools in lean production and identify the prominence of data-related technologies within integration narratives. Complementing this, Horsthofer-Rauch, Schumann, Milde, Vernim and Reinhart (2022) review digitalized VSM approaches and discuss how digitally captured production data can reduce the cost and limitations of manual VSM data gathering while also enabling more frequent or dynamic mapping practices. Other review works also identify a strong research stream focused on how Industry 4.0 technologies interact with Lean principles and methods, including opportunities and implementation challenges (Bittencourt, Alves and Leão, 2021; Schumacher, Hall, Bildstein and Bauernhansl, 2023). More broadly, the Lean–Industry 4.0 relationship has generated multiple systematic reviews and research agendas, including work emphasizing key recommendations and future research needs (Saad, Bahador, Bhovar and Zhang, 2024).

At the implementation level, recent empirical work also emphasizes the role of data acquisition and computation in sustaining Lean routines. Ilyer et al. (2023) present digitalization as a tool for the long-term adoption of Lean Manufacturing,

emphasizing that improved data capture and processing capability support ongoing monitoring and improvement. While such approaches may use different technological means depending on context, the underlying principle remains that data quality and availability condition the feasibility and sustainability of Lean and Quality Engineering interventions.

Additionally, recent contributions aim to model data and information flows explicitly within VSM frameworks. Teriete, Böhm, Sai, Erlach and Bauernhansl (2022) propose an event-based framework for digitalization of value stream mapping with a standardized interface between data acquisition and processing, targeting scalable applications and customized reporting. Building on the "VSM 4.0" stream, Bega et al. (2023) extend Value Stream Mapping 4.0 to comprehensively identify data and information flows within manufacturing, reflecting a shift from mapping only physical and basic informational flows towards capturing the logic and technical requirements of modern data landscapes.

Research also increasingly integrates sustainability perspectives with digital VSM. Horsthofer-Rauch et al. (2024) discuss sustainability-integrated value stream mapping approaches, including process mining-based and sustainability-integrated VSM, reflecting broader Quality Engineering concerns with measurement systems that support multi-dimensional performance (e.g., time, waste, and broader value criteria).

While not all manufacturing contexts have rich event-log data available, this research stream reinforces a central methodological point: the reliability and structure of operational data is increasingly a limiting factor (or enabling condition) for credible and scalable mapping and improvement.

The reviewed literature supports three points relevant to Quality Engineering and Management:

(1) Lean Manufacturing provides a robust improvement philosophy that supports process stability and quality outcomes but requires reliable measurement and disciplined feedback loops (Ohno, 1988; Womack and Jones, 2003).

(2) VSM is a powerful improvement method, yet its diagnostic accuracy and usefulness depend strongly on the quality and accessibility of operational data (Rother and Shook, 1999; Hines and Rich, 1997; Arnheiter and Maleyeff, 2005).

(3) Recent research increasingly examines the complementarity between Lean and digitalization, including digitalized VSM approaches and Industry 4.0 integration narratives (Bega et al., 2023; de Oliveira, Santos and Detro, 2025; Gallo, Cagnetti, Silvestri and Ruggieri, 2021; Horsthofer-Rauch, Schumann, Milde, Vernim and Reinhart, 2022; Horsthofer-Rauch et al., 2024; Ilyer, Sangwan and Dhiraj, 2023; Rossini, Cifone, Kassem, Costa and Portioli-Staudacher, 2021; Teriete, Böhm, Sai, Erlach and Bauernhansl, 2022; Teriete, Erlach, Bauernhansl and van der Aalst, 2026; Zangiacomi, Pessot, Fornasiero, Bertetti and Sacco, 2020); however, much of this literature focuses on advanced technological solutions and data-rich environments, with comparatively limited empirical evidence on how incremental and context-adapted improvements in data reliability can support Lean and Quality Engineering initiatives in low-digital-maturity settings.

Taken together, these insights highlight a relevant gap in the literature. While the integration of Lean and digitalization has received increasing attention, comparatively little empirical evidence exists on how incremental, context-adapted improvements in data reliability can support the effective application of established Lean tools, such as Value Stream Mapping, in low-digital-maturity environments. This study addresses this gap by examining the role of data reliability in enabling VSM-based improvement through an industrial case study in a traditional manufacturing setting.

III. METHODOLOGY

A. Research Design

This study adopts a industrial case study design, which is appropriate for investigating complex production systems within their real operational context. Case studies are particularly suitable in Lean Manufacturing and Quality Engineering research, as they allow in-depth analysis of process behaviour, data collection practices, and improvement interventions under realistic organizational and technological constraints.

The methodological approach follows a structured continuous improvement logic, combining diagnosis, intervention design, and performance evaluation. Value Stream Mapping (VSM) was used as the primary analytical framework to assess the production system and identify improvement opportunities. Improvements in data collection practices were implemented as a methodological enabler to support reliable process analysis and evaluation, rather than as an independent technological objective. Fig. 1 summarizes the methodological steps followed by the study.

B. Case Study Context

The case study was conducted at a manufacturing unit dedicated to producing technical cork stoppers for wine and champagne applications. Fig. 2 shows an example of the final product in the analysed manufacturing unit. The production system comprises multiple interconnected stages, including raw material reception, disc production, shank production, gluing, rectification, washing, selection, and packaging. These stages involve both batch and continuous operations, with significant material handling and intermediate storage.

Before the intervention, production data were primarily recorded using manual and spreadsheet-based systems. While operationally functional, these practices resulted in inconsistencies, limited traceability, and reduced confidence in performance indicators, constraining systematic process analysis. This context provided suitable conditions to examine how improved data reliability could enhance VSM-based continuous improvement interventions.

Additionally, intermediate products were stored in multiple locations throughout the factory, resulting in dispersed inventories that were difficult to quantify accurately. The limited visibility of stock levels and movements created additional

challenges for production planning and process monitoring, reinforcing the need for more reliable and structured data collection practices.

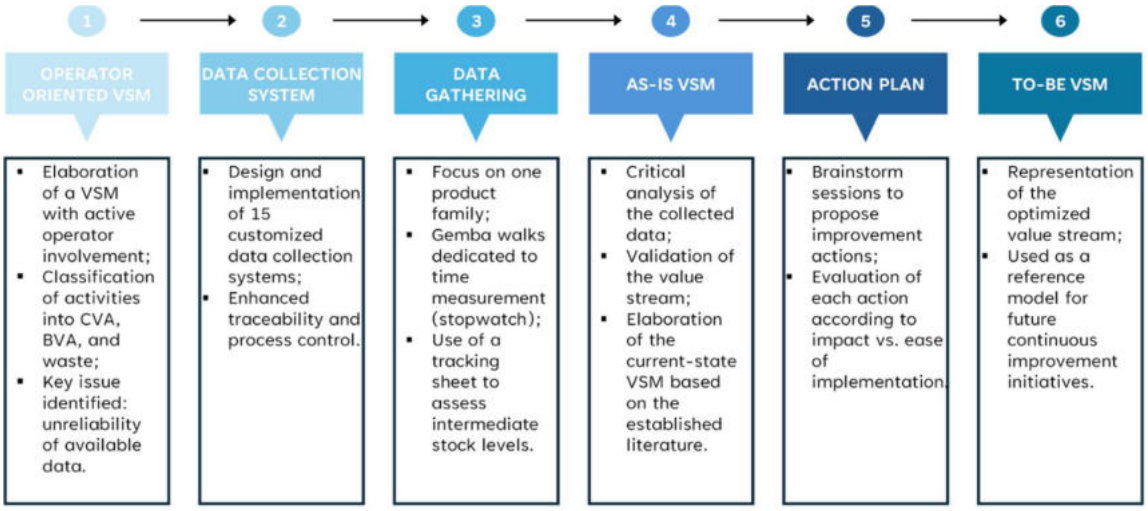


Fig. 1. Summary of the study methodology.

Fig. 2. Technical cork stopper.

C. *Value Stream Mapping Procedure*

Value Stream Mapping was applied following the structured approach proposed by Rother and Shook (1999). A representative product family was selected based on its production volume, strategic relevance within the manufacturing unit, and representativeness of the main production flow analysed in the study. The selected product family corresponded to one of the most operationally significant product streams in the factory, involving multiple interconnected production stages and substantial intermediate inventory flows. This selection ensured that the VSM analysis captured the main operational challenges associated with production planning, material flow, and inventory management within the manufacturing system.

The As-Is VSM captured both material and information flows, including process times, waiting times, inventory levels, batch sizes, and information exchange mechanisms. Activities were classified as customer value-added (CVA), business value-added (BVA), or non-value-added, enabling a quantitative assessment of waste and lead time across the value stream. This analysis supported the identification of bottlenecks and inefficiencies, with particular emphasis on inventory accumulation and waiting times. Lead time was measured in working days required for a production batch to move through the complete value stream, including processing, waiting, transportation, and storage periods. Process times were obtained through direct observation and analysis of production records collected during regular operating conditions. Inventory levels and material flows were assessed through physical verification and comparison with existing production control records.

Based on the As-Is analysis, a To-Be VSM was developed to represent an improved future state. The To-Be scenario incorporated Lean principles such as flow improvement, inventory reduction, and workload balancing, while explicitly considering operational constraints identified during the diagnosis phase.

D. *Data Collection Improvements*

To support reliable VSM analysis and performance evaluation, data collection practices were improved to enhance consistency and traceability. Structured digital data entry mechanisms were implemented to standardize information capture across production stages and reduce common manual recording errors observed during the As-Is analysis.

A VBA-based interface was developed as a low-cost operational data collection tool integrated into the existing production control workflow. The system was implemented using Microsoft Excel and Visual Basic for Applications (VBA), enabling operators and supervisors to register production-related information through standardized digital forms. Input records

included production batch identification, process stage, production quantities, inventory movements, waiting periods, and process completion status. The collected information was automatically consolidated into structured spreadsheets that supported inventory monitoring, process tracking, and performance analysis throughout the value stream.

To improve data reliability and reduce inconsistencies associated with manual records, the interface incorporated predefined input formats, mandatory data fields, dropdown selection menus, and basic validation rules designed to minimize incomplete or incorrect entries. The system also improved traceability by standardizing production records across different production stages and facilitating centralized access to operational information during the VSM analysis and implementation phases.

These mechanisms were designed to be compatible with existing infrastructure and operator practices, focusing on data standardization, validation, and batch identification. The purpose of these improvements was to ensure that performance indicators used in the VSM analysis accurately reflected process behaviour and could be consistently monitored over time.

E. Definition and Implementation of Improvement Actions

Improvement actions were defined based on the waste and inefficiencies identified in the As-Is VSM. These actions focused on operational changes aimed at reducing lead time and non-value-added activities, including adjustments to inventory levels, refinements in production planning parameters, and organizational changes to reduce waiting and unnecessary handling.

Implementation was carried out incrementally to minimize disruption to ongoing operations. Continuous interaction with production staff ensured that proposed changes were feasible and aligned with existing workflows. Improved data collection practices supported the monitoring of implementation progress and facilitated timely adjustments where necessary.

F. Performance Evaluation

The effectiveness of the proposed improvements was evaluated by comparing key performance indicators derived from the As-Is and To-Be VSM scenarios. These indicators included total lead time, proportion of non-value-added activities, and inventory levels. The analysis focused on changes attributable to process redesign rather than technological effects.

In addition to quantitative indicators, qualitative feedback from operators and supervisors was collected to assess perceived improvements in data reliability and process visibility. This combined evaluation approach supports a balanced assessment of how improved data reliability can enable effective VSM-based continuous improvement interventions in low-digital-maturity environments.

IV. RESULTS

A. As-Is Value Stream Analysis

The analysis of the current production system was conducted through the development of an As-Is Value Stream Map, based on direct shop-floor observations, time measurements, and available production records. The resulting VSM revealed a production system characterized by extended lead times, high inventory levels, and a predominance of non-value-added activities. The As-Is Value Stream Map is presented in Annex A.

Quantitative analysis indicated that the majority of the total lead time was not associated with customer value creation. Approximately 93% of the total lead time corresponded to waiting, inventory accumulation, transportation, and other non-value-added activities, while customer value-added activities accounted for less than 1% of total process time, as presented in Fig. 3. These results highlight substantial inefficiencies in flow and a significant potential for performance improvement.

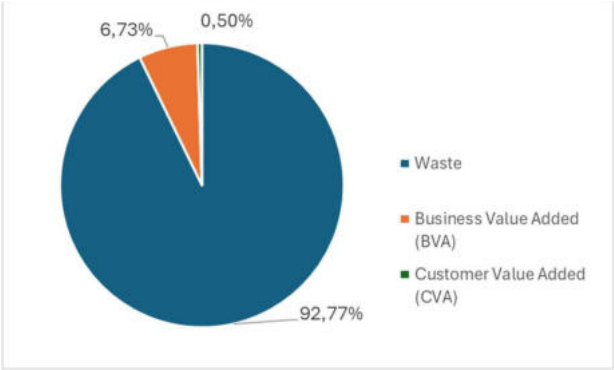


Fig. 3. Lead time distribution in As-Is Value Stream Map.

The As-Is VSM also enabled the identification of critical bottlenecks along the value stream. In particular, the disc production stage emerged as the main source of inefficiency, due to excessive intermediate inventories, long waiting times between operations, and batch-oriented production practices. These factors contributed significantly to extended lead times and limited responsiveness to production requirements. Additionally, limited visibility over process status constrained effective production planning and prioritization. The mapping exercise also highlighted fragmented information flows and limited visibility of inventory levels across the production system, complicating production planning and process synchronization.

B. *Identification of Improvement Opportunities*

Based on the As-Is analysis, improvement opportunities were identified and prioritized according to their expected impact on lead time reduction and flow efficiency. Inventory management emerged as a key leverage point, particularly regarding disc stock levels, which were significantly higher than required to meet downstream production needs.

The analysis also highlighted the need for more consistent and reliable performance information to support improvement decision-making. Inconsistent production records and limited traceability reduced confidence in performance indicators and hindered accurate assessment of process behavior. Addressing these limitations was considered necessary to support robust evaluation of improvement actions, rather than as an end in itself.

Consequently, improvement actions focused primarily on operational changes aimed at reducing waste and improving flow, supported by enhanced consistency in performance data to enable reliable analysis and monitoring.

C. *To-Be Value Stream Mapping and Performance Improvements*

A To-Be Value Stream Map was developed to represent the improved production system incorporating the proposed operational changes. The To-Be scenario included optimized inventory levels, refined production planning parameters, and reduced waiting times between operations, in accordance with Lean principles and operational constraints identified during the As-Is analysis. The To-Be Value Stream Map is presented in Annex B.

The comparative analysis between the As-Is and To-Be VSMs, presented in TABLE I. , demonstrated substantial improvements in operational performance. The total production lead time was reduced by more than 78%, reflecting significant reductions in waiting times and inventory-related delays. In addition, the proportion of non-value-added activities decreased by approximately 23%, indicating a more streamlined and efficient production flow.

Inventory reductions were particularly significant in the disc production stage, where stock levels were adjusted to better align with actual consumption rates. This contributed not only to lead time reduction but also to improved flow stability and reduced accumulation of work-in-progress. The To-Be VSM further showed improved synchronization between production stages, resulting in smoother material flow and enhanced process predictability.

TABLE I. COMPARISON OF RESULTS BETWEEN THE AS-IS AND TO-BE VSM MODELS

Metric	*As-Is*	*To-Be*	*Variation (%)*
CVA (%)	0.50	2.17	+334%
BVA (%)	6.73	26.77	+298%
Waste (%)[a]	92.77	71.06	-23.4%
Lead Time (days)	26.02	5.62	-78.4%

The substantial reduction in lead time was mainly associated with reductions in waiting periods and intermediate inventory accumulation identified during the As-Is analysis. These improvements were achieved without major modifications to core processing operations, reinforcing the importance of flow-oriented interventions in Lean improvement initiatives.

D. *Supporting Role of Improved Data Consistency*

Improved consistency and traceability of production data supported the analysis and evaluation of performance changes observed between the As-Is and To-Be scenarios. Standardized performance records enabled more accurate comparison of key indicators and facilitated ongoing monitoring of process behaviour during implementation.

To support the consistency and traceability of production records, a simple digital data collection solution was implemented using low-cost tools already available within the organization. In particular, customized data entry interfaces were developed in Microsoft Excel using Visual Basic for Applications (VBA), enabling operators to register production batches, process times, and inventory movements in a standardized format. These interfaces incorporated basic validation rules and structured fields to reduce manual recording errors and ensure consistent data entry across production stages. Fig. 4 illustrates an example of the VBA-based interface used for registering production data. Although technologically simple, this solution improved the reliability and accessibility of operational data, supporting more accurate process analysis and facilitating the monitoring of improvement actions during the implementation phase.

Fig. 4. Example of VBA-based interface used by operators and supervisors for standardized registration of production batches, inventory movements, and process progression data during shop-floor operations.

From an operational perspective, improved data reliability increased confidence in performance assessments and supported more informed decision-making during the improvement process. However, the observed performance gains were primarily attributed to process redesign and flow improvements, rather than to technological changes per se.

The reported To-Be performance indicators represent expected operational improvements derived from the Value Stream Mapping analysis and validated operational parameters observed during the implementation process. Due to the industrial scope and implementation timeframe of the study, long-term statistical post-implementation validation was not available at the time of analysis.

E. Summary of Results

Overall, the results demonstrate that the application of Value Stream Mapping led to significant improvements in operational performance, particularly in terms of lead time reduction and waste elimination. The As-Is analysis revealed extensive inefficiencies and limited flow, while the To-Be scenario showed that targeted operational interventions can substantially enhance performance when supported by reliable performance information.

These findings reinforce the role of data reliability as an enabling condition for effective Lean and Quality Engineering interventions, rather than as an independent driver of performance. The case study confirms that meaningful improvements in traditional manufacturing environments can be achieved by prioritizing process redesign and flow optimization, supported by pragmatic improvements in performance measurement.

V. DISCUSSION

The findings of this study provide important insights into the effectiveness of Value Stream Mapping as a continuous improvement tool when supported by reliable performance information. The As-Is analysis revealed a production system characterized by excessive lead times, high inventory levels, and a predominance of non-value-added activities, which is consistent with patterns commonly reported in traditional and batch-oriented manufacturing environments. However, the magnitude of inefficiency observed highlights how limited process visibility and inconsistent performance data can obscure improvement priorities and delay corrective action.

The substantial reduction in lead time achieved in the To-Be scenario confirms the effectiveness of VSM as a diagnostic and design methodology for improving flow and operational performance. The observed reduction of more than 78% in total lead time was primarily driven by reductions in waiting times and inventory-related delays, rather than by changes to core value-adding operations. This finding reinforces established Lean principles, which emphasize flow improvement and waste elimination as primary levers for performance enhancement, and is consistent with prior VSM-based studies reported in the literature.

From a Quality Engineering perspective, the results highlight the central role of reliable performance measurement in supporting systematic improvement. Improved consistency and traceability of operational data enabled a more accurate characterization of process behaviour, facilitating both the development of the To-Be VSM and the evaluation of improvement outcomes. In this context, improved data reliability contributed primarily to strengthening the accuracy and consistency of performance assessment, thereby increasing confidence in the identification and evaluation of improvement opportunities throughout the VSM process.

The case study also underscores the importance of organizational context in the success of Lean and Quality Engineering initiatives. The involvement of operators and supervisors during data collection, mapping, and implementation phases contributed to the feasibility and acceptance of improvement actions. This supports previous findings that employee engagement and contextual adaptation are critical for sustaining improvement gains, particularly in traditional manufacturing settings where standardized routines and tacit knowledge play a significant role.

VI. Conclusion

This paper investigated the application of Value Stream Mapping as a continuous improvement tool in a traditional manufacturing environment, with a particular focus on the role of reliable performance information in supporting effective improvement initiatives. Through an industrial case study conducted in a technical cork stopper manufacturing unit, the study demonstrated how systematic process analysis and redesign can lead to substantial operational performance improvements.

The As-Is Value Stream Mapping analysis revealed a production system characterized by excessive lead times, high inventory levels, and a predominance of non-value-added activities, highlighting significant inefficiencies in material flow and process coordination. By applying VSM to structure the diagnosis and guide the design of improvement actions, it was possible to identify key leverage points related primarily to inventory management, waiting times, and production flow.

The To-Be scenario showed that targeted operational interventions, grounded in Lean principles, resulted in a lead time reduction exceeding 78% and a significant decrease in non-value-added activities. These improvements were achieved without altering core value-adding processes, underscoring the effectiveness of flow-oriented redesign and waste reduction as primary drivers of performance enhancement in Quality Engineering contexts.

An important contribution of this study lies in highlighting data reliability as an enabling condition for effective VSM-based improvement rather than as an independent technological objective. Improved consistency and traceability of performance information strengthened process analysis, supported evidence-based decision-making, and increased confidence in the evaluation of improvement outcomes. From this perspective, investments in basic data governance practices, such as structured data collection, validation, and traceability, can play a critical role in enabling robust Lean and Quality Engineering initiatives.

From an academic perspective, the paper contributes to the literature by providing empirical evidence on how pragmatic and low-cost improvements in data reliability and traceability can support VSM-based continuous improvement initiatives in traditional manufacturing environments. Rather than proposing a new VSM methodology or advanced digital framework, the study contributes through an applied and context-sensitive demonstration of how structured data collection practices can enhance process visibility and support operational improvement in low-digital-maturity settings.

The study is subject to certain limitations that should be acknowledged. First, the findings are based on a single case study conducted in a traditional cork stopper manufacturing environment, which may limit direct generalisation to industries with substantially different production characteristics, product complexity, or automation levels. The observed operational challenges and improvement opportunities are therefore partly influenced by the specific material flow structure and inventory management practices of the analysed context. Second, although the proposed VBA-based data collection approach proved effective within the scope of the present study, its scalability may be limited in larger, highly automated, or data-intensive production systems that require real-time integration with enterprise-level information systems or advanced sensing infrastructures. Finally, as with many industrial case studies, potential sources of bias may exist due to researcher involvement during the implementation, data collection, and analysis phases. Although efforts were made to maintain consistency and triangulate operational information through multiple data sources and shop-floor observations, some interpretation bias may still have influenced the assessment of improvement opportunities and reported outcomes.

Overall, the findings suggest that Value Stream Mapping remains a relevant and effective continuous improvement tool when supported by reliable and well-structured operational data. By emphasizing pragmatic improvements in data reliability, traceability, and process visibility, the study contributes to a more practically grounded understanding of continuous improvement in traditional manufacturing environments.

References

Arnheiter, E.D. and Maleyeff, J. (2005), "The integration of lean management and Six Sigma", The TQM Magazine, Vol. 17, No. 1, pp. 5–18.

Bega, M., Sapel, P., Ercan, F., Schramm, T., Spitz, M., Kuhlenkötter, B. and Hopmann, C. (2023). "Extension of value stream mapping 4.0 for comprehensive identification of data and information flows within the manufacturing domain", Production Engineering, Vol. 17, No. 6, pp.915-927.

Bittencourt, V.L., Alves, A.C. and Leão, C.P. (2021), "Industry 4.0 triggered by Lean Thinking: insights from a systematic literature review", International Journal of Production Research, Vol. 59, No.5, pp.1496-1510.

de Oliveira, J.V.G., Santos, E.A.P. and Detro, S.P. (2025), "Uncovering the potential and pitfalls of process mining in manufacturing", Procedia CIRP, Vol. 132, pp.19-24.

Gallo, T., Cagnetti, C., Silvestri, C. and Ruggieri, A. (2021), "Industry 4.0 tools in lean production: A systematic literature", Procedia Computer Science, Vol. 180, pp. 394–403.

Hines, P. and Rich, N. (1997), "The seven value stream mapping tools", International Journal of Operations & Production Management, Vol. 17, No. 1, pp. 46–64.

Horsthofer-Rauch, J., Guesken, S.R., Weich, J., Rauch, A., Bittner, M., Schulz, J. and Zaeh, M.F. (2024), "Sustainability-integrated value stream mapping with process mining", Production & Manufacturing Research, Vol. 12, No. 1, pp. 2334294.

Horsthofer-Rauch, J., Schumann, M., Milde, M., Vernim, S. and Reinhart, G. (2022), "Digitalized value stream mapping: Review and outlook", Procedia CIRP, Vol. 112, pp. 244–249.

Ilyer, S.V., Sangwan, K.S. and Dhiraj (2023), "Digitalization: a tool for the successful long-term adoption of lean manufacturing", Procedia CIRP, Vol. 116, pp. 245–250.

Ohno, T. (1988), Toyota Production System: Beyond Large-Scale Production, Productivity Press, Portland, OR.

Rossini, M., Cifone, F.D., Kassem, B., Costa, F. and Portioli-Staudacher, A. (2021), "Being lean: how to shape digital transformation in the manufacturing sector", Journal of Manufacturing Technology Management, Vol. 32, No. 9, pp. 239–259.

Rother, M. and Shook, J. (1999), Learning to See: Value Stream Mapping to Add Value and Eliminate Muda, Lean Enterprise Institute, Brookline, MA.

Saad, S.M., Bahadori, R., Bhovar, C. and Zhang, H. (2024), "Industry 4.0 and Lean Manufacturing–a systematic review of the state-of-the-art literature and key recommendations for future research", International Journal of Lean Six Sigma, Vol. 15, No. 5, pp. 997–1024.

Schumacher, S., Hall, R., Bildstein, A. and Bauernhansl, T. (2023), "Lean Production Systems 4.0: systematic literature review and field study on the digital transformation of lean methods and tools", International Journal of Production Research, Vol. 61, No. 24, pp. 8751–8773.

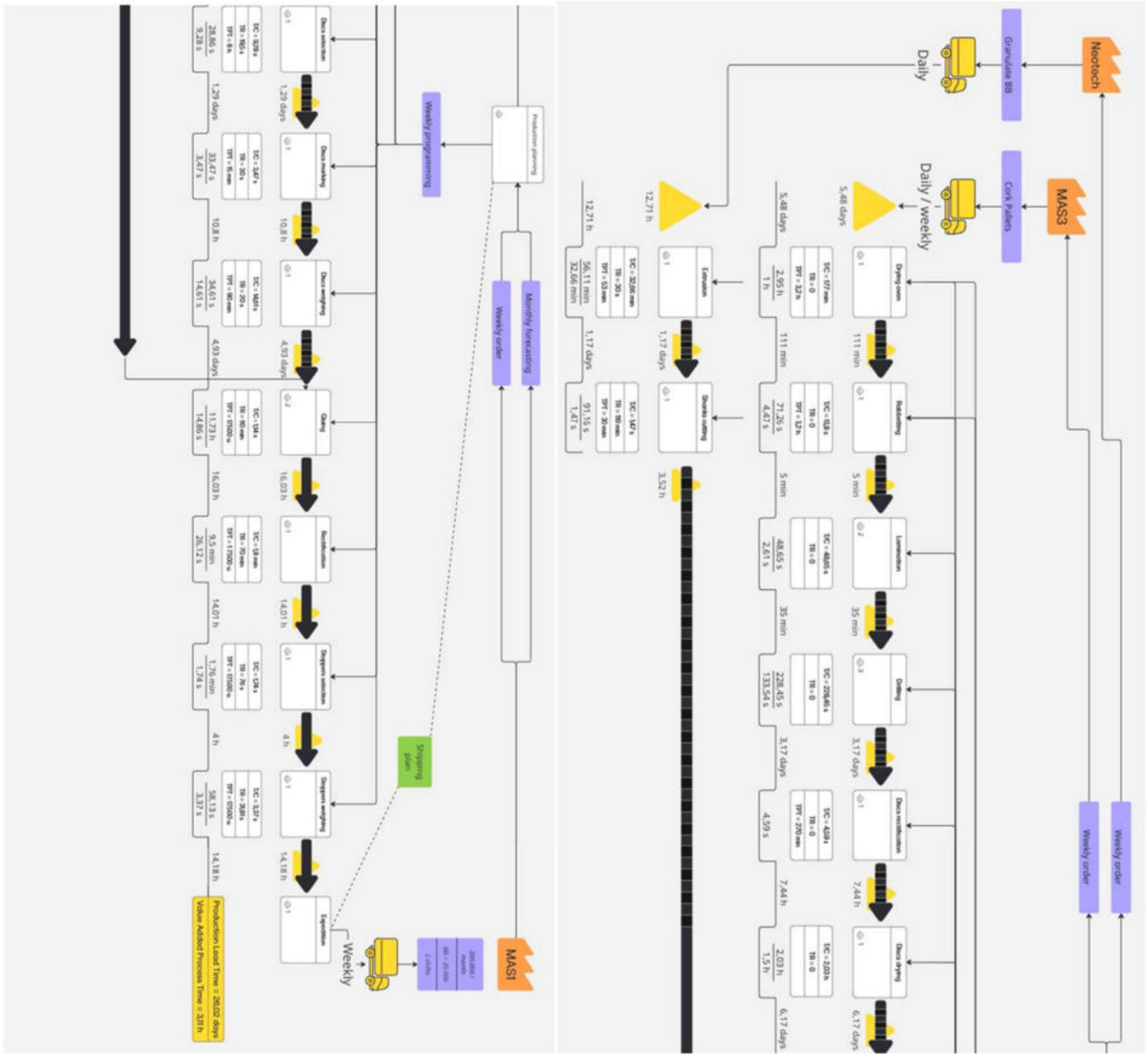
Teriete, T., Böhm, M., Sai, B.K., Erlach, K. and Bauernhansl, T. (2022), "Event-based framework for digitalization of value stream mapping", Procedia CIRP, Vol. 107, pp. 481–486.

Teriete, T., Erlach, K., Bauernhansl, T. and van der Aalst, W.M. (2026), "Combining value stream mapping and process mining in production: a systematic literature review", Discover Applied Sciences.

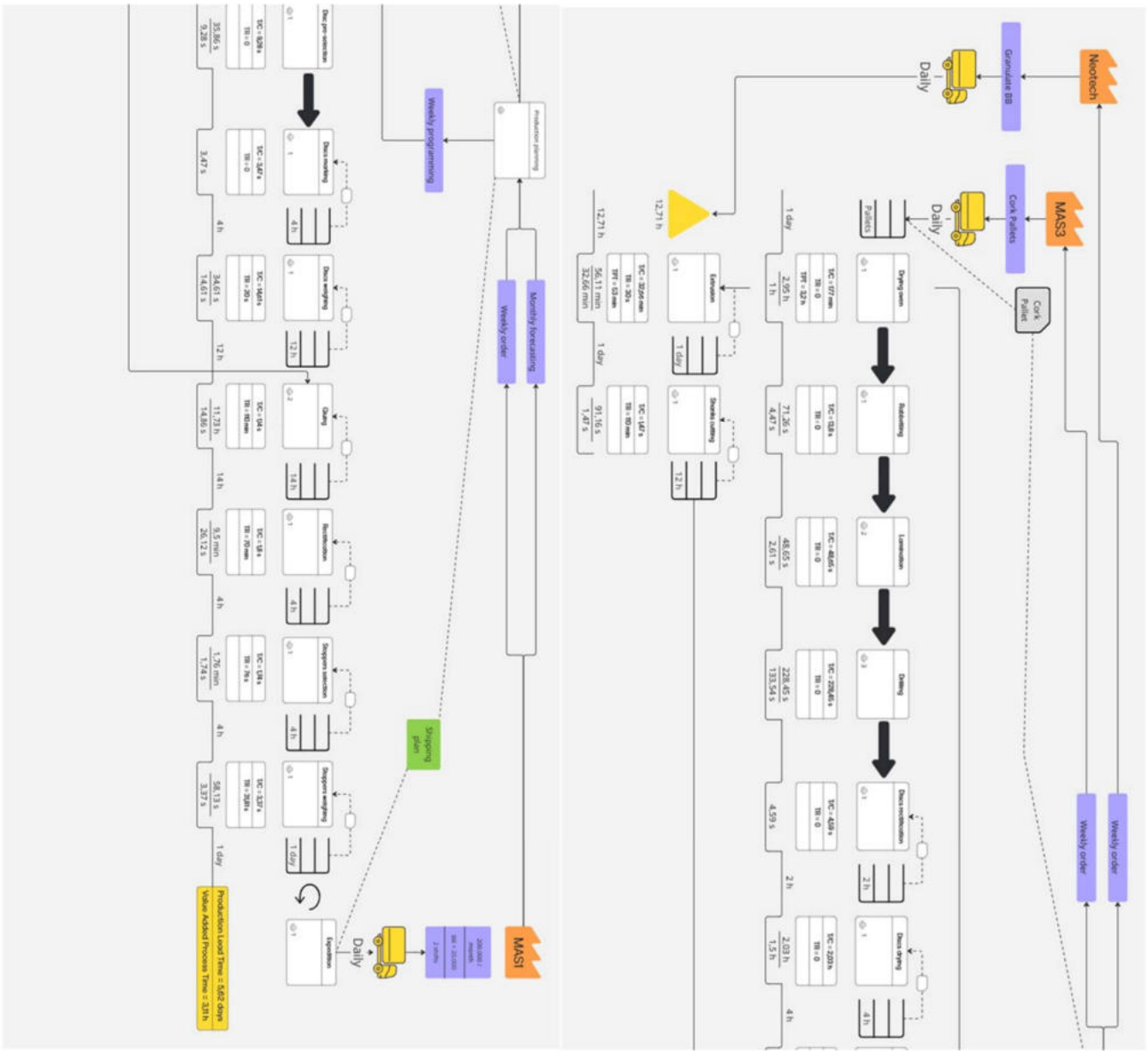
Womack, J.P. and Jones, D.T. (2003), Lean Thinking: Banish Waste and Create Wealth in Your Corporation, 2nd ed., Free Press, New York, NY.

Zangiacomi, A., Pessot, E., Fornasiero, R., Bertetti, M. and Sacco, M. (2020), "Moving towards digitalization: a multiple case study in manufacturing", Production Planning & Control, Vol. 31, No. 2–3, pp. 143–157.

A. Value Stream Map (As-Is)



B. Value Stream Map (To-Be)



7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Subsystem-oriented condition monitoring for connected vehicles using unlabelled data

Stefano Puttero [1], Elisa Verna [1], Gianfranco Genta [1] and Maurizio Galetto [1]

[1] Politecnico di Torino, Department of Management and Production Engineering, Corso Duca degli Abruzzi 24, 10129 Torino, Italy; stefano.puttero@polito.it

Abstract — Connected vehicles continuously generate large volumes of telemetry data, but it remains difficult to transform these streams into useful maintenance support when failure labels are scarce, degradation is gradual and heterogeneous, and operators require stable, interpretable and operationally manageable outputs. This paper presents a subsystem-oriented monitoring architecture for electric vehicles that is designed for this setting. Instead of collapsing the vehicle into a single global health score, the proposed approach monitors four key subsystems (battery, drivetrain, brakes and tyres/suspension) and computes for each of them a Subsystem Condition Indicator (SCI) referenced to subsystem-specific baseline behaviour. The SCI combines structural deviation, derived from reconstruction error, with short-term temporal deviation, derived from next-step forecasting error. It is then converted into stable alert states through fixed, baseline-derived thresholds, persistence logic and hysteresis. Thus, the architecture transforms unlabelled telemetry data into dashboard-oriented outputs that facilitate day-to-day monitoring, including subsystem condition, short-term trends, alarm states and concise reason-code-based cues to support initial inspections. A case study based on an electric-vehicle telemetry dataset shows that the proposed architecture preserves subsystem differentiation, produces interpretable and manageable alert behaviour, and remains stable under alternative baseline definitions and moderate missing-data perturbations, while also providing a basis for future extensions toward more prescriptive maintenance support.

Keywords – Maintenance, Condition monitoring, Fleet management, Connected vehicles

I. Introduction

The growing number of connected vehicles is expanding the scope of data-driven maintenance and Prognostics and Health Management (PHM) strategies based on operational data. Instead of being triggered mainly by inspections, complaints or failures, maintenance can now be supported by continuous telemetry streams generated during vehicle operation (Hu *et al.*, 2022). Modern Electric Vehicles (EVs) collect electrical, thermal, mechanical and contextual data over long operating periods, creating favourable conditions for data-driven condition monitoring. In practice, however, many organisations still struggle to convert this telemetry into maintenance decisions that are timely, trustworthy and economically meaningful (Mahale *et al.*, 2025). The problem is often not the lack of data itself, but the lack of monitoring outputs that align with the way maintenance decisions are made (Theissler *et al.*, 2021).

This challenge is particularly evident in fleet-based business models. Shared mobility operators, leasing companies, corporate fleets and service providers are directly exposed to the operational and economic consequences of unexpected vehicle degradation, downtime and unplanned maintenance (Sharma, 2024). They therefore require tools that can identify which vehicles require attention first, which subsystems are behaving abnormally and whether the condition observed is stable enough to justify inspection or intervention (Benhanifia *et al.*, 2025). An exclusively diagnostic approach based on isolated threshold violations can result in numerous unrelated alerts, whereas a purely prognostic approach based on time-to-failure or Remaining Useful Life (RUL) can be challenging to implement when the historical data required for supervised modelling is limited, inconsistent or unavailable (Zonta *et al.*, 2020). Reviews of predictive maintenance in the automotive sector explicitly identify data availability and quality as key challenges. Recent fleet-oriented research has also highlighted the need for unsupervised monitoring due to the frequent absence of accurate labels (Theissler *et al.*, 2021).

Electric vehicles highlight the tension between data richness and label scarcity. The main subsystems of these vehicles evolve under operating conditions that are strongly context-dependent, and their degradation patterns are often non-linear and asynchronous (Severson *et al.*, 2019). Battery-related variables, drivetrain signals, braking behaviour and tyre-related measurements are all influenced by usage conditions, load and the environment. Consequently, the same absolute sensor value can have different meanings in different contexts, limiting the usefulness of fixed thresholds and motivating monitoring strategies based on deviations from learned reference behaviour. The need for broader, context-aware, data-driven monitoring is consistent with recent work on AI-enabled EV management and unsupervised fleet maintenance approaches that explicitly model operational context (Swaroopa Rani and Jatoth, 2025).

Another practical issue concerns the level of aggregation. Vehicle-level condition scores are appealing because they provide a concise overview and support high-level reporting. However, for maintenance engineers and fleet coordinators, a single score can obscure the most relevant information for taking action (Theissler *et al.*, 2021). When a vehicle starts to deteriorate, the first operational question is rarely whether the vehicle as a whole is healthy. More useful questions are which subsystem deserves attention, how quickly its condition is changing, and whether the signal is persistent enough to justify action. For this reason, a monitoring architecture that aggregates data too much may reduce the practical value of the available telemetry. While existing work on fleet-level unsupervised predictive maintenance demonstrates the importance

of asset-level ranking and anomaly detection (Mahale *et al.*, 2025), there is still scope for solutions that provide clearer, more subsystem-oriented support for maintenance triage.

This paper addresses these issues by redefining the problem as the generation of stable and interpretable subsystem signals from unlabelled telemetry. The objective is not to introduce a radically new anomaly-detection algorithm or to perform traditional supervised prognostics. Instead, it proposes integrating standard learning-based modelling blocks, trained on baseline data representative of nominal behaviour, into a hybrid monitoring architecture that combines learned anomaly representations, statistical deviation analysis, and operational alarm governance logic. This architecture could form the analytical core of a dashboard for connected vehicle maintenance and support subsequent operational monitoring. In PHM terms, this work is more closely aligned with operational health monitoring and maintenance decision support than with label-intensive failure prediction (Hu *et al.*, 2022). For this reason, the proposed approach is built around four practical requirements: subsystem specificity, baseline-referenced interpretation, stable alarm governance and concise, drill-down information for inspection prioritisation. Each major subsystem is monitored via a dedicated Subsystem Condition Indicator (SCI), which detects structural deviations in multivariate sensor states and temporal deviations in their short-term evolution. These signals are then smoothed, thresholded, and translated into alarm states that are suitable for operational interpretation.

The remainder of the paper is organised as follows. Section II explains the research gap and clarifies why a subsystem-oriented operational approach is necessary. Section III presents the monitoring architecture, emphasising the design choices that enable deployment. Section IV describes the case study and the implementation logic used to test the architecture with electric-vehicle telemetry data. Section V discusses the main results and their implications for maintenance and fleet management. Section VI concludes the paper and outlines the next steps for multi-vehicle validation.

II. Research gap and motivation

Recent literature on predictive maintenance in the automotive sector reveals a growing interest in data-driven diagnostics, fault detection, and RUL estimation (Mahale *et al.*, 2025). However, much of this literature still relies on supervised learning settings where labelled fault events, degradation stages or run-to-failure histories are assumed to be available (Serradilla *et al.*, 2022). In practice, this assumption is often difficult to fulfil in connected-vehicle applications, where maintenance records are frequently incomplete, inconsistent, or misaligned with the actual progression of subsystem degradation (Hu *et al.*, 2022). A survey focusing on machine learning–enabled predictive maintenance in the automotive sector explicitly discusses data availability and quality as key challenges for deployment. More recent works on RUL estimation are increasingly motivated by the need for self-supervised or label-efficient solutions, precisely because labelled degradation data are limited (Theissler *et al.*, 2021).

This limitation is particularly relevant in vehicle settings with a high volume of telemetry data. Connected vehicles generate large volumes of multivariate time-series data, but these streams are not automatically accompanied by reliable labels that can support supervised prognostics (Swaroopa Rani and Jatoth, 2025). Even when maintenance records exist, they do not necessarily indicate the precise moment of failure onset, nor do they always depict the entire process from normal behaviour to degradation and failure. Consequently, establishing a robust supervised pipeline for fault classification or RUL prediction can be impractical, particularly when the available data covers only one or a few vehicles or when there are no explicit fault annotations (Lipu *et al.*, 2018). In such cases, methods that learn nominal behaviour directly from unlabelled operational data are a more practical option. This approach is consistent with recent work on unsupervised predictive maintenance for fleet management, where anomaly detection is used to address the issue of scarce labels and highly variable operating contexts (Giannoulidis and Gounaris, 2023).

At the same time, shifting from supervised prognostics to unsupervised monitoring does not automatically solve the deployment problem. While existing unsupervised monitoring and anomaly-detection approaches are useful when labelled failures are scarce, they often remain centred on deviation detection, generic anomaly scores, or asset-level ranking (Serradilla *et al.*, 2022). In practice, such outputs can be challenging to act upon as they do not always indicate which subsystem requires attention, whether the deviation is significant enough to warrant inspection and how alerts should be managed. Therefore, maintenance personnel need more than just knowledge of unusual behaviour; they require outputs with stable semantics, understandable scales, and clear alert conditions. Recent work on explainable anomaly detection for predictive maintenance explicitly acknowledges that limited interpretability and reduced user trust are significant barriers to adoption. This suggests that the practical value of an unsupervised monitoring method depends on two things: its ability to detect deviations and its ability to translate these deviations into condition signals that can be consistently interpreted over time (Resanovic *et al.*, 2026).

A second gap concerns the level at which monitoring results are presented (Benhanifia *et al.*, 2025). While many existing approaches provide asset-level anomaly outputs, which can be useful for generic ranking or screening, they are often insufficient for making maintenance decisions in vehicles. In practice, inspections and interventions are usually organised around subsystems and components, such as the battery, drivetrain, braking system, or tyre and suspension systems. While a single vehicle-level score may indicate that something abnormal is happening, it does not necessarily provide the required level of detail to support triage, prioritisation and follow-up inspection (Mahale *et al.*, 2025). While the literature on fleet-oriented unsupervised maintenance recognises the importance of context-aware and streaming monitoring, there is still scope for solutions that connect anomaly detection more directly to operational actions at the subsystem level (Giannoulidis and Gounaris, 2023).

A third gap relates to the robustness of real telemetry streams. Reports from predictive maintenance reviews consistently indicate that real-world data is affected by noise, missing values, variability across operating conditions and other quality

issues that complicate operational deployment (Taylor *et al.*, 2016). This is particularly pertinent in the context of connected vehicles, where measurements are influenced by usage patterns, ambient conditions and the occasional loss of data. In this context, a monitoring architecture must be able to detect deviations, remain stable enough to avoid oscillating alerts, and be interpretable enough to preserve operator trust. Therefore, calibration, alert persistence, and alarm-clearing logic are integral to the monitoring process itself (Nunes *et al.*, 2023).

In light of these considerations, there is still a gap between unsupervised telemetry analytics and monitoring solutions that can be effectively used in the maintenance of connected vehicles. While recent studies have demonstrated the value of label-free and context-aware monitoring (Lipu *et al.*, 2018), less attention has been given to structuring such outputs for stable operational interpretation. This is particularly important when maintenance decisions require subsystem-level visibility rather than generic vehicle-level anomaly evidence. This is especially critical in situations where reliable labels are unavailable, and monitoring outputs must support triage, inspection planning, and intervention prioritisation. Against this background, this paper proposes a monitoring architecture for subsystems based on unlabelled telemetry. Instead of introducing a new anomaly-detection algorithm, it integrates established unsupervised learning components with baseline-referenced statistical monitoring and alarm governance logic. In this architecture, deviation signals are consolidated into the SCI indicator, which is designed for operational use. The work thus aims to translate anomaly evidence into a more structured and actionable representation to support maintenance.

III. Subsystem Condition Indicators

The proposed monitoring architecture is organised into four subsystem groups reflecting both engineering structure and maintenance relevance: battery, drivetrain, brakes, and tyres/suspension. This grouping is intentionally simple and transparent, ensuring that the analytical outputs remain aligned with typical maintenance planning and execution (Zonta *et al.*, 2020). Instead of monitoring the vehicle as a single, aggregated asset, the architecture preserves visibility at the subsystem level, which is more consistent with inspection planning, prioritisation, and maintenance triage. In addition to signals specific to each subsystem, the monitoring logic incorporates contextual variables such as ambient temperature, humidity, speed, load, idle time, distance travelled, and route roughness. Accordingly, the objective is not to flag demanding but legitimate usage, but to identify behaviour that is unusual relative to the operating context.

Fig. 1 summarises the overall workflow of the proposed subsystem-oriented monitoring architecture. Starting with EV telemetry data, the pipeline first groups the available variables by subsystem, before extracting two complementary forms of deviation evidence: a static component obtained from autoencoder (AE) reconstruction error and a dynamic component obtained from Gated Recurrent Unit (GRU)-based next-step forecasting error. These two signals are then fused and normalised with respect to subsystem-specific baseline behaviour to obtain a bounded SCI. The SCI is then translated into stable alert states via the alert governance logic and supplemented with inspection-support outputs.

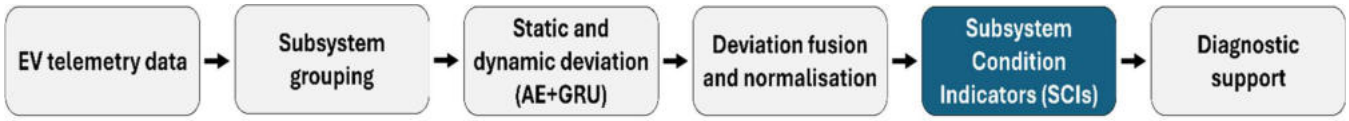


Fig. 1. Subsystem-oriented operational monitoring workflow.

To clarify the learning-based nature of the proposed approach, the monitoring architecture has been organised into two distinct stages: an offline model training stage, and an online operational monitoring stage. During the initial stage, the AE and GRU models are trained using baseline data that is representative of the nominal behaviour of the subsystem. The AE learns compact representations of nominal multivariate patterns and produces reconstruction errors when incoming observations deviate from this learned behaviour. Similarly, the GRU model learns the temporal evolution of nominal signals and provides forecasting errors when observed dynamics depart from expected temporal patterns. Once trained, these models are deployed for operational monitoring in the second stage, where reconstruction and forecasting errors are computed on new telemetry streams and converted into condition indicators and alerts oriented towards the subsystem.

Therefore, the framework should be interpreted as a hybrid monitoring architecture rather than a purely learning-driven anomaly detection model. Learning-based components provide compact representations of nominal multivariate and temporal behaviour. Meanwhile, statistical monitoring and alarm governance components transform the resulting deviation measures into stable, interpretable, subsystem-level condition indicators. Table I summarises the functional decomposition, distinguishing the learning-based components from the statistical monitoring and operational alarm governance logic.

TABLE I. Functional Role of the Main Components in the Proposed Hybrid Monitoring Architecture.

Component	*Role in the framework*	*Type of logic*
AE model	Learns nominal multivariate relationships and produces reconstruction-error deviations	Learning-based
GRU model	Learns short-term temporal dynamics and produces forecasting-error deviations	Learning-based
Error calibration and robust scaling	Convert model errors into comparable subsystem-level deviation scores	Statistical monitoring
Smoothing and thresholds	Reduce short-term fluctuations and support stable interpretation of deviations	Statistical monitoring

Component	*Role in the framework*	*Type of logic*
Alarm governance	Translates deviation patterns into operational alerts and inspection-oriented information	Heuristic/ operational logic

In the case study, the same modelling logic was applied to each subsystem independently. Prior to training, the selected subsystem variables were robustly scaled using baseline data so that deviations could be compared across heterogeneous signals. The AE was used to reconstruct the current state of the subsystem, while the GRU model was used to forecast its short-term evolution. The resulting reconstruction and forecasting errors were calibrated with respect to the baseline distribution and converted into SCI. The main implementation settings of the learning-based components are reported in Table II. The training used an 80%-20% partition of the baseline data in chronological order, with early stopping based on the validation loss. This configuration was used for all the subsystem-specific models.

TABLE II. MAIN IMPLEMENTATION SETTINGS OF THE LEARNING-BASED COMPONENTS.

Aspect	*AE*	*GRU*
Input	Current subsystem feature vector	Rolling subsystem feature sequence
Main role	Reconstruction of nominal subsystem states	Short-term forecasting of subsystem dynamics
Training data	Baseline nominal data	Baseline nominal data
Architecture	Dense encoder-decoder with 64–32–16 bottleneck	GRU layer with 64 hidden units and dense output
Error used	Mean absolute reconstruction error	Mean absolute one-step forecasting error
Use in SCI	Structural deviation component	Temporal deviation component

The second step in the workflow is to define subsystem-specific feature groups. These groups reflect the physical structure of the vehicle and the operational logic of maintenance activities. Battery-related signals capture electrochemical and thermal behaviour, drivetrain variables describe mechanical and energetic performance, brake-related features reflect braking workload and efficiency, and the tyres/suspension group captures conditions associated with load transfer, handling, and chassis behaviour. While the exact features available depend on the telemetry dataset being analysed, Table III summarises the representative variable groups considered by the architecture for each subsystem (Taylor *et al.*, 2016).

TABLE III. REPRESENTATIVE VARIABLE GROUPS CONSIDERED BY THE MONITORING ARCHITECTURE.

Subsystem	*Representative variables*	*Operational relevance*
Battery	State of Charge (SoC), State of Health (SoH), battery voltage/current, battery temperature, charge cycles, contextual variables	Supports monitoring of electrochemical stress, thermal behaviour, and charging-related degradation
Drivetrain	Motor temperature, vibration, torque, revolutions per minute (RPM), power consumption, contextual variables	Captures efficiency losses, abnormal load transfer, and evolving dynamic behaviour
Brakes	Brake pad wear, pressure, brake temperature, regenerative braking efficiency, contextual variables	Links alerts to braking workload and possible imbalance between regenerative and friction braking
Tyres/Suspension	Tyre pressure, tyre temperature, suspension load, contextual variables	Provides a route to identify handling, load-distribution, and chassis-related issues

The monitoring workflow then proceeds to identify a baseline specific to the subsystem in question, i.e. an initial reference window representing its normal behaviour. In realistic deployment scenarios, nominal periods are rarely labelled explicitly, meaning that it is not possible to select the reference behaviour manually in a reliable or scalable way (Jardine *et al.*, 2006). Therefore, the pipeline evaluates several candidate baseline windows located at the beginning of the time series. These include fixed fractions of the initial history and a scan window representing the initial operating period. For each candidate, a robust sanity score is computed based on the proportion of extreme values observed across the relevant variables, using interquartile-range fences. The candidate with the lowest score is then selected as the baseline for that subsystem. This reduces analyst subjectivity and recognises that different subsystems may stabilise at different rates, even within the same vehicle.

Once the baseline specific to the subsystem has been identified, the data preparation stage is designed to preserve temporal consistency and prevent information leakage from later, potentially degraded operating periods into the nominal reference. First, numeric variables are converted into a consistent numerical format. Then, missing values are managed using a leakage-safe procedure: forward filling is employed to preserve temporal continuity and any remaining missing values at the beginning of the acquisition history are replaced with medians estimated from only the selected baseline rows. This strategy is intentionally conservative. The aim is not to reconstruct an ideal signal, but to ensure that the reference remains anchored to nominal behaviour. After imputation, the variables are scaled using robust statistics derived from the baseline (i.e. medians and interquartile ranges) so that heterogeneous sensor units can be compared while limiting sensitivity to outliers (Chandola *et al.*, 2009). The main choices regarding preprocessing and calibration were made to prioritise operational robustness rather than optimising performance against labelled fault outcomes, which are unavailable in this setting. Robust scaling, baseline-derived calibration, quantile-based thresholds and persistence-based alarm activation were therefore adopted to reduce sensitivity to outliers, limit unstable alerts and preserve interpretability across subsystems. The same parameter-selection logic was consistently applied to all subsystem-specific models.

After the preparation of the data, the SCI is constructed by combining two complementary deviation channels. The first of these is a structural deviation channel, which captures how far the current multivariate sensor configuration deviates from patterns observed during nominal operation. In the current implementation, this component is obtained from an AE that has been trained using only baseline data (Heger *et al.*, 2020). Under reference conditions, the reconstruction error is expected to remain relatively small. However, when the subsystem moves away from the learned nominal patterns, the reconstruction error increases. This first channel is useful for identifying atypical states, such as unusual combinations of electrical or mechanical variables that may be associated with emerging degradation.

This structural view is complemented by a temporal deviation channel. Although reconstruction error indicates whether the current state is unusual, it does not explicitly capture whether the short-term evolution of the subsystem is becoming abnormal. For this reason, the architecture also includes a next-step predictor that is trained using baseline sequences. In the reference implementation, this predictor is a GRU model which estimates the expected change in the scaled sensor vector over the next time step using a rolling lookback window (Chung *et al.*, 2014). The resulting prediction error quantifies how unusual the local dynamics are relative to normal behaviour. This temporal component is particularly important when the first signs of degradation manifest as a change in short-term evolution and not as an obvious, instantaneous abnormality.

As a monitoring pipeline intended for practical use should remain functional under non-ideal conditions, the implementation includes controlled fallback options. For example, if the AE cannot be reliably trained or the available baseline is too short, the structural deviation channel can revert to a Principal Component Analysis (PCA)-based reconstruction model. Similarly, if the recurrent predictor cannot be trained under the available conditions, the temporal channel can revert to a ridge regression predictor. These alternatives are not intended to replace the preferred models in all situations, but to ensure continuous monitoring and prevent unreliable behaviour when the pipeline is first used, when there is limited data, or when computer settings are restricted.

Once the two deviation channels have been normalised and fused, the resulting signal is mapped onto the SCI via a calibrated sigmoid transformation (Warton, 2008):

$$\mathrm{SCI}(t) = \frac{1}{1+e^{\beta\left(z(t)-z_{ref}\right)}} \tag{1}$$

where $z(t)$ is the fused deviation signal at time t and z_{ref} is the baseline reference level, β is a calibration parameter estimated from baseline behaviour. This mapping is important because it enables maintenance users to interpret a bounded and monotonic condition scale more easily than raw model errors. In the implemented formulation, nominal operation remains in a clearly non-critical state while increasing deviation pushes the SCI towards lower values. Therefore, the SCI is an indicator of relative condition, based on learned reference behaviour, and not as a direct measure of probability of failure.

As telemetry streams may contain local fluctuations and short-lived excursions (Mahale *et al.*, 2025), the SCI is smoothed using an exponentially weighted moving average before alert generation. This improves readability, stabilises trend interpretation and enables more reliable alarm governance. The smoothed SCI is then translated into discrete operating states using thresholds derived from the SCI distribution observed during the baseline period. In practice, the baseline is used to estimate the range of values associated with normal subsystem behaviour, and fixed, quantile-based cut-offs are extracted from the resulting distribution. These cut-offs define a warning state and a more severe critical state. The warning state highlights behaviour that requires attention, even if it is not severe enough to justify an alarm. The critical threshold, on the other hand, is used for alarm activation. To avoid alarms caused by isolated fluctuations, activation is persistence-based: the SCI must remain below the critical threshold for a predefined number of consecutive time steps before the alarm is raised. The process of alarm clearance is governed by hysteresis, which signifies that recovery necessitates the SCI to rise above a less severe clearance threshold, as opposed to merely crossing back over the alarm-on threshold. This threshold design reduces chattering and limits alarm fatigue, making the alerting logic more suitable for day-to-day maintenance support.

At each telemetry step, the tool provides a concise set of maintenance-oriented monitoring outputs for each subsystem. These include the current SCI value, the current alarm state, the short-term SCI trend, and the threshold band used for interpretation. The architecture also provides simple reason codes, obtained by aggregating absolute reconstruction residuals during alarm periods and identifying the variables with the highest average contribution. These reason codes should be interpreted as inspection-support cues rather than root-cause identifications. They are intended to improve operational clarity and facilitate maintenance triage by indicating which variables and subsystem groups may require inspection first. In this sense, the proposed monitoring layer can also be seen as a first step towards supporting prescriptive maintenance (Ansari *et al.*, 2019), since persistent alarms and their associated inspection-support cues may later be linked to recommended inspection or intervention actions. The outputs returned by the pipeline in dashboard format at each time step are summarised in Table IV.

TABLE IV. OPERATIONAL OUTPUTS PROVIDED BY THE SUBSYSTEM MONITORING PIPELINE.

Pipeline output	*Generated by the pipeline*	*Interpretation*	*Maintenance use*
SCI value	Bounded SCI in [0,1]	Lower values indicate stronger deviation from baseline behaviour	Quick view of current subsystem status
Trend over time	Gradient of the SCI stream	Negative trend indicates worsening behaviour	Prioritise assets before the condition becomes persistent
Alarm state	Persistence and hysteresis applied to the SCI	Binary indication of active/non-active critical condition	Stable trigger for inspection or escalation

Pipeline output	Generated by the pipeline	Interpretation	Maintenance use
Reason codes	Top residual contributors during alarm windows	Inspection-support cues associated with subsystem deviations	Guide first inspection checks and maintenance triage

IV. Case Study

The proposed monitoring architecture was evaluated using the publicly available EVIoT Predictive Maintenance dataset on Kaggle (Kaggle, 2025). This electric-vehicle telemetry dataset is sampled at 15-minute intervals and includes electrical, thermal, mechanical and operational measurements. It represents a realistic connected-vehicle monitoring scenario in which explicit fault labels are unavailable for supervised modelling. Consequently, the case study does not seek to quantify classical fault-detection accuracy, the false-positive rate or time to detection. Instead, the evaluation focuses on the operational behaviour of the monitoring architecture. This includes the stability of SCI trajectories, manageable alarm rates, sensitivity to baseline definition and robustness under moderate missing-data perturbations. These criteria were selected because they are directly related to deployment usability in label-scarce monitoring settings. Thus, the investigation's objective is to evaluate if the suggested subsystem-level framework is capable of producing significant, practically beneficial maintenance assistance results from unlabelled telemetry.

A first indication of applicability is provided by the subsystem SCI trajectories in Fig. 2, where the battery, drivetrain, brakes, and tyres/suspension subsystems are tracked together with the alarm-state markers and the fixed alarm-on and clearance thresholds. In this case study, the two thresholds are derived from the baseline SCI distribution using the 0.25 and 0.35 quantiles, respectively, to define conservative alarm and recovery regions. This representation preserves subsystem-specific behaviour, rather than collapsing the monitored asset into a single aggregate signal. This makes it possible to distinguish localised deviations from more widespread abnormal behaviour within the same operating period. At the same time, the plotted alarm states reflect the persistence-and-hysteresis logic adopted by the architecture: isolated excursions do not immediately generate alarms, and recovery is only recognised once the indicator rises above the clearance threshold. The resulting output is therefore stable enough for day-to-day monitoring yet still highlights subsystem-specific deterioration patterns that could be relevant for maintenance.

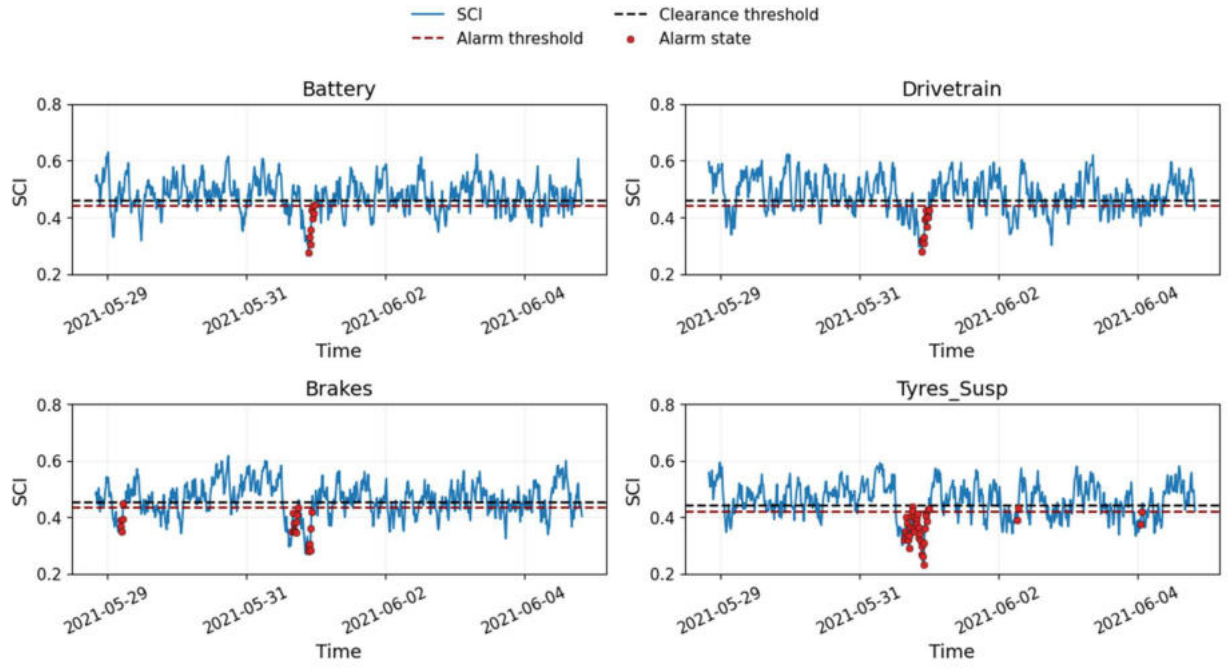


Fig. 2. Subsystem SCI trajectories with frozen thresholds and alarm markers.

A second indication of applicability relates to the burden and manageability of the alerts generated. While a monitoring architecture may be technically feasible, it could still be impractical if it produces excessive, unstable or ambiguous alarms. For this reason, the case study examines the mean number of alarm episodes per day for each subsystem (see Fig. 3), as well as a broader set of operational metrics reported in Table V. In this study, an alarm episode is defined as the start of an alarm period following a non-alarm state and not as the total number of consecutive samples below the threshold. This definition is more meaningful from a maintenance perspective because it reflects how often the monitoring system actually requires attention. Both Fig. 3 and Table V cover the entire case study analysis period. In practical deployments, the same metrics could be recalculated using rolling daily or weekly windows depending on the intended monitoring workflow.

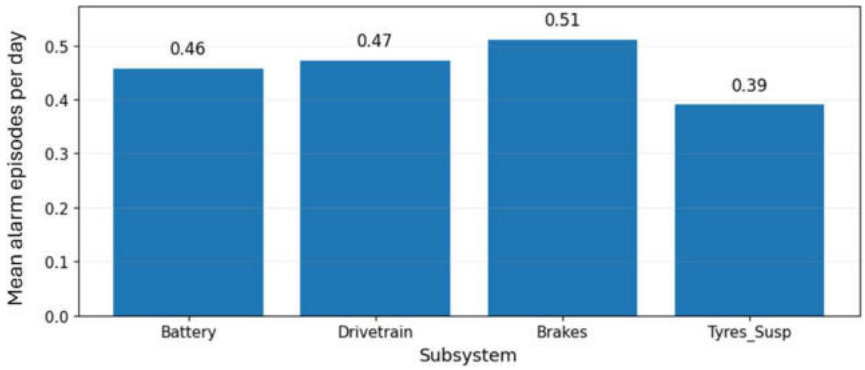


Fig. 3. Mean alarm episodes per day for each subsystem, computed over the full case-study horizon.

As shown in Fig. 3, the alert burden remains relatively low and consistent across subsystems throughout the analysis period, with mean rates of alarm episodes ranging from approximately 0.39 to 0.51 per day. These values are obtained by dividing the total number of alarm activations observed for each subsystem by the corresponding observation period in days. Brakes exhibit the highest alert frequency, tyres/suspension the lowest and battery/drivetrain lie in between. This suggests that the proposed architecture disperses deviations across multiple alert streams without concentrating them in a single dominant stream. Consequently, it provides a differentiated view of subsystem behaviour while maintaining an alert volume that remains manageable in operational terms.

Table V provides a summary of the main operational characteristics of the alerts over the same analysis horizon as Fig. 3. As well as the mean number of alarm episodes per day, it reports the median SCI, the proportion of time spent in the critical state, the mean and maximum durations of alarm episodes, and the median time between consecutive alarm activations. The SCI is presented as the median to provide a more robust description of typical subsystem conditions in the presence of short-lived excursions and alarm-related deviations. Taken together, these metrics suggest that the generated alerts correspond to structured periods of deviation and not sporadic threshold crossings. Mean alarm durations remain at around two hours, maximum durations remain limited and the typical time between events remains well above one day for all subsystems. This is consistent with an alerting logic designed to support observation and prioritisation over the generation of unstable alarms.

TABLE V. MAIN OPERATIONAL ALERT METRICS PER SUBSYSTEM, COMPUTED OVER THE FULL CASE-STUDY HORIZON.

Subsystem	*Median SCI*	*Mean episodes per day*	*Time fraction in critical state (%)*	*Mean alarm duration (h)*	*Max alarm duration (h)*	*Median inter-event time (h)*
Battery	0.474	0.46	4.07	2.13	13.50	34.88
Drivetrain	0.465	0.47	4.06	1.91	9.50	34.25
Brakes	0.472	0.51	4.09	2.02	12.50	34.38
Tyres/Suspension	0.475	0.39	3.29	1.92	9.25	42.50

In line with the scope of this evaluation, these alerts should not be interpreted as confirmed faults or as automatic indications that maintenance is required immediately. Their purpose is to provide an early warning that requires observation, contextual assessment and, where appropriate, further inspection. In practice, some alerts may correspond to transient or easily recoverable deviations, some may be deemed non-actionable, and some may prompt maintenance planning or diagnostic follow-up. Therefore, the objective is not to optimise alarm counts against an external ground truth, which is unavailable in this label-scarce setting, but to verify that the architecture produces outputs that are both interpretable and operationally governable.

Once the alerting outputs have been characterised, it is important to assess their stability under plausible non-ideal conditions, without relying on a specific baseline choice or complete data. Table VI therefore summarises the robustness of the outputs under two moderate perturbation settings: an alternative baseline definition and additional missing data. For baseline variation, the reference SCI is compared with the SCI obtained when the baseline window is set to 20% of the available time series. For missingness stress testing, 10% of the values in the multivariate telemetry matrix are randomly masked, after which the matrix is reprocessed through the same monitoring pipeline. SCI rank correlation is measured using Spearman's rank correlation coefficient, evaluating stability in terms of preserving the overall SCI trajectory ordering over exact pointwise agreement (Stephanou and Varughese, 2021). Values close to one therefore indicate that the overall subsystem monitoring pattern remains essentially unchanged. The remaining quantities report the absolute change in the mean number of alarm episodes per day and the proportion of time spent in the critical state under the 10% missingness condition. This captures the practical effect of the perturbation on the burden of alerts.

The results show that the monitoring outputs remain highly stable under both considered perturbation settings. Under the alternative baseline condition, where the baseline window is set to 20% of the available time series, the Spearman SCI rank correlations are near one for all subsystems. This indicates that the indicator's overall ordering behaviour is preserved. In the case of 10% missingness, Spearman SCI rank correlations remain at least 0.92, while the magnitude of the absolute changes in mean alarm episodes per day and critical-state time fraction remains limited. This suggests that moderate

baseline variation and additional missingness do not significantly affect the operational picture produced by the architecture.

TABLE VI. ROBUSTNESS SUMMARY UNDER ALTERNATIVE BASELINE DEFINITION AND ADDITIONAL MISSINGNESS (I.E. MISSING DATA).

Subsystem	*Spearman SCI rank correlation (20% baseline condition)*	*Spearman SCI rank correlation (10% missingness)*	*Absolute change in mean alarm episodes/day (10% missingness)*	*Absolute change in critical state (%) (10% missingness)*
Battery	1.00	0.93	+0.01	+0.15
Drivetrain	0.99	0.92	-0.05	-0.13
Brakes	0.99	0.92	-0.08	-0.38
Tyres/Suspension	0.99	0.92	-0.01	-0.10

V. OPERATIONAL IMPLICATIONS

Section IV demonstrated that the proposed architecture can produce interpretable and consistent outputs in a realistic, label-scarce case study. The following question is whether these outputs are useful in practice. While not aiming for automatic fault diagnosis, the proposed approach organises raw telemetry into early-warning information at the subsystem level that can support monitoring, triage and initial inspection planning.

Compared with simpler alternatives such as fixed sensor thresholds, vehicle-level anomaly scores, and single-channel reconstruction-based monitoring, the proposed architecture is designed to provide more operationally usable outputs. While fixed thresholds are easier to implement, they may be sensitive to context and generate unrelated alerts across multiple variables. Vehicle-level anomaly scores support generic ranking, but they provide limited information about where to begin inspection. Single-channel, reconstruction-based monitoring captures unusual multivariate states, but it can miss deviations that primarily emerge through abnormal short-term dynamics. The proposed architecture combines subsystem-level grouping, reconstruction and forecasting evidence, baseline-referenced calibration and alarm-governance logic to offer a more structured compromise between technical anomaly detection and maintenance usability.

One operational advantage lies in the subsystem-level structure of the outputs. Since one SCI is generated for each monitored subsystem, the user is not limited to a generic indication that the vehicle is deviating from normal behaviour. Instead, the monitoring interface can indicate whether the most significant deviation is primarily associated with the battery, drivetrain, brakes or tyres/suspension. This is important in practice, as inspection and maintenance activities are typically organised around subsystem responsibilities rather than a single, undifferentiated, vehicle-level score.

Unlike the aggregate metrics reported in Section IV, which are computed over the full case-study horizon for evaluation purposes, Fig. 4 provides a dashboard-oriented view based on a rolling window set to the last 24 hours. This reflects how outputs could be monitored on a daily basis. Panel (a) provides an immediate overview of the recent condition of the subsystem by reporting the mean SCI over the last 24 hours. This is computed from the corresponding 15-minute SCI updates together with the associated alarm-on and clearance thresholds. Panel (b) complements this view by reporting the mean short-term SCI trend over the same 24-hour period. This indicates whether recent behaviour is stable, improving (increasing SCI) or deteriorating (decreasing SCI). In the example, none of the subsystems are currently in alarm mode, but panel (a) still allows comparison of how close each subsystem is to the threshold region. The trend lines in Panel (b) show slight negative trends in drivetrain, brakes and tyres/suspension, with tyres/suspension showing the steepest decline. From an operational perspective, this enables the user to transition from a general condition overview to subsystem prioritisation and closer observation within the same interface.

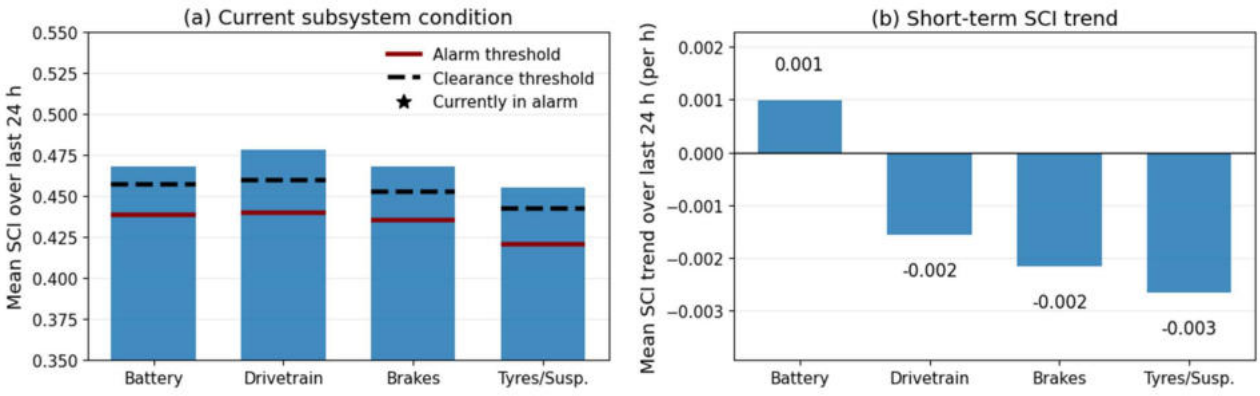


Fig. 4. Example operational interpretation of the monitoring outputs. Panel (a) summarises recent subsystem SCI levels, while panel (b) shows the corresponding short-term SCI trends.

More broadly, the architecture converts generic anomaly signals into structured, subsystem-level monitoring outputs that can be interpreted within an operational context. These outputs can then be used to support the prioritisation of tasks, the planning of inspections, and the decision-making process for maintenance. Its practical value lies not only in signalling that a deviation exists, but also in helping to determine where to look first and which subsystem deserves closer attention. It can also help to establish whether the observed behaviour is stable or progressively deteriorating. In this sense, the

dashboard represents an initial step towards providing more prescriptive maintenance support, as persistent alarms and subsystem-specific cues could eventually be linked to recommended inspection or intervention actions.

Table VII supplements Fig. 4 by reporting representative first-inspection cues derived from reason codes. These cues identify the variables most strongly associated with the current subsystem state or with a detected deviation. They should be interpreted as cues to support inspections and not to explain the root cause. They improve operational clarity, reduce ambiguity in alarm handling, and support maintenance triage by indicating where technical assessment could begin. These cues are useful even when no alarm is active, as they help to focus attention on the variables that contribute most to the monitoring output at the subsystem level.

Together, Fig. 4 and Table VII demonstrate how the proposed architecture can serve as a practical maintenance support layer between raw telemetry and engineering judgement. The SCI provides a concise subsystem-level condition summary, the short-term trend adds anticipatory value, and the reason-code-based cues support first-line drill-down. Overall, the result is a monitoring output that is more maintenance-oriented and can be used directly for practical supervision, inspection prioritisation, and supporting first-line maintenance decisions in connected-vehicle settings where labels are scarce.

TABLE VII. REPRESENTATIVE FIRST-INSPECTION CUES DERIVED FROM SUBSYSTEM REASON CODES.

Subsystem	***Current alarm state***	***Suggested first inspection cue***
Battery	No alarm	SoC
Drivetrain	No alarm	Motor vibration
Brakes	No alarm	Regenerative braking efficiency
Tyres/Suspension	No alarm	Load weight

VI. CONCLUSIONS AND FUTURE WORK

This paper presented a subsystem-oriented monitoring architecture for connected vehicles based on unlabelled telemetry. The proposed tool transforms multi-sensor vehicle data into SCIs designed to support maintenance interpretation through bounded condition signals, stable alert logic, and concise first-inspection cues. In this way, the architecture does more than detect deviations: it organises them into dashboard-oriented outputs that can support monitoring, prioritisation, and initial maintenance decision-making at subsystem level, while also providing a basis for future prescriptive-maintenance extensions.

The main contribution of the work lies in showing how established modelling components can be integrated into an operationally meaningful monitoring layer for label-scarce settings, which represents a practically relevant contribution in the field. The EV telemetry case study demonstrated that the proposed architecture preserves subsystem differentiation and generates interpretable, manageable alert behaviour. It also remains stable under moderate baseline variation and additional missing data. These properties are important because they transform the output from a generic anomaly signal into a structured representation that can help to determine where attention should be directed, which subsystem may require escalation and how initial inspection efforts can be prioritised.

Although the present evaluation is based on a single-vehicle dataset, the architecture is designed for fleet-oriented deployment. In this sense, this work does not constitute fleet-level validation, but the definition and initial evaluation of an interpretable, subsystem-level monitoring layer that could support inspection prioritisation, maintenance coordination and more scalable, connected-vehicle supervision. Future work will extend the evaluation to multi-vehicle datasets to assess cross-vehicle comparability and prioritisation strategies at fleet level. When partial maintenance outcomes are available, they can be used to refine alert governance and evaluate the practical usefulness of the generated alerts, eliminating the need for dense failure labels. Further developments will also address refreshing the baseline and operational drift, including the effects of seasonality, ageing-related shifts and sensor recalibration. Finally, the link between alarm patterns and reason-code information and recommended inspection or intervention actions will be explored as a step towards providing more prescriptive maintenance support.

References

Ansari, F., Glawar, R. and Nemeth, T. (2019), "PriMa: a prescriptive maintenance model for cyber-physical production systems", International Journal of Computer Integrated Manufacturing, Taylor & Francis, Vol. 32 No. 4–5, pp. 482–503.

Benhanifia, A., Cheikh, Z. Ben, Oliveira, P.M., Valente, A. and Lima, J. (2025), "Systematic review of predictive maintenance practices in the manufacturing sector", Intelligent Systems with Applications, Vol. 26, p. 200501.

Chandola, V., Banerjee, A. and Kumar, V. (2009), "Anomaly detection: A survey", ACM Computing Surveys (CSUR), ACM New York, NY, USA, Vol. 41 No. 3, pp. 1–58.

Chung, J., Gulcehre, C., Cho, K. and Bengio, Y. (2014), "Empirical evaluation of gated recurrent neural networks on sequence modeling", ArXiv Preprint ArXiv:1412.3555.

Giannoulidis, A. and Gounaris, A. (2023), "A context-aware unsupervised predictive maintenance solution for fleet management", Journal of Intelligent Information Systems, Vol. 60 No. 2, pp. 521–547.

Heger, J., Desai, G. and El Abdine, M.Z. (2020), "Anomaly detection in formed sheet metals using convolutional autoencoders", Procedia CIRP, Vol. 93, pp. 1281–1285.

Hu, Y., Miao, X., Si, Y., Pan, E. and Zio, E. (2022), "Prognostics and health management: A review from the perspectives of design, development and decision", Reliability Engineering & System Safety, Vol. 217, p. 108063.

Jardine, A.K.S., Lin, D. and Banjevic, D. (2006), "A review on machinery diagnostics and prognostics implementing condition-based maintenance", Mechanical Systems and Signal Processing, Vol. 20 No. 7, pp. 1483–1510.

Kaggle. (2025), "EVIoT-PredictiveMaint Dataset", available at: https://www.kaggle.com/datasets/datasetengineer/eviot-predictivemaint-dataset (accessed 30 December 2025).

Lipu, M.S.H., Hannan, M.A., Hussain, A., Hoque, M.M., Ker, P.J., Saad, M.H.M. and Ayob, A. (2018), "A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations", Journal of Cleaner Production, Elsevier, Vol. 205, pp. 115–133.

Mahale, Y., Kolhar, S. and More, A.S. (2025), "A comprehensive review on artificial intelligence driven predictive maintenance in vehicles: technologies, challenges and future research directions", Discover Applied Sciences, Vol. 7 No. 4, p. 243.

Nunes, P., Santos, J. and Rocha, E. (2023), "Challenges in predictive maintenance – A review", CIRP Journal of Manufacturing Science and Technology, Vol. 40, pp. 53–67.

Resanovic, D., Softic, S. and Balc, N. (2026), "Explainable Anomaly Detection for Predictive Maintenance Using Synthetic Benchmarks and Human-in-the-Loop Labeling BT - Advances in Production Management Systems. Cyber-Physical-Human Production Systems: Human-AI Collaboration and Beyond", in Mizuyama, H., Morinaga, E., Nonaka, T., Kaihara, T., von Cieminski, G. and Romero, D. (Eds.), , Springer Nature Switzerland, Cham, pp. 89–103.

Serradilla, O., Zugasti, E., Rodriguez, J. and Zurutuza, U. (2022), "Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects", Applied Intelligence, Springer, Vol. 52 No. 10, pp. 10934–10964.

Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., et al. (2019), "Data-driven prediction of battery cycle life before capacity degradation", Nature Energy, Nature Publishing Group UK London, Vol. 4 No. 5, pp. 383–391.

Sharma, A. (2024), "Making electric vehicle batteries safer through better inspection using artificial intelligence and cobots", International Journal of Production Research, Vol. 62 No. 4, pp. 1277–1296.

Stephanou, M. and Varughese, M. (2021), "Sequential estimation of Spearman rank correlation using Hermite series estimators", Journal of Multivariate Analysis, Vol. 186, p. 104783.

Swaroopa Rani, B. and Jatoth, C. (2025), "Deep Learning and Blockchain-Enabled Predictive Maintenance in Electric Vehicles: A Comprehensive Review", Concurrency and Computation: Practice and Experience, Wiley Online Library, Vol. 37 No. 25–26, p. e70298.

Taylor, P., Griffiths, N., Bhalerao, A., Anand, S., Popham, T., Xu, Z. and Gelencser, A. (2016), "Data mining for vehicle telemetry", Applied Artificial Intelligence, Taylor & Francis, Vol. 30 No. 3, pp. 233–256.

Theissler, A., Pérez-Velázquez, J., Kettelgerdes, M. and Elger, G. (2021), "Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry", Reliability Engineering & System Safety, Vol. 215, p. 107864.

Warton, D.I. (2008), "Penalized normal likelihood and ridge regularization of correlation and covariance matrices", Journal of the American Statistical Association, Taylor & Francis, Vol. 103 No. 481, pp. 340–349.

Zonta, T., da Costa, C.A., da Rosa Righi, R., de Lima, M.J., da Trindade, E.S. and Li, G.P. (2020), "Predictive maintenance in the Industry 4.0: A systematic literature review", Computers & Industrial Engineering, Vol. 150, p. 106889.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Preliminary Framework for Verifying the Syntactic Quality of Digital Voice-of-Customer for Product Quality Monitoring

Barravecchia Federico[1], Mastrogiacomo Luca[2] and Franceschini Fiorenzo[3]

[1] Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy; federico.barravecchia@polito.it
[2] Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy; luca.mastrogiacomo@polito.it
[3] Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy; fiorenzo.franceschini@polito.it

***Abstract*—Voice-of-Customer (VoC) data, such as online customer reviews and unsolicited customer feedback, has become a valuable complement to traditional quality information in manufacturing contexts and within the broader paradigm of Quality 4.0. In particular, digital VoC provides post-market evidence on product quality, supporting the monitoring of field performance and the early detection of customer-perceived defects. However, the reliability of digital VoC-based monitoring depends critically on the quality of the underlying data. When digital VoC datasets are noisy, incomplete, inconsistent, or manipulated, even advanced analytics may generate misleading signals, reflecting the well-known "garbage in, garbage out" effect. This paper addresses the problem of digital VoC data reliability for product quality monitoring by grounding the discussion in established data-quality principles, with particular reference to ISO 8000. The standard distinguishes three complementary perspectives on data quality: syntactic, semantic, and pragmatic, which together determine whether data can support reliable analysis and decision-making. Building on this conceptual structure, the paper develops a preliminary framework for assessing the syntactic quality of digital VoC datasets. Syntactic quality concerns the formal correctness of data, independently of their meaning or usefulness. The proposed framework operationalizes syntactic quality through a set of subdimensions and associated metrics, providing a foundation for assessing the reliability of VoC datasets before performing downstream analytics in manufacturing quality monitoring. While the broader problem of VoC reliability includes semantic and pragmatic considerations, this work focuses specifically on the syntactic dimension as a necessary first step toward a comprehensive reliability assessment framework for digital VoC.**

Keywords – Digital Voice of Customer; Data Quality; Quality 4.0; Customer Reviews; ISO 8000; Reliability; Product Quality Monitoring.

I. Introduction

Digital Voice-of-Customer (VoC) has established itself as a rich and timely source of customer feedback (Özdağoğlu et al., 2018; Barravecchia et al., 2023). Digital VoC is produced continuously and without prompting, providing a stream of information about real-world usage, perceived performance, and emerging dissatisfaction on products and services. In manufacturing industries, these data are particularly valuable because they provide post-market evidence on how products perform in real usage conditions, complementing internal quality indicators such as non-conformities, warranty claims, service reports, and in-process quality data. In Quality 4.0, commonly understood as the digitalization of quality management through connectivity, advanced analytics, and new data sources (Zonnenshain and Kenett, 2020; Broday, 2022), digital VoC can act as a complementary "external sensor" that enriches internal quality signals (Barravecchia et al., 2022b).

Quality management research has explored how digital VoC can support monitoring and tracking quality (Barravecchia et al., 2023; Barravecchia et al., 2022a). In particular, online reviews and user-generated feedback have been used as a complementary source of information to detect early signals of quality evolution and customer-perceived issues. Because these data reflect real-world product usage, they can provide indications of emerging problems that may not yet appear in traditional internal quality metrics(Barravecchia et al., 2023). In the manufacturing domain, such analyses are especially relevant when digital VoC is used to identify recurring defect-related complaints, detect early signs of product quality deterioration, and support the prioritization of engineering investigations. Several studies have proposed approaches for quality tracking based on the analysis of review texts, using techniques such as sentiment analysis (Kim, 2021), topic modelling (Yan et al., 2025), and large language model-based methods (Park et al., 2026) to identify recurring defect-related discussions or dissatisfaction patterns. Research have also combined star ratings and textual content to detect shifts in customer perception over time (Barravecchia et al., 2026), applying time-series analysis (See-To and Ngai, 2018) or control-chart approaches (Kakde et al., 2024) to identify anomalies and potential quality degradation. These contributions illustrate that digital VoC analytics can help identify quality-relevant patterns. At the same time, the practical value of VoC is already evident in many organizational contexts, where actionable insights are obtained from reviews, and in many cases digital VoC can reveal issues earlier than conventional post-market surveillance channels (Kim and Park, 2017). The goal of this paper is to show that digital VoC-based quality monitoring is stronger and more defensible when it incorporates an explicit layer of input-data reliability assessment.

In practice digital VoC approaches frequently emphasize model development and insight extraction more than the evaluation of dataset quality. In domains like online reviews, the need is particularly critical because there is limited quality control over who can contribute and what they can write; this creates conditions for low-quality content and deliberate manipulation.

In this context, the core premise is that monitoring results are only as reliable as the data used to generate them. This is a general principle in information systems and decision-making: if data is incomplete, inconsistent, or unfit for purpose, decisions degrade correspondingly (Shankaranarayanan and Cai, 2006). In digital VoC, the risk of "garbage in, garbage out" is amplified by unstructured language, off-topic content, selection effects, and deception. In product quality monitoring, this issue is particularly critical because unreliable digital VoC data may distort the identification of defect patterns, mask emerging field failures, or trigger unnecessary alarms, thereby affecting downstream quality decisions.

This paper addresses the general problem of VoC data reliability for quality monitoring by grounding the discussion in established data-quality principles and, in particular, in the ISO 8000 standard (International Organization for Standardization, 2011) that conceptualizes data quality through three complementary perspectives: syntactic, semantic, and pragmatic, each capturing a different aspect of how data support analysis and decision-making. Within this broader perspective, the present study focuses specifically on syntactic quality, which concerns the formal correctness of data. The paper develops a preliminary framework that operationalizes syntactic quality for digital VoC datasets through a set of subdimensions and associated metrics.

Against this background, the study addresses the following research questions:

- RQ1: How can the syntactic quality of digital VoC data for product quality monitoring be conceptually defined and structured?
- RQ2: How can syntactic be operationalized through measurable subdimensions, metrics, and a structured verification-and-improvement protocol for digital VoC datasets?

The remainder of the paper is structured as follows. Section II discusses why Digital VoC reliability constitutes a distinct challenge in quality monitoring contexts. Section III reviews established data-quality principles. Section IV introduces ISO 8000 as the conceptual foundation of the proposed framework. Section V operationalizes the syntactic quality dimension for Digital VoC datasets. Section VI introduces a structured approach for improving syntactic quality through data cleaning activities. Section VII presents an application case to illustrate the practical implementation of the proposed framework. Finally, Section VIII concludes the paper and outlines directions for future research.

II. Why VoC Reliability Is a Distinct Problem in Quality Monitoring

Digital VoC data differs from many traditional quality datasets because it is user-generated, largely unstructured, and not governed by the organization that uses it. This creates different structural reliability challenges.

First, review texts can be extremely brief, off-topic, or highly subjective, and real-world review corpora often contain a substantial amount of spam content, or unhelpful information (Mewada and Dewang, 2023). A second challenge is a systematic bias in who writes reviews and when. Research has highlighted that online ratings often follow a J-shaped distribution, reflecting self-selection and incentives that are not random samples of the customer population (Ullah et al., 2016). In monitoring terms, this implies that the dataset may exaggerate extreme experiences, complicating the interpretation of trends as "true" quality shifts (Subroto and Christianis, 2021). A third challenge is manipulation. Opinion spam and the trustworthiness of online opinions have been identified as critical issues in review analysis (Hu et al., 2012). Early large-scale studies of reviews argued that trustworthiness was a neglected problem in review mining and showed how untruthful opinions, brand-only reviews, and non-reviews can distort the dataset (Esposito et al., 2021). Finally, even when reviews are genuine, the practical usefulness of text is influenced by clarity and interpretability. Writing quality, readability, and error-related signals have been treated as important factors that shape perceived helpfulness and, by extension, the practical value of reviews as information (Santhiran et al., 2024).

These issues do not reduce the usefulness of digital VoC. Instead, they highlight the importance of carefully assessing the reliability of the data. If digital VoC is used as an input to quality monitoring, especially for decisions that might influence engineering investigations, design changes, supplier actions, or customer communication, then the dataset should be treated like an "information product" whose quality must be defined and validated for use.

In manufacturing contexts, these limitations are not merely analytical inconveniences. When digital VoC is incorporated into product quality monitoring systems, syntactic and structural defects in the data can affect defect detection, distort product-level comparisons, and weaken the evidential basis for decisions concerning root-cause analysis, supplier assessment, redesign priorities, or corrective actions. For this reason, Digital VoC data should be treated as a quality-relevant information asset whose reliability must be verified before being used in product quality evaluation.

III. Recognized Data-Quality Principles

A key reason why digital VoC reliability is often hard to discuss is the lack of shared vocabulary. Data-quality research has long established that quality is multidimensional. A widely cited perspective is that data quality is not only accuracy: it includes intrinsic dimensions (e.g., accuracy, believability), contextual dimensions (e.g., relevance, timeliness), representational dimensions (e.g., interpretability, consistency of representation), and accessibility-related dimensions (Wand and Wang, 1996). The "Beyond Accuracy" framework articulates these categories and shows that data consumers evaluate quality across multiple criteria, not a single one (Wang and Strong, 1996).

Several surveys emphasize the breadth of dimensions and the need to interpret them relative to context, particularly for large-scale and heterogeneous datasets (Fox et al., 1994; Ridzuan and Zainon, 2024; Wang et al., 2024).

Within this broader landscape, ISO 8000 (International Organization for Standardization, 2011) provides a relevant foundation for a Quality 4.0 setting because it makes data quality conceptually compatible with quality management thinking. ISO vocabulary defines data quality as the degree to which inherent characteristics of data fulfil requirements (International Organization for Standardization, 2015).

IV. ISO 8000 as Conceptual foundation for measuring Digital VoC Data Quality

According to ISO 8000-8 (International Organization for Standardization, 2015), information and data quality can be understood through three complementary perspectives, each addressing a different aspect of how data support decision-making and analysis (see Figure 1):

- Syntactic quality refers to the formal correctness of data: it concerns whether data conform to predefined schemas, formats, and integrity rules, ensuring that records are structurally valid and processable by information systems, independently of their meaning or usefulness.
- Semantic quality concerns whether data correspond correctly, unambiguously, and meaningfully to the real-world domain they are intended to represent. In ISO 8000-8, semantic quality is expressed in terms of a correct and complete mapping between identifiable data units and the entities or concepts of the domain of interest.
- Pragmatic quality concerns whether data are suitable for a specific purpose or application. It focuses on whether the data, even if formally correct and semantically meaningful, are actually useful for supporting the decisions or activities they are intended to inform.

This triad is particularly useful for digital VoC because it clarifies that "being well-formed" is not equivalent to "being trustworthy," and "being correct" is not equivalent to "being fit for monitoring decisions". This ISO structure suggests a structured approach: digital VoC quality should be evaluated at all three levels, because failures at any level can undermine monitoring conclusions.

ISO 8000 also distinguishes verification and validation (International Organization for Standardization, 2015): some aspects of data quality are checked against specified requirements (verification), while others are judged against intended use (validation). This distinction becomes central when digital VoC data is syntactically well-formed but still unreliable for monitoring.

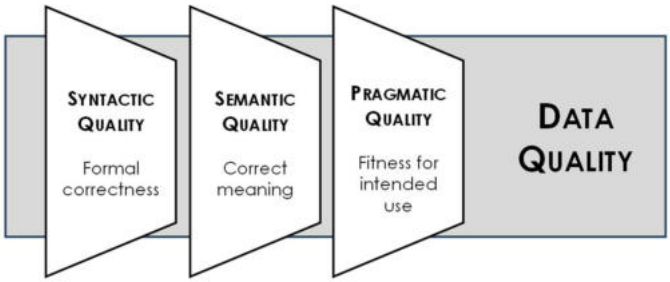


Figure 1. Conceptual representation of data quality according to ISO 8000-8 (International Organization for Standardization, 2015)

V. Operationalizing Syntactic Quality of Digital VoC Data

As seen in the previous section, ISO 8000-8 conceptualizes data quality through three complementary dimensions (International Organization for Standardization, 2015). While all three dimensions are relevant for assessing the reliability of Digital VoC data, the focus of this article is exclusively on the syntactic dimension. According to ISO 8000-8, syntactic quality refers to the degree to which data conform to formally specified metadata (International Organization for Standardization, 2015). In this study, syntactic quality is operationalized by translating the ISO 8000 notion of conformance to formally specified metadata into classes of verifiable constraints for Digital VoC datasets. Specifically, the six subdimensions were derived by identifying the main types of metadata requirements that can be formally specified and checked before analysis: data-type requirements, mandatory-field requirements, domain and format requirements, identifier uniqueness requirements, inter-attribute integrity requirements, and text representation requirements. These requirement classes were then mapped respectively onto six syntactic subdimensions: structural validity, completeness, format validity, entity integrity, cross-field consistency, and text processability. In particular, they were derived by grouping the main categories of formally verifiable defects discussed in the literature, such as type mismatches, missing mandatory values, domain and format violations, duplicate records, inter-field inconsistencies, and text representation problems (Wang and Strong, 1996). Consistent with this perspective, these dimensions were defined to capture typical situations of syntactic non-conformance, such as incorrect data types, missing mandatory values, domain or format violations, duplicate records, inconsistencies between related

attributes, and problems in text representation. Similar categories of data defects are widely discussed in data cleaning and data quality assessment research (Rahm and Do, 2000). Each subdimension therefore represents a specific category of syntactic defect.

Table 1 provides an overview of the syntactic subdimensions and associated metrics proposed in this study. While these subdimensions are informed by established data-quality literature and the ISO 8000 conceptualization of syntactic quality, their specific definition and organization represent an original contribution of this work.

Table 1. Syntactic quality subdimensions and releted metrics for digital VoC data. All indicators are normalized in the interval [0,1], where 1 represents full compliance.

Syntactic Subdimension	ISO 8000 Metadata Requirement	What It Verifies	Metric
STRUCTURAL VALIDITY	Data type	Conformance of attribute values to declared data types	*Type Conformance Rate (TCR)*
COMPLETENESS	Mandatory attribute	Presence of mandatory attributes defined in metadata	*Mandatory Attribute Completeness Ratio (MACR)*
FORMAT VALIDITY	Domain / representation	Compliance with declared domains and representation formats	*Domain Compliance Rate (DCR)*
ENTITY INTEGRITY	Uniqueness	Uniqueness of records according to declared identifiers	*Uniqueness Rate (UR)*
CROSS-FIELD CONSISTENCY	Integrity	Logical consistency between related attributes	*Compliance Score (CS)*
TEXT PROCESSABILITY	Representation	Compliance of textual data with encoding and normalization standards	*Text Representation Compliance Rate (TRCR)*

A. Structural Validity

Structural validity concerns adherence to declared data types. ISO 8000-8 explicitly anchors syntactic quality in conformance to requirements stated in metadata. Data type inconsistencies are recognized as a primary source of structural defects, particularly in multi-source environments (Rahm & Do, 2000).

In Digital VoC contexts, structural validity is critical because review datasets are typically aggregated from heterogeneous platforms. For example, one platform may encode ratings as integers (1–5), while other stores ratings as textual strings (e.g., "5 stars"). If data type harmonization is incomplete, data analytics may produce failures. Such defects represent violations of formally declared structure and therefore fall within syntactic quality.

To illustrate violations of structural validity, Table 2 reports examples of records that do not conform to the declared data types. These records contain type mismatches across multiple attributes. The Rating field includes non-integer values, the Timestamp field contains a non-standard textual expression, and Helpful_Votes is expressed as a string. Such violations prevent correct parsing and compromise the structural integrity of the dataset.

Structural validity can be operationalized through a Type Conformance Rate (TCR), defined as the proportion of records that conform to the declared data types. Let N be the total number of records and N_s the number of records satisfying all constraints. The metric can be defined as:

$$TCR = \frac{N_s}{N} \qquad \text{TCR} \in [0,1] \tag{1}$$

A record is considered-compliant if all attributes respect the declared data types.

Table 2. Example of structural validity violations. The table reports review identifier (Review_ID), unstructured textual content (Review_Text), product identifier (Product_ID), customer rating (Rating), timestamp (Timestamp), and helpfulness indicators (Helpful_Votes, Total_Votes).

Review_ID	Review_Text	Product_ID	Rating	Timestamp	Helpful_Votes	Total_Votes
R1	I've been using this device for a couple of weeks and overall it works quite well, especially considering the price.	P100	"five"	2026-01-10	3	5
R2	The product arrived on time and works as expected, although the battery life could be slightly better.	P101	4	"yesterday"	2	3
R3	The performance is decent, but I experienced occasional lag when running multiple apps at the same time.	P102	3.5	2026-02-01	"ten"	12

B. Completeness

Completeness refers to the presence of mandatory attributes as declared in metadata. In data-quality assessment literature, completeness is a core measurable dimension and is commonly operationalized through missing-value ratios (Pipino et al., 2002). ISO/IEC 25012 similarly identifies completeness as an inherent data-quality characteristic (Guerra-García et al., 2023).

Within this framework, completeness is syntactic when it is defined relative to formally declared requirements. If a digital VoC dataset specifies that each record must include a timestamp, rating, and product identifier, then missing values constitute non-conformance.

Table 3 illustrates typical completeness violations, where mandatory attributes defined in the metadata are missing. In these cases, mandatory fields such as Timestamp, Product_ID, and Rating are missing. Since these attributes are required by the

metadata specification, their absence constitutes syntactic non-conformance and reduces the usability of the dataset for monitoring purposes.

Completeness can be measured through a Mandatory Attribute Completeness Ratio (MACR). Let M be the set of attributes declared mandatory in the metadata specification. For each attribute $m \in M$, completeness can be computed as:

$$C_m = \frac{\text{Number of non-null values for } m}{N} \tag{2}$$

An overall completeness index can then be defined as:

$$MACR = \frac{1}{\mathrm{M}} \sum_{m \in M} C_m \qquad MACR \in [0,1] \tag{3}$$

Table 3. Example of completeness violations. The table reports review identifier (Review_ID), unstructured textual content (Review_Text), product identifier (Product_ID), customer rating (Rating), timestamp (Timestamp), and helpfulness indicators (Helpful_Votes, Total_Votes).

Review_ID	Review_Text	Product_ID	Rating	Timestamp	Helpful_Votes	Total_Votes
R4	Excellent product, very easy to use and the build quality feels solid. I would definitely recommend it.	P200	5	NULL	2	3
R5	Works well overall, but I had some issues with the initial setup which took longer than expected	NULL	4	2026-01-12	1	2
R6	The product is okay for basic use, but it does not meet expectations for more advanced tasks.	P202	NULL	2026-01-15	0	1

C. Format Validity

Format validity evaluates whether individual attribute values respect declared domains. Format validity concerns domain constraints and standardized representations. Data-cleaning research identifies domain violations as recurring classes of dirty data (Rahm & Do, 2000).

In Digital VoC datasets, format violations often arise in timestamp and categorical fields. A crawler may collect dates in ambiguous forms such as "03/04/2026" without day/month specification. If interpreted inconsistently in processing, such ambiguity can distort monitoring. Similarly, ratings outside the permitted domain (e.g., "6" in a five-point scale) represent domain violations. These defects are verifiable against declared format and domain rules and thus belong to syntactic quality. Table 4 provides examples of violations of domain and format constraints in digital VoC data.

These records violate predefined domains and representation formats. The Rating value exceeds the allowed range, and Timestamp values are expressed using inconsistent formats instead of the required standard. Such inconsistencies can lead to errors in aggregation and temporal analysis.

Format validity can be assessed using a Domain Compliance Rate (DCR), representing the proportion of attribute values that fall within the permitted domain and match declared formats. For a given attribute m with defined domain constraints:

$$DCR_m = \frac{\text{Number of values conforming to domain/format}}{\text{Total values of } m} \qquad DCR_a \in [0,1] \tag{4}$$

An aggregate format validity score can be computed across all constrained attributes:

$$DCR = \frac{1}{K} \sum_{k=1}^{K} DCR_{m_k} \qquad \mathrm{DCR} \in [0,1] \tag{5}$$

where K is the number of attributes with domain constraints.

Table 4. Example of format validity violations. The table reports review identifier (Review_ID), unstructured textual content (Review_Text), product identifier (Product_ID), customer rating (Rating), timestamp (Timestamp), and helpfulness indicators (Helpful_Votes, Total_Votes).

Review_ID	Review_Text	Product_ID	Rating	Timestamp	Helpful_Votes	Total_Votes
R7	The product exceeded my expectations in almost every aspect, especially performance and ease of use.	P300	6	2026-01-10	2	3
R8	Good overall experience, but the packaging was slightly damaged when it arrived.	P301	4	10/01/2026	1	2
R9	Not very satisfied with this purchase, as the device tends to overheat after extended use.	P302	2	Jan-10-2026	0	1

D. Entity Integrity

Entity integrity addresses uniqueness constraints and the absence of duplicate records. ISO 8000-8 explicitly recognizes that violations of entity integrity rules affect syntactic quality when uniqueness is formally specified (International Organization for Standardization, 2015). In data-quality research, duplicate detection and record linkage are established problem domains. The work of Fellegi and Sunter (1969) formalized record linkage as the identification of records representing the same real-world entity, while Elmagarmid et al. (2007) survey duplicate detection as a central data-quality challenge.

In Digital VoC environments, duplicates frequently arise from repeated crawling or syndicated reviews. If identical review texts are stored multiple times, sentiment frequencies and topic distributions may be artificially inflated. Such distortions can

trigger false monitoring alarms. Since uniqueness can be formally specified (e.g., one record per review identifier), duplicate presence constitutes a syntactic integrity violation.

To illustrate violations of entity integrity, Table 5 shows duplicate records that violate uniqueness constraints. The two records share the same Review_ID, violating the uniqueness constraint defined in the metadata. Such duplicates may arise from repeated data collection and can distort statistical analyses by artificially inflating the presence of specific reviews.

Entity integrity can be quantified through a Uniqueness Rate (UR). Let N_d be the number of detected duplicate records. Then:

$$UR = 1 - \frac{N_d}{N} \qquad \text{DR} \in [0,1] \tag{6}$$

Table 5. Example of entity integrity violations. The table reports review identifier (Review_ID), unstructured textual content (Review_Text), product identifier (Product_ID), customer rating (Rating), timestamp (Timestamp), and helpfulness indicators (Helpful_Votes, Total_Votes).

Review_ID	Review_Text	Product_ID	Rating	Timestamp	Helpful_Votes	Total_Votes
R10	This is one of the best products I've purchased recently, very reliable and easy to configure.	P400	5	2026-01-10	2	3
R10	This is one of the best products I've purchased recently, very reliable and easy to configure.	P400	5	2026-01-10	2	3

E. Cross-field Consistency

Cross-field consistency concerns referential and inter-attribute integrity constraints. Consistency is recognized as a measurable data-quality dimension in established assessment frameworks (Pipino et al., 2002), and integrity constraint enforcement is a foundational principle in database systems and data cleaning (Rahm and Do, 2000). ISO/IEC 25012 also identifies consistency as an inherent characteristic of data quality (Guerra-García et al., 2023).

In VoC datasets, cross-field inconsistencies occur when attribute combinations violate declared logical rules. For instance, a record may report more helpful votes than total votes, or it may indicate a specific platform while containing attributes exclusive to another platform. Table 6 reports examples of violations of cross-field integrity rules defined between related attributes. These records violate a temporal consistency constraint between related attributes. The Review_Date precedes the Purchase_Date, which is logically inconsistent under the assumption that a review can only be written after a purchase.

Cross-field consistency can be measured using an Integrity Violation Rate (IVR). Let R be the set of declared cross-attribute or referential integrity rules (e.g., helpful_votes ≤ total_votes). For each rule r, define:

$$IVR_r = \frac{\text{Number of records violating } r}{N} \qquad IVR_r \in [0,1] \tag{7}$$

An overall Consistency Score (CS) can be defined as:

$$CS = 1 - \frac{1}{\mathrm{R}} \sum_{r \in R} IVR_r \qquad CS \in [0,1] \tag{8}$$

Table 6. Example of cross-field consistency violations. The table reports review identifier (Review_ID), unstructured textual content (Review_Text), product identifier (Product_ID), review date (Review_Date), purchase date (Purchase_Date), and customer rating (Rating).

Review_ID	Review_Text	Product_ID	Review_Date	Purchase_Date	Rating
R11	I am very satisfied with the product, but it is strange that the review date appears before the purchase date.	P500	2026-01-10	2026-01-15	5
R12	The product works fine overall, but the recorded dates seem inconsistent with the actual purchase.	P501	2026-01-05	2026-01-20	4

F. Text Processability

Digital VoC differs from many traditional datasets due to the presence of unstructured textual content. Nevertheless, text fields remain subject to formal representation constraints. ISO 8000-8 frames syntactic quality in terms of conformance to formal expression rules (International Organization for Standardization, 2015). At the representation level, standards such as RFC 3629 (UTF-8) and the Unicode Standard define encoding and normalization rules that ensure consistent machine-readable representation.

In practice, encoding inconsistencies can distort analytics. A review containing the word "caffè" may appear in different Unicode forms or may be mis-decoded (e.g., "caffÃ¨"). Without normalization, tokenization and dictionary matching treat these variants as distinct terms, fragmenting frequency counts and affecting performance of text analytics.

Table 7 illustrates violations related to text encoding and normalization requirements. These records exhibit encoding inconsistencies, escaped Unicode sequences, and non-standard symbolic content. Such issues hinder text normalization and tokenization processes, affecting the reliability of downstream text analytics.

Text processability can be assessed through a Text Representation Compliance Rate (TRCR). This metric evaluates the proportion of textual records conforming to declared encoding and normalization standards. Let N_t be the number of records containing text, and N_p the number of text records successfully validated, then:

$$TRCR = \frac{N_p}{N_t} \qquad TRCR \in [0,1] \tag{9}$$

Table 7.Example of text processability violations. The table reports review identifier (Review_ID), unstructured textual content (Review_Text), product identifier (Product_ID), customer rating (Rating), timestamp (Timestamp), and helpfulness indicators (Helpful_Votes, Total_Votes).

Review_ID	Review_Text	Product_ID	Rating	Timestamp	Helpful_Votes	Total_Votes
R14	The batterÃ-a life is excellent and the device performs consistently even under heavy usage.	P600	5	2026-01-10	2	3
R15	The product is \u00E8asy to use and the interface is intuitive, although some features could be improved.	P601	4	2026-01-11	1	2
R16	👍 👍 👍 Works fine for basic tasks, but lacks some advanced functionalities expected at this price point.	P602	3	2026-01-12	0	1

VI. Data Cleaning for Improving Syntactic Quality in Digital VoC

In practical Digital VoC analytics, data cleaning is almost always performed. Records with obvious defects are removed, duplicates are filtered, missing values are handled, and textual preprocessing steps such as lowercasing or token normalization are applied. However, these activities are typically conducted in an ad hoc manner, rather than within a formally declared verification framework. Cleaning decisions are often motivated by practical convenience rather than by explicit conformance to declared metadata and integrity constraints.

This lack of structure creates two problems. First, cleaning procedures become difficult to reproduce and audit. Second, there is no explicit link between cleaning actions and formally defined data-quality requirements. As a result, sophisticated analytics can be applied to datasets whose syntactic reliability has never been systematically verified.

Within the ISO 8000 perspective adopted in this paper, syntactic quality is not an emergent property of preprocessing but the outcome of verification against formally specified rules. Data cleaning, therefore, should not be viewed as a heuristic preprocessing step, but as a structured verification-and-correction process aimed at restoring conformance to declared metadata and integrity constraints.

To operationalize syntactic quality improvement in Digital VoC datasets, this study proposes a structured four-stage protocol (see Figure 2).

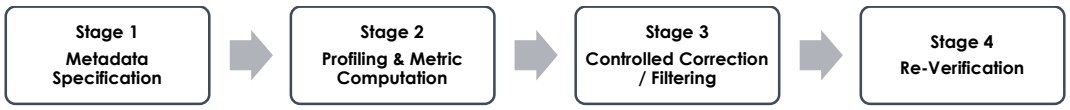


Figure 2. Proposed syntactic quality improvement protocol

A. Stage 1 – Metadata Specification

The protocol begins with the formal specification of metadata. Before any inspection or correction can be performed, the dataset must be associated with a declared structural definition. This includes the schema (attributes and data types), identification of mandatory fields, specification of permitted domains and format standards, explicit uniqueness constraints, cross-field integrity rules, and declared text representation requirements. Without such specification, it is not possible to distinguish between compliant and non-compliant records, and cleaning decisions inevitably become arbitrary.

To make Stage 1 operational and reproducible, Table 2 specifies the metadata components that must be formally declared.

Table 8. Stage 1: Metadata Specification Requirements for Syntactic Verification.

Metadata Component	What Must Be Defined
Data type Definition	List of attributes and associated data types (e.g., integer, string, datetime, boolean)
Mandatory Attributes	Identification of fields required for syntactic completeness (e.g., timestamp, rating, product ID)
Permitted Domains	Allowed value ranges or enumerations for constrained attributes (e.g., rating ∈ {1..5})
Format Standards	Required representation formats (e.g., ISO 8601 for timestamps, standardized language codes)
Uniqueness Constraints	Definition of primary key(s) or uniqueness rule (e.g., one record per reviewer ID)
Cross-field Integrity Rules	Logical consistency rules between attributes (e.g., helpful_votes ≤ total_votes)
Text Representation Standards	Encoding standard and normalization form for textual content (e.g., UTF-8, NFC normalization)

B. Stage 2 – Profiling and Violation Detection

Once the specification is established, the dataset is subjected to systematic profiling in order to detect violations. At this stage, the syntactic metrics introduced in Section V are computed. This step transforms qualitative concerns about "dirty data" into quantifiable metrics of structural non-conformance. Rather than directly modifying the dataset, the objective is first to measure the magnitude and distribution of defects across different constraint families. Profiling thus provides a diagnostic baseline that guides subsequent corrective actions.

C. Stage 3 – Controlled Correction or Filtering

Corrective interventions are subsequently carried out in a controlled manner. At this stage, each detected violation is addressed through actions that are explicitly linked to a declared syntactic constraint. Cleaning is therefore not guided by heuristic judgment or convenience, but by formal verification requirements established during metadata specification. The objective is to restore conformance to declared structural rules.

Corrective actions may include:

- Type harmonization, when structural validity is violated. This may involve aligning heterogeneous attribute names across sources, enforcing declared data types (e.g., converting numeric strings into integers when unambiguous), and removing records that cannot be parsed according to the declared schema.
- Removal or formally justified imputation of incomplete records, when mandatory attribute constraints are not satisfied. In contexts where a missing mandatory field cannot be deterministically reconstructed, records should be excluded rather than arbitrarily completed.
- Normalization of domain values and format standardization, in cases of format validity violations. This includes enforcing declared value ranges (e.g., rating bounds), converting dates to a standardized representation, and correcting enumerations when a one-to-one mapping to the declared domain is clearly defined.
- Duplicate detection and elimination according to declared uniqueness constraints, when entity integrity is compromised. Removal decisions should be based on explicit matching criteria (e.g., identical reviewer identifier).
- Enforcement of cross-field integrity constraints, when inter-attribute logical inconsistencies are detected. Records violating declared rules (e.g., helpful_votes exceeding total_votes) should either be corrected or excluded.
- Encoding normalization and text representation standardization, when textual processability requirements are not satisfied. This includes enforcing declared character encoding (e.g., UTF-8), applying Unicode normalization forms, and removing non-parseable or corrupted characters that violate representation standards.

D. Stage 4 – Re-Verification

Following correction, all syntactic metrics must be recomputed. By comparing pre- and post-cleaning metrics, it becomes possible to quantify improvement and evaluate the effectiveness of corrective actions. This step closes the verification loop and ensures that syntactic quality is demonstrably enhanced rather than assumed.

VII. Application Case

To illustrate the practical relevance of the proposed framework, a Digital VoC dataset concerning a smartphone in the consumer electronics sector was analyzed. The dataset was constructed from publicly available online customer reviews posted on three major e-commerce and product-review platforms. Reviews were collected through a structured web-scraping procedure over a 12-month observation window, using the smartphone model identifier and related product-page URLs as selection criteria. The final raw dataset included 52,438 reviews. The dataset was selected because it reflects a typical post-market monitoring scenario in which customer feedback is integrated from heterogeneous external sources rather than generated within a controlled internal quality system. Each record contains the following attributes: Review_ID, Reviewer_ID, Platform, Product_ID, Rating, Review_Text, Timestamp, Helpful_Votes, and Total_Votes. The objective of this application is not only to assess syntactic quality, but to demonstrate how syntactic defects can affect the reliability of product quality monitoring.

A. Stage 1 – Metadata Specification

A formal metadata specification was defined prior to analysis, including: (i) rating domain (integer $\in \{1–5\}$), (ii) timestamp format (ISO 8601), (iii) mandatory attributes (Review_ID, Product_ID, Rating, Timestamp), (iv) uniqueness constraint (one record per Review_ID), (v) cross-field consistency rules, and (vi) UTF-8 encoding with NFC normalization for textual data. This specification provides the formal reference for distinguishing compliant from non-compliant records.

B. Stage 2 – Profiling and Impact of Syntactic Defects

The dataset was profiled against the declared constraints. Several categories of syntactic defects were identified. *Structural validity* issues included ratings expressed as strings (e.g., "five") or non-integer values (e.g., 3.5). *Format violations* were observed in timestamps represented in inconsistent forms (e.g., "08/14/2025" instead of ISO format). *Completeness* issues mainly involved missing timestamps (5.6% of records). *Entity integrity* violations were due to duplicate Review_IDs (4.1%), primarily caused by repeated crawling. *Text processability* problems included encoding errors (e.g., "batterÃa", "\u00E8") and non-normalized Unicode strings. Overall, only 86.7% of records satisfied all syntactic constraints.

These defects were not only structural but had a direct impact on monitoring outputs. This effect becomes particularly relevant in manufacturing contexts, where digital VoC data is used to identify product-related defects and support engineering decisions. For example, several reviews referring to battery overheating issues were duplicated due to repeated data extraction, artificially amplifying the perceived frequency of this defect. Similarly, inconsistent timestamps distorted the temporal distribution of complaints related to device malfunctions, creating misleading signals of sudden quality degradation. In addition, format and encoding inconsistencies fragmented defect-related information. For instance, reviews mentioning "battery issue", "battery problem", and mis-encoded variants (e.g., "batterÃ-a issue") were treated as distinct terms prior to

normalization, reducing the accuracy of defect aggregation and topic identification. These issues illustrate how syntactic defects can directly affect the identification, quantification, and prioritization of product quality problems in manufacturing environments.

C. Stage 3 – Controlled Correction

Corrective actions were applied in relation to the violated constraints. Type casting harmonization were applied to 1,214 records to restore structural validity. A total of 2,937 records with missing mandatory attributes were removed. Domain and format violations were addressed by excluding 947 out-of-range ratings and standardizing timestamp representations. Entity integrity was restored by removing 1,508 duplicate records. Cross-field inconsistencies (936 records) were either corrected (421 cases) or removed. Finally, 2,468 records required text normalization to ensure proper encoding and representation. After correction, the dataset size decreased from 52,438 to 47,526 records (−9.4%). All interventions were directly derived from the metadata specification and applied in a reproducible manner.

D. Stage 4 – Re-Verification and Impact on Monitoring

Following correction, syntactic quality metrics were recomputed. The results are summarized in Table 9.

Table 9. Syntactic Quality Metrics Before and After Cleaning

Metric	Pre-Cleaning	Post-Cleaning	Change
Type Conformance Rate (TCR)	0.962	0.998	+3.6%
Mandatory Attribute Completeness Ratio (MACR)	0.944	1.000	+5.6%
Domain Compliance Rate (DCR)	0.960	1.000	+4.0%
Uniqueness Rate (UR)	0.971	1.000	+2.9%
Compliance Score (CS)	0.982	0.997	+1.5%
Text Representation Compliance Rate (TRCR)	0.953	0.999	+4.6%

All syntactic dimensions improved after correction. The proportion of fully compliant records increased from 86.7% to 99.4%. More importantly, the removal of syntactic defects led to more stable and reliable monitoring signals. In particular, the reduction of duplicate records decreased artificially inflated defect frequencies, while timestamp normalization reduced noise in time-based analyses. This resulted in more consistent product-level comparisons and a more reliable identification of potential quality issues. This application shows that syntactic verification is not merely a preprocessing step, but a necessary condition for reliable Digital VoC-based product quality monitoring. Without such verification, monitoring systems may generate misleading signals, potentially leading to incorrect engineering decisions, such as unnecessary investigations or missed detection of emerging defects.

VIII. Conclusions

This paper addressed the problem of digital VoC data reliability in product quality monitoring by adopting a data-quality perspective grounded in ISO 8000. While prior research has extensively explored methods for extracting insights from digital VoC, limited attention has been devoted to the explicit verification of the quality of the underlying data. The main contribution of this study is the development of a structured and operational framework for assessing the syntactic quality of digital VoC datasets. In particular, the paper provides three original elements. First, it translates the ISO 8000 conceptualization of data quality into the specific context of digital VoC, clarifying the role of syntactic quality as a prerequisite for reliable analytics. Second, it proposes a set of six syntactic subdimensions—structural validity, completeness, format validity, entity integrity, cross-field consistency, and text processability—that organize the main categories of formally verifiable defects in VoC data. Third, it operationalizes these subdimensions through measurable metrics and integrates them into a structured four-stage protocol for syntactic verification and improvement. These subdimensions are not intended as an exhaustive taxonomy of data quality, but as an operational decomposition of ISO 8000 syntactic conformance into verifiable metadata requirement classes for Digital VoC datasets.

The application case demonstrated the practical relevance of the proposed approach, showing that non-negligible levels of syntactic defects may be present even in large-scale VoC datasets and that systematic verification can substantially improve data compliance prior to analysis. In this sense, the framework contributes to strengthening the reliability and auditability of digital VoC-based quality monitoring.

The study has some limitations. First, it focuses exclusively on syntactic quality and does not address semantic aspects, such as the correctness or relevance of review content, nor pragmatic aspects related to fitness for specific decision-making contexts. Second, although the proposed metrics are formally defined, the identification of acceptable thresholds and governance rules remains context-dependent and requires further investigation.

Future research may extend this work by integrating semantic and pragmatic dimensions into a comprehensive reliability assessment framework for digital VoC. Additional studies could explore cross-domain applications, benchmark ranges for syntactic quality metrics, and the integration of syntactic verification into automated VoC analytics pipelines for Quality 4.0 systems.

References

Barravecchia, F., Franceschini, F., & Mastrogiacomo, L. (2018). A service network perspective to evaluate service matching in early design. *Journal of Service Theory and Practice*, *28*(3), 356-383.

Barravecchia, F., Mastrogiacomo, L., & Franceschini, F. (2022a). KA-VoC map: classifying product key-attributes from digital voice-of-customer. *Quality Engineering*, *34*(3), 344-358.

Barravecchia, F., Mastrogiacomo, L., & Franceschini, F. (2022b). Digital voice-of-customer processing by topic modelling algorithms: insights to validate empirical results. *International Journal of Quality & Reliability Management*, *39*(6), 1453-1470.

Barravecchia, F., Mastrogiacomo, L. and Franceschini, F. (2023) 'Product quality tracking based on digital Voice-of-Customers', Total Quality Management & Business Excellence, 34 (11–12), 1386–1409.

Barravecchia, F., Mastrogiacomo, L. and Franceschini, F. (2026) 'Digital VoC analysis for product/service quality tracking in the era of Quality 4.0', International Journal of Quality & Reliability Management, 43 (2), 499–515.

Broday, E. E. (2022) 'The evolution of quality: from inspection to quality 4.0', International Journal of Quality and Service Sciences, 14 (3), 368–382.

Elmagarmid, A. K., Ipeirotis, P. G. and Verykios, V. S. (2007) 'Duplicate record detection: A survey', IEEE Transactions on knowledge and data engineering, 19 (1), 1–16.

Esposito, C., Moscato, V. and Sperlí, G. (2021) 'Trustworthiness assessment of users in social reviewing systems', IEEE Transactions on Systems, Man, and Cybernetics: Systems, 52 (1), 151–165.

Fellegi, I. P. and Sunter, A. B. (1969) 'A theory for record linkage', Journal of the American statistical association, 64 (328), 1183–1210.

Fox, C., Levitin, A. and Redman, T. (1994) 'The notion of data and its quality dimensions', Information processing & management, 30 (1), 9–19.

Franceschini, F., Galetto, M., Maisano, D., & Neely, A. D. (2019). Designing performance measurement systems, Theory and Practice of Key Performance Indicators. *Springer, Cham (Switzerland)*

Guerra-García, C., Nikiforova, A., Jiménez, S., Perez-Gonzalez, H. G., Ramírez-Torres, M. and Ontañon-García, L. (2023) 'ISO/IEC 25012-based methodology for managing data quality requirements in the development of information systems: Towards Data Quality by Design', Data & Knowledge Engineering, 145 , 102152.

Hu, N., Bose, I., Koh, N. S. and Liu, L. (2012) 'Manipulation of online reviews: An analysis of ratings, readability, and sentiments', Decision support systems, 52 (3), 674–684.

International Organization for Standardization (2011) 'ISO 8000-1: Data quality — Part 1: Overview'.

International Organization for Standardization (2015) 'ISO 8000-8: Data quality — Part 8: Information and data quality: Concepts and measuring'.

Kakde, S. T., Roychowdhury, S., Bhosale, A. T. and Maiti, J. (2024) 'CPAN chart: A novel customer perception analysis system using natural language processing and attribute control charting', IEEE Transactions on Engineering Management, 71 , 7609–7622.

Kim, R. Y. (2021) 'Using online reviews for customer sentiment analysis', IEEE Engineering Management Review, 49 (4), 162–168.

Kim, W. G. and Park, S. A. (2017) 'Social media review rating versus traditional customer satisfaction: Which one has more incremental predictive power in explaining hotel performance?', International Journal of Contemporary Hospitality Management, 29 (2), 784–802.

Mastrogiacomo, L., Barravecchia, F., Franceschini, F., & Marimon, F. (2021). Mining quality determinants of product-service systems from user-generated contents. *Quality Engineering*, *33*(3), 425-442.

Mewada, A. and Dewang, R. K. (2023) 'A comprehensive survey of various methods in opinion spam detection', Multimedia Tools and Applications, 82 (9), 13199–13239.

Özdağoğlu, G., Kapucugil-İkiz, A. and Çelik, A. F. (2018) 'Topic modelling-based decision framework for analysing digital voice of the customer', Total Quality Management & Business Excellence, 29 (13–14), 1545–1562.

Park, S., Joung, J. and Kim, H. (2026) 'Large Language Model-Based Online Review Classification for Subfeature-Level Customer Opinion Analysis', Journal of Mechanical Design, 148 (4), 041701.

Pipino, L. L., Lee, Y. W. and Wang, R. Y. (2002) 'Data quality assessment', Communications of the ACM, 45 (4), 211–218.

Rahm, E. and Do, H. H. (2000) 'Data cleaning: Problems and current approaches', IEEE Computer Society Data Engineering Bulletin, 23 (4), 3–13.

Ridzuan, F. and Zainon, W. M. N. W. (2024) 'A review on data quality dimensions for big data', Procedia computer science, 234 , 341–348.

Santhiran, R., Varathan, K. D. and Chiam, Y. K. (2024) 'Feature extraction from customer reviews using enhanced rules', Peerj Computer Science, 10 , e1821.

See-To, E. W. K. and Ngai, E. W. T. (2018) 'Customer reviews for demand distribution and sales nowcasting: a big data approach', Annals of Operations Research, 270 (1), 415–431.

Shankaranarayanan, G. and Cai, Y. (2006) 'Supporting data quality management in decision-making', Decision Support Systems, 42 (1), 302–317.

Subroto, A. and Christianis, M. (2021) 'Rating prediction of peer-to-peer accommodation through attributes and topics from customer review', Journal of Big Data, 8 (1), 9.

Ullah, R., Amblee, N., Kim, W. and Lee, H. (2016) 'From valence to emotions: Exploring the distribution of emotions in online product reviews', Decision Support Systems, 81 , 41–53.

Wand, Y. and Wang, R. Y. (1996) 'Anchoring data quality dimensions in ontological foundations', Communications of the ACM, 39 (11), 86–95.

Wang, J., Liu, Y., Li, P., Lin, Z., Sindakis, S. and Aggarwal, S. (2024) 'Overview of data quality: Examining the dimensions, antecedents, and impacts of data quality', Journal of the Knowledge Economy, 15 (1), 1159–1178.

Wang, R. Y. and Strong, D. M. (1996) 'Beyond accuracy: What data quality means to data consumers', Journal of management information systems, 12 (4), 5–33.

Yan, L., Tao, B., Han, Z. and Ouyang, L. (2025) 'Online reviews analysis in product defects and customer requirements via two-stage model', Total Quality Management & Business Excellence, 36 (7–8), 788–810.

Zonnenshain, A. and Kenett, R. S. (2020) 'Quality 4.0—the challenging future of quality engineering', Quality Engineering, 32 (4), 614–626.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Analysis of Non-Conformities Raised in ISO 9001 Audits in Portuguese Municipal Chambers

Roberth Pereira [1], Pedro Domingues[2] and Carina Pimentel [3]

1 University of Minho, Master's Degree in Engineering and Quality Management ; roberth.ufma@hotmail.com
2 University of Minho, Professor and Researcher at the School of Engineering of the University of Minho; pdomingues@dps.uminho.pt
3 University of Minho, Professor and Researcher at the School of Engineering of the University of Minho¸ carina.pimentel@dps.uminho.pt

***Abstract*— The present study investigated the non-conformities (NCs) raised in external audits of the ISO 9001:2015 standard in Portuguese municipal councils during the period from 2019 to 2023. The general objective of the study was to identify and understand the dynamics of the NCs found in the audits. Through a quantitative and qualitative analysis of the audit reports and questionnaires applied to Quality managers, the results identified that the most problematic clauses were related to the management of documented information (7.5), monitoring and measuring performance (9.1), and production and service provision (8.5). The main difficulties encountered included inadequacies in the control of documented information, a lack of effective mechanisms for monitoring and analyzing processes, and failures in the management and control of service delivery operations. These issues highlight the need for greater attention to the structuring and management of processes, as well as continuous training for employees involved in Quality management systems. The study also emphasized the importance of committed leadership and effective communication for the successful implementation of the ISO 9001 standard in public institutions. Based on the presented evidence, it is recommended that municipal councils invest in capacity-building strategies and improvements in their Quality management systems, with the aim of reducing NCs and promoting more efficient and transparent management. In summary, this work contributes to the understanding of the dynamics of non-conformities and the challenges of Quality management in the public sector, offering valuable insights for continuous improvement and maintaining compliance with the ISO 9001 standard.**

Keywords – Non-Conformity, ISO 9001:2015, Municipal Chambers and Auditing.

I. Introduction

In an increasingly competitive and dynamic organisational context, quality plays a strategic role in promoting competitive advantage and organisational sustainability, requiring an effective response to customers' needs and expectations. In this regard, the adoption of Quality Management Systems (QMS), particularly through ISO 9001:2015 certification, has become a fundamental practice for ensuring process standardisation, preventing non-conformities (NC), and promoting continuous organisational improvement (ISO, 2015; Pires, 2007).

ISO 9001:2015 establishes requirements aimed at the identification, recording, and treatment of non-conformities, reinforcing the implementation of corrective actions and the prevention of recurring failures, thereby promoting a systematic and preventive approach to quality management (Vieira, 2022). In this context, non-conformities represent deviations from established requirements and are often associated with processes that generate unsatisfactory outcomes (Campos, 2013).

In Public Administration, particularly within municipalities, the implementation of QMS has contributed to more efficient, transparent, and continuously improving management practices, with compliance being assessed through internal and external audits that identify non-conformities and opportunities for improvement (Marques, 1997).

Despite the growing adoption of ISO 9001 within Portuguese Public Administration, research on non-conformities identified during external audits remains limited. Existing studies have focused primarily on the benefits of certification and the critical success factors for the implementation of Quality Management Systems, while knowledge regarding the most frequent non-conformities and the challenges faced by municipalities in maintaining compliance with the standard's requirements remains scarce. López-Lemus et al. (2023) highlight that significant challenges and underexplored aspects still exist regarding the implementation of ISO 9001:2015-based Quality Management Systems in the public sector, emphasizing the need for further research in this area. In this context, it becomes relevant to analyse the non-conformities identified in external audits, contributing to the identification of critical areas for improvement and to strengthening the effectiveness of Quality Management Systems within the municipal sector.

Therefore, the present study aims to analyse the dynamics of non-conformities identified in Portuguese municipalities certified under ISO 9001:2015 between 2019 and 2023, seeking to generate knowledge that contributes to the improvement of certification processes and quality management practices within the municipal sector. The research covered 41 municipalities and involved statistical data analysis using the Statistical Package for the Social Sciences (SPSS).

II. Literature Review

ISO 9001 is the leading international reference standard for Quality Management Systems (QMS), promoting continuous improvement, process control, and stakeholder satisfaction (Hoyle, 2017). The ISO 9001:2015 version strengthened the

integration of quality management into organisational strategy by introducing a greater emphasis on risk-based thinking and leadership (ISO, 2015).

In Public Administration, the implementation of Quality Management Systems has contributed to administrative modernisation, improved service efficiency, increased transparency, and a stronger citizen-oriented approach (Poksinska, 2010; Moumene & Benhrimida, 2025). At the municipal level, quality management has been recognised as a relevant instrument for enhancing reliability, responsiveness, and citizens' trust in local public services (Rivera-Fernández & Fragoso-Martínez, 2025). Existing studies have mainly focused on the benefits of certification and the challenges associated with its implementation (Boiral, 2012; Oliveira, 2018).

However, knowledge regarding non-conformities identified during external audits remains limited, particularly concerning their frequency, distribution across the clauses of the standard, and evolution over time. Analysing these aspects may contribute to identifying weaknesses in Quality Management Systems and supporting the development of improvement actions.

In this context, the present study seeks to answer the following research questions:

RQ1: What are the most frequent non-conformities identified in external ISO 9001:2015 audits conducted in Portuguese municipalities?

RQ2: Which clauses of ISO 9001:2015 are most frequently associated with the occurrence of non-conformities?

RQ3: How did the identified non-conformities evolve between 2019 and 2023?

III. Methodology

The database used in this research consisted of 21 external audit reports of Quality Management Systems (QMS) conducted in accordance with the ISO 9001:2015 standard, as well as information collected through questionnaires administered to Portuguese municipalities certified under the same standard. Based on data obtained from the Certified Companies Directory (Guia de Empresas Certificadas – GEC), 77 Portuguese municipalities holding ISO 9001:2015 certification were identified. Initially, efforts were made to obtain audit reports from certified municipalities and their respective certification bodies. However, due to confidentiality concerns and limitations regarding access to such documentation, only 21 audit reports covering the period from 2019 to 2023 were obtained.

To complement the available information and enhance the representativeness of the study, a questionnaire was developed and distributed to the 77 identified municipalities. A total of 41 valid responses were received, corresponding to a response rate of 53.2%. Data were collected anonymously, without identifying the participating entities. Prior to the distribution of the questionnaire, a pilot test was conducted with three municipalities to assess the clarity and adequacy of the questions. After incorporating the necessary adjustments, the questionnaire was distributed to the remaining certified municipalities. Data collection took place between March and June 2024. Quality managers were selected as respondents due to their knowledge of the processes and the operation of the Quality Management Systems within their respective organisations.

The selected external audit reports covered the period between 2019 and 2023. The analysis focused on identifying the most frequent non-conformity (NC) themes and the clauses of the standard with the highest number of NC occurrences. Each report included information regarding the audited organisation, the type of audit conducted, the certification scope, recommendations for decision-making, requirements considered not applicable, identification of NCs according to NP EN ISO 9001:2015, an audit summary highlighting strengths and areas for improvement, details of NCs identified by clause/subclause, and the corresponding opportunities for improvement (OFIs).

To analyse the collected data, both qualitative and quantitative approaches were employed. The combination of these methods provided a broader and more robust analysis of the phenomenon under study. Statistical analysis of the quantitative data was performed using the Statistical Package for the Social Sciences (SPSS) software.

IV. Results and discussion

This chapter presents and discusses the results obtained from the study. To address the research objectives and examine the research hypothesis, a series of analyses was conducted, and their results are presented in the following sections.

Figure 1 presents the geographical distribution of the sample, namely the percentage of municipalities belonging to each district and autonomous region, considering that the sample consists of 41 municipalities. The sample is composed of municipalities certified under ISO 9001:2015, regardless of the scope of certification.

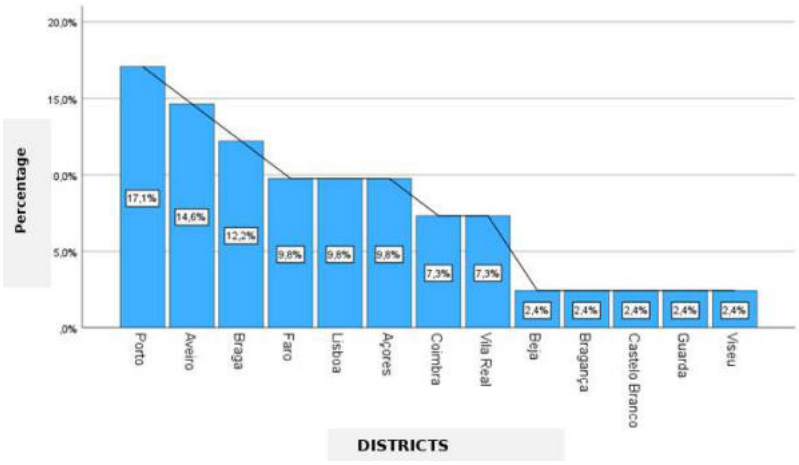


Fig. 1. Percentage of municipalities in the sample by district/autonomous region.

Analysis of Figure 1 shows that the Porto district has the highest percentage (17.1%) of certified municipalities included in the sample. The implementation of ISO 9001 is frequently observed in larger districts. Boiral (2012) highlights that in large cities, where demand for public services is higher, ISO 9001 certification helps standardise processes, increasing customer satisfaction and promoting more efficient resource management.

Among the 82 ISO 9001:2015 certification scopes and the number of municipalities certified in each scope, customer service was the area with the highest number of certified municipalities (18), suggesting a strong concern with service quality, namely face-to-face service (12), one-stop shop service (4), online service (1), and call centre service (1). This reflects the growing demand for more efficient and accessible public services for the population. Marcondes (2018) states that "ISO 9001 certification in customer service has been widely adopted in the service sector and Public Administration, especially to ensure greater transparency, standardisation, and better fulfilment of citizens' needs."

Figure 2 presents the responses of municipal quality managers regarding their familiarity with the non-conformity treatment procedures established by ISO 9001.

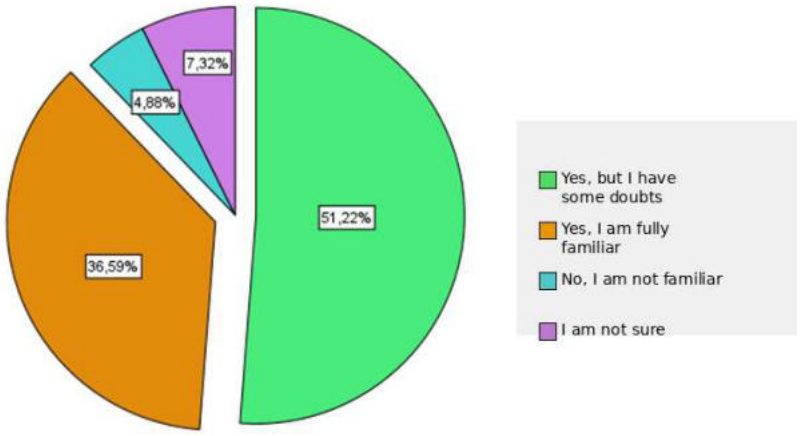


Fig. 2. Familiarity of municipal quality managers with the non-conformity treatment procedures established by ISO 9001.

Based on Figure 2, it was observed that 51.22% indicated that they are familiar with the procedures but still have some doubts. This demonstrates a solid knowledge base; however, it also highlights the need for additional clarification to ensure full understanding and confidence in the implementation of these procedures, which may be mitigated through continuous training initiatives. According to Poksinska (2010), the lack of adequate training and resistance to change are also factors that contribute to uncertainty during implementation.

In the following Figures (3, 4, 5, 6, and 7), the clauses with the highest percentage of non-conformities identified in external audits are presented, highlighting the areas in which municipalities experienced the greatest difficulties in complying with the requirements of ISO 9001:2015 during the years 2019, 2020, 2021, 2022, and 2023.

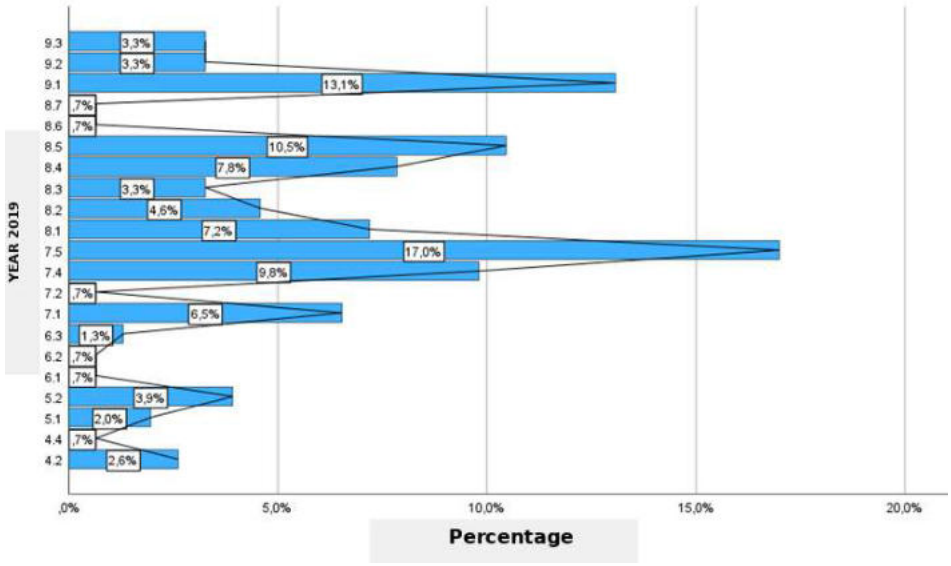


Fig. 3. Clauses with the highest number of non-conformities in 2019.

In Figure 3, Clause 7.5 (Documented Information) was the clause with the highest number of non-conformities identified in municipalities during 2019, corresponding to 17% of the non-conformities recorded in ISO 9001:2015 audits.

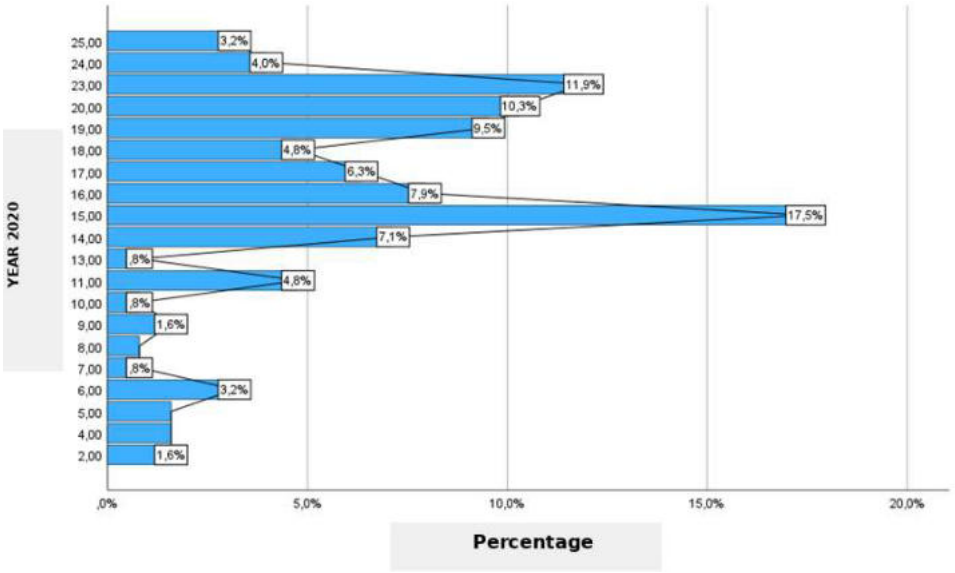


Fig. 4. Clauses with the highest number of non-conformities in 2020.

In 2020 (Figure 4), it can be observed that the clauses with the highest number of non-conformities identified were the same as those in 2019 (Clauses 7.5, 9.1, and 8.5), although with slightly higher and lower incidence levels compared to the previous year.

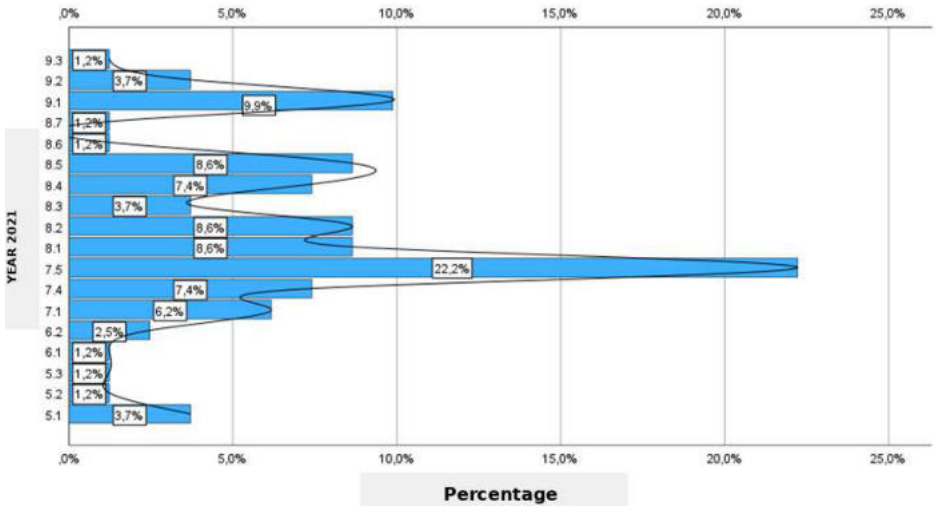


Fig. 5. Clauses with the highest number of non-conformities in 2021.

In 2021 (Figure 5), it can be observed that the clauses with the highest number of non-conformities identified were also broadly similar to those found in previous years (2019 and 2020), although Clause 7.5 (Documented Information) presented a comparatively higher percentage, corresponding to 22.2%.

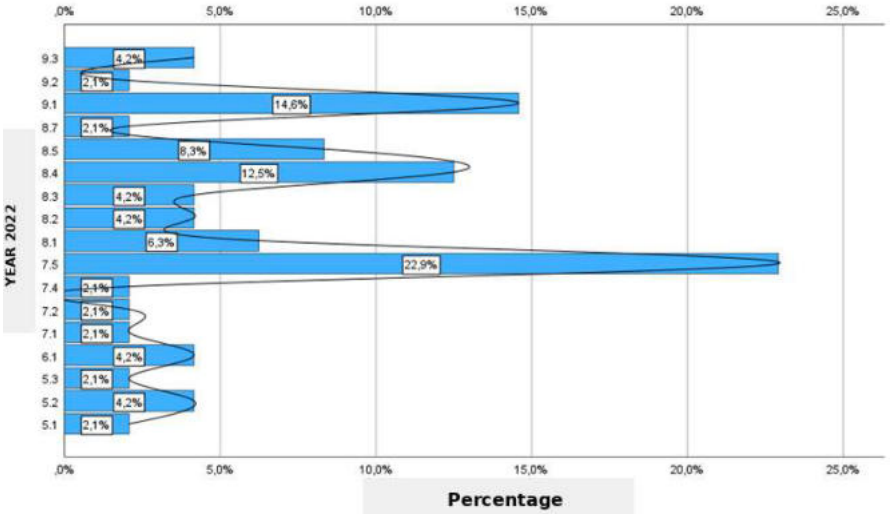


Fig. 6. Clauses with the highest number of non-conformities in 2022.

In the year 2022 (Figure 6), it is observed that the clauses with the highest number of NCs were clauses 7.5 (22.9%), 9.1 (14.6%), and 8.4 (Control of externally provided processes, products and services) (12.5%).

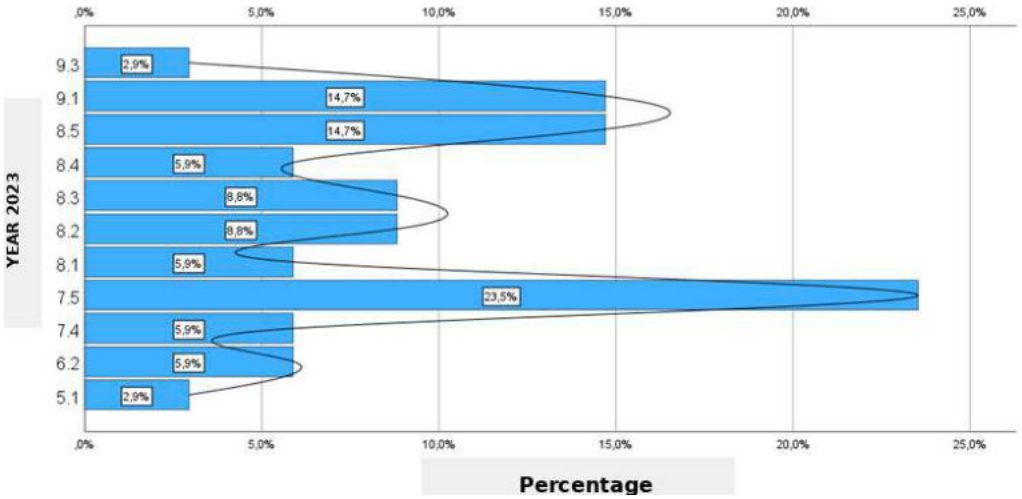


Fig. 7. Clauses with the highest number of non-conformities in 2023.

In 2023 (Figure 7), the clauses with the highest number of non-conformities were Clause 7.5 (23.5%), Clause 9.1 (14.7%), and Clause 8.5 (14.7%).

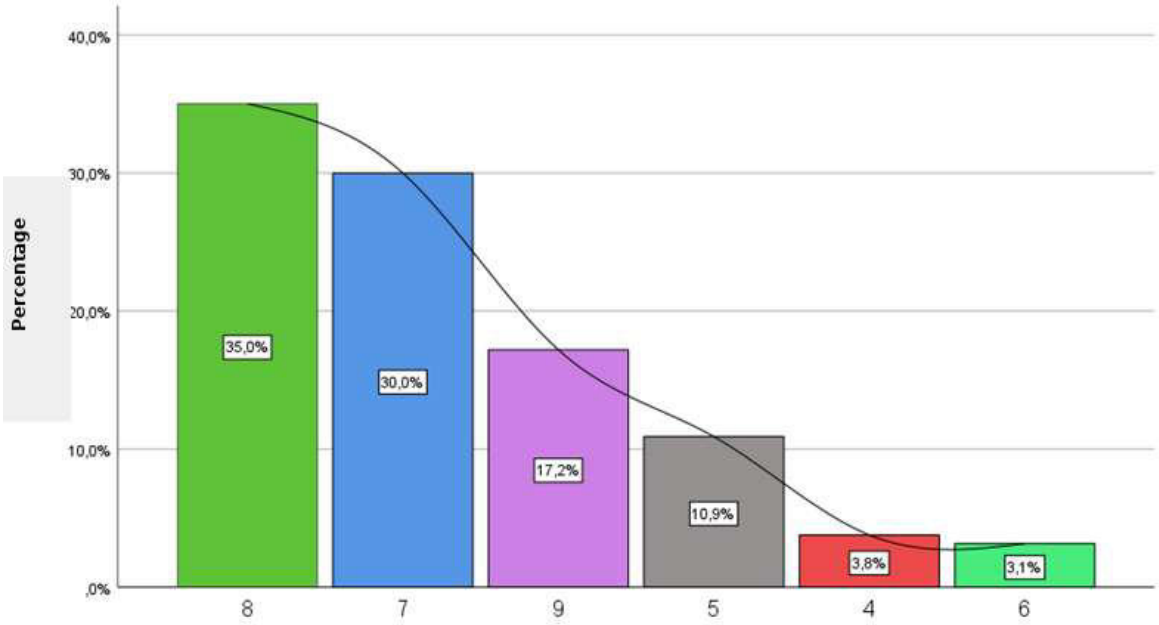
Analysing the non-conformities across all years (2019 to 2023), the data indicate that the most frequent non-conformities identified in ISO 9001:2015 audits during this period were related to Clauses 7.5 (Documented Information), 9.1 (Monitoring, Measurement, Analysis and Evaluation), 8.5, and 8.4. The high frequency of non-conformities associated with Clause 7.5 indicates that many municipalities experienced difficulties in organising, maintaining, and controlling the documentation required for the QMS. This may include issues related to document creation, version control, filing, and accessibility.

In this study, Clause 7.5 recorded the highest number of non-conformities (85 throughout the analysed period), representing 19.23% of the total. This reflects the importance of ensuring that the documentation required for ISO 9001 implementation is maintained accurately and kept up to date. Hoyle (2017) discusses the importance of documentation (Clause 7.5), emphasising that failure to maintain adequate and updated records may compromise the entire QMS. According to the same author, documentation must be accessible to ensure process traceability and compliance with ISO 9001 requirements, particularly during external audits. Non-conformities related to Clause 9.1 indicate challenges in the ability to monitor, measure, and analyse processes to ensure compliance with requirements.

The predominance of non-conformities associated with documented information and process control suggests that the implementation of ISO 9001 continues to require robust monitoring and continuous improvement mechanisms. These findings are consistent with Moumene and Benhrimida (2025), who highlight ISO 9001 as a tool for transforming public services through the systematic improvement of organisational processes.

Rodríguez-Antón (2020) highlights the crucial role of monitoring, measurement, and evaluation (Clause 9.1) in ensuring that continuous improvement processes are effective. Problems and non-conformities related to Clause 8.5 may indicate difficulties in the consistent delivery of services provided by municipalities. Zimon (2020) also addresses the issue of production and service provision (Clause 8.5), noting that the integration of external suppliers into the organisation's QMS is essential. The same author found that organisations that fail to implement robust verification processes for externally provided products and services tend to face recurring non-conformities during audits.

Figure 10 illustrates the distribution of non-conformities identified in external ISO 9001:2015 audits in Portuguese municipalities, by chapter of the standard, throughout the period from 2019 to 2023.

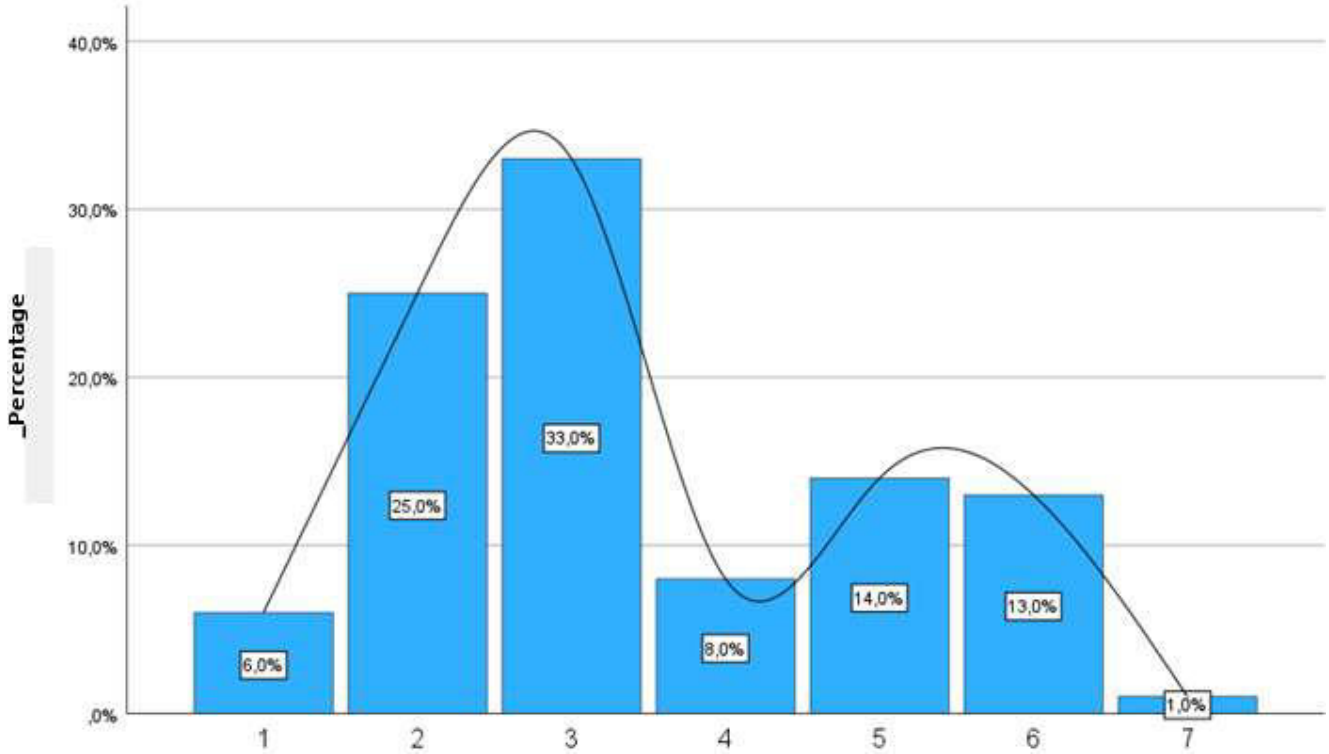


Legend: 8 - Operation; 7 - Support; 9 - Performance Evaluation; 5 - Leadership; 4 - Context; 6 - Planning.

Fig. 8. Percentage of non-conformities by chapter identified in external ISO 9001:2015 audits in municipalities.

Figure 10 shows that the highest percentages of non-conformities are concentrated in Chapters 8 (35%), 7 (30%), and 9 (17%), which correspond to crucial aspects of ISO 9001:2015, such as documentation, operations, resource management, and performance monitoring. These results are consistent with the study conducted by Sampaio (2008), which found higher percentages of non-conformities in Chapter 7, which in the current ISO 9001:2015 version has been reorganised and redistributed across Chapters 7 and 8.

Figure 11 illustrates the percentage distribution of the reasons that prioritized the resolution of nonconformities in a functional department. The reasons are listed from 1 to 7 and range from the opinion of the Quality Manager onwards.



Legend: 1 - The opinion of the Quality Manager; 2 - The potential impact on product or service quality; 3 - The financial cost associated with correcting the NC; 4 - The risks associated with each NC; 5 - Feedback from customers or stakeholders; 6 - Opportunities for improvement; 7 - None of the above.

Fig. 9. . Reasons for prioritizing the resolution of nonconformities.

As shown in Figure 11, the most relevant reason, accounting for 33% of the votes, was the financial cost associated with correcting the nonconformity (3), followed by the potential impact on product or service quality (2), with 25%. According to Balassiano (2023), nonconformities result in increased operational costs due to failures such as rework, material waste, and production downtime. These costs may increase further when organizations fail to identify and correct the root causes of failures, leading to the recurrence of problems.

Figure 12 shows whether municipalities are distinguished or recognized for holding ISO 9001:2015 certification.

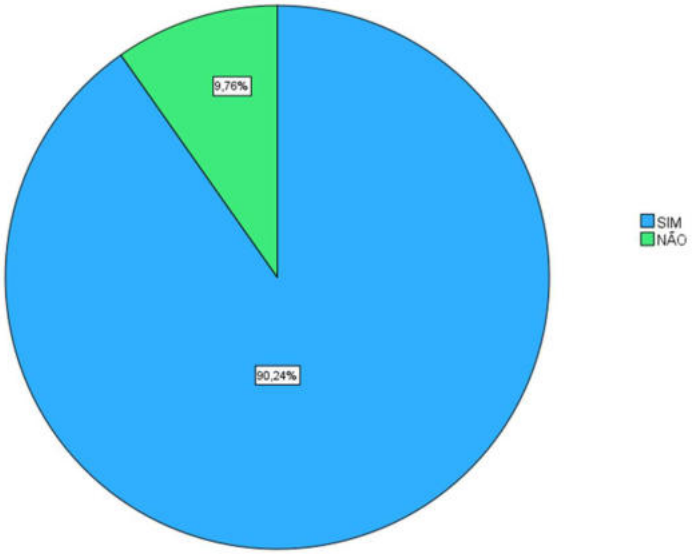


Fig. 10. .Percentage of municipalities that are distinguished and recognized for holding ISO 9001:2015 certification.

According to the data presented in Figure 12, 90.24% of municipalities state that they are distinguished, valued, and recognized for holding this certification, while 9.76% indicate that they do not receive such recognition.

This high percentage of positive recognition suggests that ISO 9001:2015 certification has a significant impact on the public and institutional perception of municipalities. As noted by Oliveira (2018), "ISO 9001:2015 certification in public institutions is a quality milestone that contributes to the credibility and trust of citizens, while also strengthening institutional image." The recognition given to certified municipalities reflects the value attributed to efficient management and commitment to the continuous improvement of processes, which may bring benefits to both the administration and the citizens.

V. FUTURE RESEARCH

Future research could further explore recurring non-conformities identified in external audits, distinguishing between major and minor non-conformities. Comparative studies involving other public sector organisations may help determine whether the challenges identified in municipalities are common across public administration. Additionally, longitudinal studies could assess the effectiveness of corrective actions over time. Further research may also examine the impact of continuous training, leadership, organisational culture, and continuous improvement practices on reducing non-conformities and enhancing the effectiveness of Quality Management Systems in the public sector.

VI. CONCLUSIONS

The present study analysed the non-conformities identified in external ISO 9001:2015 audits conducted in Portuguese municipalities between 2019 and 2023, demonstrating that certification contributed positively to improving quality management, operational efficiency, and citizen satisfaction.

The clauses with the highest incidence of non-conformities were Clause 7.5 (Documented Information), Clause 9.1 (Monitoring, Measurement, Analysis and Evaluation), Clause 8.5 (Production and Service Provision), and Clause 8.4 (Control of Externally Provided Processes, Products and Services), with Clause 7.5 standing out as the most recurrent throughout the analysed period.

The results revealed an overall downward trend in non-conformities over the years, suggesting a positive evolution in the maturity of the implemented management systems. However, the recurrence of certain non-conformities in both consecutive and non-consecutive audits was observed, particularly those associated with document management, indicating persistent difficulties in consolidating effective corrective actions.

It is concluded that ISO 9001 constitutes an important instrument for modernisation and continuous improvement in Public Administration, contributing to more efficient, transparent, and citizen-oriented management. However, the effectiveness of

certification depends on organisations' ability to consolidate corrective practices, strengthen monitoring mechanisms, and ensure proper document management.

VII. REFERENCES

Balassiano, R.S., I.W.E. and N.J.M. (2023), "Efeitos das práticas de ESG no custo de capital das empresas brasileiras", *REUNIR - Revista de Administração, Contabilidade e Sustentabilidade*, pp. 197–217.

Boiral, O. (2012), "Quality of the ISO 9000 series of standards: Perceptions of quality management experts", *International Journal of Quality & Reliability Management*.

Campos, V.F. (2013), *Gestão da Qualidade Total*, Atlas, São Paulo.

Deming, W.E. (1982), *Out of the Crisis*, MIT Press, Cambridge.

Fonseca, L.M. and Domingues, J.P. (2017), "Reliable and Flexible Quality Management Systems in the Automotive Industry: Monitor the Context and Change Effectively", *Procedia Manufacturing*, Vol. 11, pp. 1200–1206, available at: https://doi.org/10.1016/j.promfg.2017.07.245

Hoyle, D. (2017), *ISO 9001:2015: Understand, Implement, Succeed*, Routledge, New York, NY.

ISO (2015), *NP EN ISO 9000:2015 Sistemas de Gestão da Qualidade – Fundamentos e Vocabulário*, Instituto Português da Qualidade, Lisboa.

Juran, J.M. and Godfrey, A.B. (1998), *Juran's Quality Handbook*, McGraw-Hill, New York.

López-Lemus, J.A., Moreno-Enguix, M.R. and Valls Martínez, M.C. (2023), "ISO 9001 and the public service: An investigation of the effect of quality management systems on public service performance", International Journal of Organizational Analysis, Vol. 31 No. 4, pp. 1143-1161.

Manders, B., de Vries, H.J. and Blind, K. (2016), "ISO 9001 and product innovation: A literature review and research framework", *Technovation*, pp. 41–55.

Marcondes, R.B. (2018), *Fundamentos do Sistema de Gestão da Qualidade Aplicados aos Processos Produtivos de uma Empresa de Projetos de Engenharia e Arquitetura*, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Marques, M. (1997), *Auditoria e Gestão*, Editorial Presença, Lisboa.

Moumene, A. and Benhrimida, M. (2025), "ISO 9001 as a tool for transforming public services", Journal of Economics, Finance and Management, Vol. 2 No. 1, pp. 201–220.

Nogueira, P. and Moutinho, F. (2001), "Primeiros passos da ISO 9001 na Administração Pública em Portugal", *Revista de Qualidade e Gestão*, pp. 33–49.

Oliveira, A. and S.M. (2018), *ISO 9001 e a Qualidade na AP: Conceitos e Aplicações*, Editora Universitária, Lisboa.

Pires, A.R. (2007), *Sistemas de Gestão da Qualidade*, 3rd ed., Edições Sílabo, Lisboa.

Poksinska, B. (2010), "Challenges and misconceptions in the implementation of ISO 9001 in organizations", *Total Quality Management & Business Excellence*.

Rivera-Fernández, G. and Fragoso-Martínez, F.J. (2025), "Quality Management in Public Services: Lessons for Municipal Governance and Citizen-Centric Administration", Lex Localis – Journal of Local Self-Government.

Rodríguez-Antón, J.M., Ruiz-Aguirre, A.L. and Cegarra-Navarro, M.S. (2020), "Innovation and quality management: ISO 9001:2015 standard".

Vieira, S., Costa, E.S.P., Santos, C.A. dos and Berbara, R.L.L. (2022), "Certificação ISO 9001 em uma instituição pública de ensino e pesquisa agrícola: um estudo de caso", *Research, Society and Development*, Vol. 11 No. 3, e0911325973, available at: https://doi.org/10.33448/rsd-v11i3.25973

Zimon, D., Tyan, J. and Sroufe, R. (2020), "Implementing sustainable supply chain management: Reactive, cooperative, and dynamic models", *Sustainability*.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Measurement System Analysis and Robustness Assessment of an AI-Based Vision Inspection System: Evidence from Automotive Production

Frederico Santos [1], Eduardo e Oliveira [2], Hermilio Vilarinho [2] and Henriqueta Nóvoa [2]

[1] Faculdade de Engenharia, Universidade do Porto; up202006435@edu.fe.up.pt
[2] INEGI, Faculdade de Engenharia, Universidade do Porto; eoliveira@inegi.up.pt
[2] INEGI, Faculdade de Engenharia, Universidade do Porto; hvilarinho@inegi.up.pt; hcfernandes@fe.up.pt
[2] INEGI, Faculdade de Engenharia, Universidade do Porto; hnovoa@fe.up.pt

***Abstract*—The industrial adoption of AI-based vision inspection systems has increased significantly in recent years, particularly in quality-critical manufacturing environments. Despite their promising performance, systems that demonstrate high predictive accuracy during offline development may still exhibit unstable behaviour when exposed to real production variability. Ensuring their reliability therefore requires validation approaches aligned with quality engineering principles.**

This study presents an industrial case study focused on the structured validation of an AI-based Automated Visual Inspection (AVI) system deployed on an automotive assembly line. The proposed validation strategy integrates established quality engineering tools with statistical modelling and operational monitoring in order to evaluate inspection repeatability, robustness to process variability, and stability during real production operation. In particular, the approach combines Attribute Agreement Analysis (AAA) to assess inspection consistency, General Linear Modeling (GLM) to analyse sensitivity to operational factors, and confusion-matrix-based performance monitoring during system deployment.

The results illustrate how AI-based inspection solutions can be assessed as measurement systems operating within dynamic production environments, rather than solely as predictive models evaluated through offline metrics. The study provides empirical evidence from an industrial application demonstrating how structured quantitative validation methods can support the reliable deployment of AVI systems in high-volume manufacturing. The proposed approach contributes to bridging the gap between advances in AI-based inspection technologies and established practices in quality engineering and measurement system validation.

Keywords – AI-based vision systems; Quality engineering; Measurement system analysis; Industrial inspection; attribute agreement analysis

I. Introduction

Artificial Intelligence (AI)-based vision inspection systems are increasingly deployed as in-line quality control solutions in high-volume manufacturing environments. In sectors such as automotive production, where inspection decisions directly affect customer safety, production flow, and cost performance, these systems are often positioned as critical quality gates. Their ability to process large image datasets in real time and automate defect detection has made them attractive alternatives to manual inspection.

Despite their growing adoption, a persistent challenge remains: strong predictive performance under offline testing conditions does not necessarily guarantee reliable operation in real production environments. While AI models are typically evaluated using metrics such as accuracy, precision, recall, or F1-score on static validation datasets, these metrics do not capture important industrial requirements such as repeatability, robustness to operational variability, and performance stability over time. This limitation is particularly relevant in industrial applications where AI-based inspection systems are deployed as proprietary or vendor-supplied solutions, and the underlying algorithms operate as "black-box" models without direct access to their internal logic. In such contexts, validation must focus on the observable behaviour of the inspection system as a measurement process rather than on the internal properties of the algorithm itself.

In industrial practice, inspection systems must be evaluated not only as predictive algorithms but as measurement systems operating within dynamic production processes. Variations in environmental conditions, part positioning, process states, and operational disturbances can influence inspection outputs in ways that are not observable during offline controlled model development. When unexpected inspection errors occur, it becomes difficult to determine whether they originate from algorithmic limitations or from external process variability. This ambiguity reduces trust in AI-based inspection systems and complicates root cause analysis and quality assurance activities. Systems that perform well during offline validation may still generate false positives or false negatives once integrated into the production line.

This study addresses this challenge through a quantitative industrial case study conducted in an automotive assembly environment. The objective is to evaluate the repeatability, robustness, and deployment stability of an AI-based vision inspection system using structured statistical methods.

The contribution of this paper is threefold. First, it provides quantitative evidence of inspection repeatability under identical operating conditions. Second, it evaluates the sensitivity of the system to controlled variations in environmental and operational factors using General Linear Modelling (GLM). Third, it assesses performance stability during a real three-week production deployment, using confusion-matrix-based key performance indicators.

By presenting detailed statistical results from a real automotive production context, this study contributes empirical evidence on how AI-based vision inspection systems can be validated beyond conventional evaluation metrics typically used during model development, such as accuracy, precision, recall, and F1-score derived from confusion matrices. While these metrics are useful to assess predictive performance on static datasets, they do not capture important industrial requirements such as inspection repeatability, robustness to production variability, and stability during continuous operation. Although the findings are derived from a specific industrial application, the validation logic demonstrated in this case may inform similar deployments in other manufacturing environments.

The remainder of this paper is organized as follows. Section II reviews the relevant literature on AI-based visual inspection and measurement system validation in manufacturing contexts. Section III presents the industrial context and describes the AI-based vision inspection system analysed in this study. Section IV details the methodological approach adopted for the validation, including the repeatability assessment, robustness experiments, and deployment monitoring strategy. Section V presents the results obtained from the industrial case study and discusses the implications of these results for the validation of AI-based inspection systems in production environments. Finally, Section VI concludes the paper and outlines directions for future research.

II. Literature Review And Research Gap

The application of Artificial Intelligence (AI) in industrial visual inspection has expanded substantially over the past decade, largely driven by advances in deep learning. Deep learning refers to a class of machine learning methods based on multi-layer artificial neural networks that automatically learn hierarchical feature representations directly from raw data, eliminating the need for manually engineered features (LeCun, Bengio and Hinton, 2015). Within this paradigm, Convolutional Neural Networks (CNNs) have become the dominant architecture for image-based tasks. CNNs are specialized neural network models designed to process grid-structured data such as images, using convolutional layers to automatically extract spatial patterns and visual features through learnable filters (Rawat and Wang, 2017). In manufacturing environments, deep learning-based inspection systems have demonstrated superior performance compared to traditional rule-based or handcrafted feature approaches, particularly in complex surface defect detection applications where defect morphology and texture variations are difficult to model explicitly (Huang, Yang, Zheng and Wang, 2018).

Recent reviews confirm the widespread adoption of deep learning in Automated Visual Inspection (AVI) across automotive, electronics, steel and semiconductor industries (Tao, Zhang, Liu and Nee, 2018; Hütten et al., 2024). These studies highlight that performance improvements are frequently achieved through transfer learning strategies, improved data curation and augmentation techniques rather than through fundamental modifications of inspection logic (Weiher et al., 2023; Zhao, Zheng, Xu and Wu, 2019). However, most published results rely on static validation datasets evaluated under controlled laboratory conditions.

Recent comprehensive surveys have examined industrial surface defect detection methods powered by deep learning, highlighting both classification and localization approaches as well as challenges related to dataset scarcity and real production variability (Ameri, 2024; Ma, 2024). These surveys emphasize the importance of robust model design, evaluation metrics beyond offline accuracy, and the practical role of visual object detection techniques in industrial quality settings (Czimmermann, 2020). Furthermore, systematic reviews of surface defect detection methods reveal that deep learning approaches must address real-world concerns such as imbalance, small sample issues, and operational deployment constraints in manufacturing contexts (Jin and Chen, 2022; Qi, 2020).

The discrepancy between benchmark accuracy and operational robustness has been documented across machine learning domains. Hendrycks and Dietterich (2019) demonstrated that neural networks with high benchmark performance can experience significant degradation under common corruptions and perturbations. Similarly, Bergmann, Fauser, Sattlegger and Steger (2019) highlighted the challenges associated with realistic industrial anomaly detection datasets, where natural variability and defect rarity complicate reliable evaluation. These findings suggest that conventional accuracy-based metrics may overestimate real-world inspection reliability.

In dynamic manufacturing systems, inspection models are exposed to distribution shift and concept drift. Gama, Žliobaitė, Bifet, Pechenizkiy and Bouchachia (2014) and Lu et al. (2019) provide comprehensive reviews of learning under concept drift, emphasizing that evolving data distributions can significantly affect model performance over time. Koh et al. (2021) further demonstrated that models trained using empirical risk minimization may not generalize reliably under real-world distribution shifts. Although these studies advance understanding of algorithmic robustness, they rarely frame inspection validation within established quality engineering methodologies.

Parallel to robustness research, the discipline of Machine Learning Operations (MLOps) has emerged to address deployment monitoring and lifecycle management of AI systems. Sculley et al. (2015) emphasized the hidden system complexity in machine learning systems deployed in production. Breck, Cai, Nielsen, Salib, and Sculley (2017) proposed structured evaluation criteria for assessing production readiness. While these contributions improve software reliability practices, they do not explicitly incorporate industrial measurement system capability or inspection risk evaluation.

From a quality engineering perspective, inspection systems are treated as measurement systems that must demonstrate repeatability, reproducibility, and capability. Measurement System Analysis (MSA) is widely established in

manufacturing industries, particularly in automotive contexts aligned with IATF standards (AIAG, 2010). Within MSA, Attribute Agreement Analysis (AAA) is used for categorical or binary inspection systems (e.g., OK/NOK). AAA evaluates the consistency of classification decisions by measuring agreement across repeated inspections, between different appraisers, and against a known reference standard. It quantifies repeatability (within-system consistency), reproducibility (between-appraiser consistency), and misclassification rates such as false positives and false negatives, which are directly linked to producer and customer risk. Despite this mature body of knowledge, systematic application of MSA principles, and particularly AAA, to AI-based vision inspection systems remains limited in the literature.

Some empirical studies have begun bridging this gap. Guha, Moore and Huyghe (2023) applied structured statistical validation methods to a machine vision system in a regulated biomedical manufacturing context, demonstrating the importance of reliability assessment beyond predictive accuracy.

Despite the growing body of research on AI-based visual inspection and machine learning deployment practices, a clear methodological gap remains regarding the industrial validation of such systems once deployed on production lines. Existing studies largely emphasize model-level performance and algorithmic improvements, while the evaluation of AI-based inspection systems as measurement systems embedded in manufacturing processes remains limited. In quality-critical production environments, however, inspection reliability must be assessed in terms of repeatability, robustness to operational variability, and stability during continuous operation. Addressing this challenge requires validation approaches that integrate machine learning evaluation with established quality engineering principles.

The present study contributes to this need by presenting a quantitative industrial case study that applies structured validation methods to assess the performance of an AI-based vision inspection system deployed in an automotive production environment.

III. Industrial Context and Inspection System

The case study was conducted in an automotive manufacturing plant producing structural components for seat frames (Fig.1). These components are assembled through multiple fastening and positioning operations and are subject to strict quality requirements due to their functional and safety relevance. Non-conforming parts may compromise structural integrity and generate significant downstream costs associated with rework, scrap, or potential customer risk.

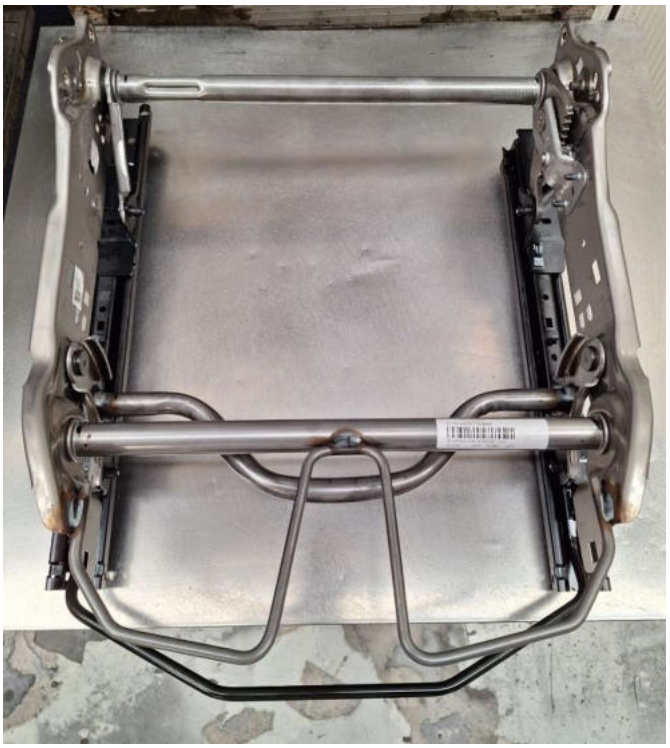


Fig. 1. Seat Frame produced in the manufacturing plant

The production line operates under high-volume conditions, with continuous flow and limited buffering between stations. As a result, inspection decisions must be both rapid and reliable, as they directly influence production continuity. The AI-based vision inspection system analysed in this study was deployed as an in-line quality gate responsible for detecting assembly-related defects associated with fastening elements (e.g., lockbolt presence and positioning). An example of a part defect is shown in Fig. 2.

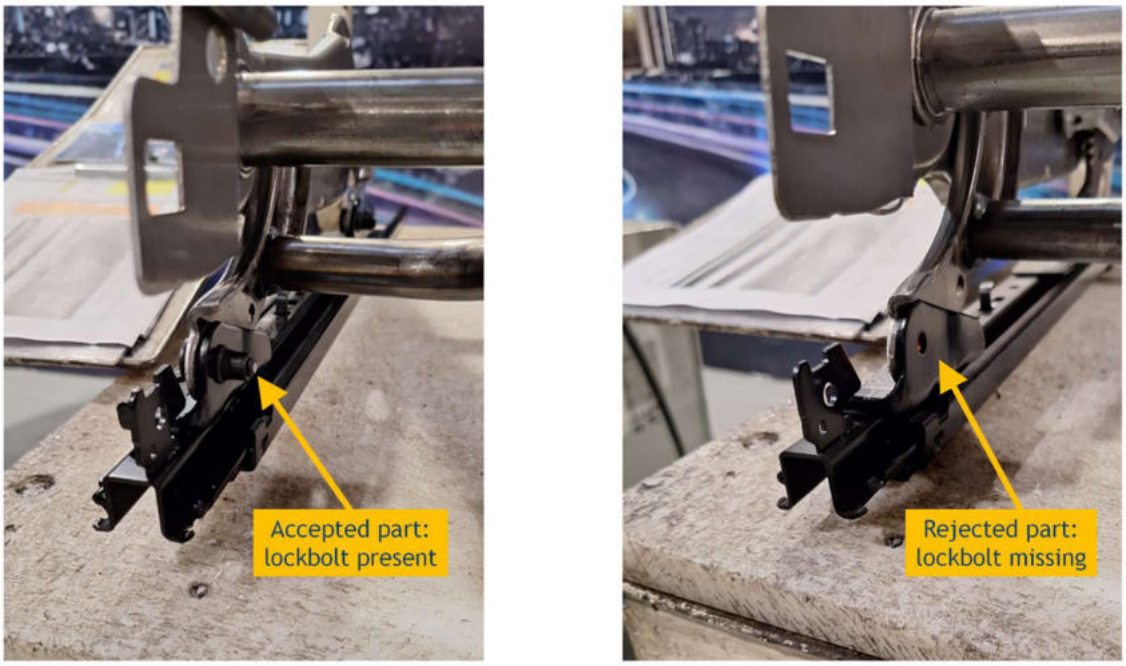


Fig. 2. Example of a defect to be detected by the Automated Visual Inspection (Left: accepted part | Right: rejected part due to a missing lockbolt)

The inspection system consists of an industrial camera, controlled internal illumination, mechanical part positioning via carriers, and a deep-learning-based classification model that outputs both a binary decision (OK/NOK) and a continuous confidence score ranging from 0 to 100, indicating the model assurance regarding the predicted class (OK / NOK). A predefined operational threshold (Score $\geqslant$ 90) determines the acceptance of a part. If the confidence score falls below this threshold, the part is classified as non-conforming.

During the initial deployment phase, the inspection system exhibited several false positives (FP) and false negatives (FN), indicating that the system was not yet properly integrated into the production environment. At that stage, the system was effectively operating as a standalone hardware solution rather than as a fully calibrated inspection system embedded within the production process.

Although offline validation during model development showed high predictive performance, plant stakeholders expressed uncertainty regarding system stability under real production variability. Specifically, questions emerged regarding:

- Whether the system produced consistent decisions under identical conditions;
- Whether environmental or operational variations influenced the algorithm's confidence output;
- Whether inspection performance remained stable over time during continuous operation.

A structured analysis of potential failure causes was therefore conducted, focusing not only on algorithmic behaviour but also on external factors that could influence inspection performance. Since the internal algorithmic logic was not accessible (black-box model), the analysis concentrated on operational and environmental conditions affecting the inspection process.

Based on this analysis, several improvement actions were implemented, including replacement of the part carriers, adjustment of camera exposure and lens aperture, and additional training of the AI model using representative borderline cases. These actions significantly improved system stability and enabled the implementation of the structured validation procedure presented in this study.

This improvement phase highlights the importance of considering inspection systems as socio-technical measurement systems embedded within production environments rather than as isolated predictive models.

IV. Methodological Approach

The validation strategy adopted in this study was structured to quantitatively assess three complementary dimensions of inspection performance: repeatability under identical conditions: (i) repeatability assessment under controlled conditions, (ii) robustness and sensitivity analysis under controlled operational variability, and (iii) deployment monitoring during real production operation. The methodological design was defined in collaboration with plant quality engineers to ensure industrial relevance and operational feasibility.

A. Repeatability Assessment

To evaluate intrinsic system consistency, an Attribute Agreement Analysis (AAA) was conducted. A sample of 20 representative parts (including both conforming and non-conforming cases) was inspected twice under identical and controlled conditions, resulting in 40 inspection observations (20 parts × 2 repetitions). Environmental parameters, carrier positioning, lighting conditions, and system configuration were kept constant.

Inspection outcomes (OK/NOK) were analysed using attribute-based Measurement System Analysis principles to determine agreement across repetitions and alignment with the predefined reference classification.

B. Robustness and Sensitivity Analysis

To evaluate system sensitivity to realistic operational variability, a designed experiment was conducted considering three potentially influential factors identified through cause-and-effect analysis:

- Block (external environmental/temporal conditions);
- Carrier (mechanical positioning conditions);
- Light (internal illumination level).

The factors used in the experiment with their corresponding levels are presented in Table I.

TABLE I. FACTORS AND LEVELS IN EXPERIMENTS

Factor	*levels*
Block	1. Day 1 – Morning 2. Day 1 – Afternoon 3. Day 2 – Morning 4. Day 3 – Night 5. Day 4 - Night
Carrier	1. Normal 2. With Deviation
Light	1. Light Off 2. Light On

For each experimental run, both the binary classification result and the continuous Confidence Score (0–100 scale) were recorded. The analysis focused on the Confidence Score to capture subtle variations that may not immediately change the OK/NOK decision.

A General Linear Model (GLM) was applied to evaluate the influence of the experimental factors on the Confidence Score. Statistical significance of the main effects was assessed using Analysis of Variance (ANOVA), with p-values used to determine whether variations in factor levels produced statistically significant changes in the response variable. Estimated coefficients were analysed to interpret the direction and magnitude of the factor effects. Model adequacy was evaluated through residual diagnostics to verify the assumptions of normality, independence, and homoscedasticity.

A second experimental repetition was conducted to evaluate the reproducibility of initially observed effects. Additionally, an outlier analysis was performed to assess the influence of extreme observations on model behaviour.

C. Deployment and Monitoring

Following experimental validation, the inspection system was monitored during three consecutive production weeks. Real inspection data were collected under normal operating conditions.

Performance was evaluated using key performance indicators derived from the confusion matrix. These correspond to the offline evaluation metrics typically used during model development. In this context, True Positives (TP) represent defective parts correctly identified as non-conforming, True Negatives (TN) correspond to conforming parts correctly classified as acceptable, False Positives (FP) indicate conforming parts incorrectly classified as defective, and False Negatives (FN) correspond to defective parts incorrectly classified as acceptable.

Based on these quantities, the following performance metrics were computed:

- Accuracy = (TP + TN) / (TP + TN + FP + FN)
- Precision = TP / (TP + FP)
- Recall (Sensitivity) = TP / (TP + FN)
- Specificity = TN / (TN + FP)
- F1-score = 2 × (Precision × Recall) / (Precision + Recall)

Temporal stability of the inspection system was assessed by comparing these indicators across the monitoring period and by analysing the occurrence of false positives and false negatives.

This multi-stage quantitative approach enabled systematic evaluation of inspection reliability from controlled testing to real industrial deployment.

V. Results And Discussion

This section presents the results obtained from the repeatability assessment, robustness experiments, and deployment monitoring. The analysis distinguishes between statistical significance and operational relevance, which is critical in industrial inspection contexts.

A. Repeatability Results

The Attribute Agreement Analysis (AAA) was conducted using 20 representative parts that were previously inspected and classified by experienced human inspectors to establish the reference standard. Each part was then inspected twice by the AI-based vision system under identical conditions, resulting in 40 inspection observations. Across all repetitions, the system produced fully consistent classification results.

For every part, the OK/NOK decision remained identical between repetitions and was fully aligned with the predefined reference classification. No classification changes were observed, and no contradictory decisions occurred.

This corresponds to a 100% repeatability rate under controlled conditions.

Importantly, repeatability was confirmed for both conforming and non-conforming parts, including borderline cases. The absence of intra-system variability indicates that the inspection output is not affected by short-term instability or internal state fluctuations when operating under stable environmental and mechanical conditions.

These results establish a reliable performance baseline and justify progression to robustness testing under controlled variability.

B. Robustness and Sensitivity Analysis

1) First Experiment

The first designed experiment evaluated the influence of three factors—Block, Carrier, and Light—on the continuous Confidence Score output, assuming a 5% significance level.

The General Linear Model (GLM) analysis revealed:

- Block: $p = 0.003$ (statistically significant)
- Carrier: $p = 0.084$ (not significant)
- Light: $p = 0.766$ (not significant)

The fitted model explained a substantial portion of the variability in the Confidence Score ($R^2 = 71.79\%$). The most influential factor was Block 2, which exhibited a negative coefficient of approximately -1.026 relative to the overall mean.

The mean Confidence Score across all conditions was 99.154 (on a 0–100 scale). Considering the operational acceptance threshold of 90, this corresponds to a safety margin of approximately 9.15 score units.

Although the Block factor was statistically significant, the maximum observed deviation (-1.026) to the overall mean represents only about 11% of the available operational margin. Therefore, from a practical perspective, the effect does not approach the acceptance threshold and does not induce classification instability.

This distinction between statistical significance and operational relevance is critical: while measurable differences in score were detected, they do not translate into decision-level risk.

2) Second Experiment

To assess reproducibility, the experiment was repeated under comparable conditions.

In contrast to the first experiment, none of the tested factors were statistically significant:

- Block: $p = 0.618$
- Carrier: $p = 0.202$
- Light: $p = 0.363$

The model explained only a limited portion of the variability, suggesting that the previously observed Block effect was not consistently reproducible.

Further analysis revealed the presence of an extreme observation (Confidence Score ≈ 93), substantially below the overall mean (~99). This observation inflated residual variance and reduced model explanatory power.

3) Re-analysis Without Outlier

After excluding the extreme observation, the GLM was re-estimated. The revised analysis showed that none of the evaluated factors were statistically significant at the 95% confidence level. The obtained p-values were:

- Block: $p = 0.099$
- Carrier: $p = 0.349$

• Light: p = 0.946

Although the Block factor presented the lowest p-value, its effect remains non-significant and therefore does not indicate a consistent influence on the Confidence Score.

Overall, the robustness experiments indicate that the system exhibits limited sensitivity to the tested operational factors. The observed variability appears primarily attributable to intrinsic algorithmic fluctuation or minor geometric variation rather than systematic process effects.

C. Deployment and Monitoring Results

The system was monitored during three consecutive production weeks (weeks 49–51). Performance indicators derived from confusion matrices were computed weekly. Table II presents the performance indicators.

TABLE II. WEEKLY PERFORMANCE METRICS OF THE AVI SYSTEM DURING PRODUCTION MONITORING

Week	Accuracy	Precision	Recall	Specificity	F1-score
49	99.98%	0	N/A	99.98%	N/A
50	100%	100%	100%	100%	1
51	100%	N/A	N/A	100%	N/A

No false-negative cases were observed during the monitoring period, while only one false positive occurred, which was attributed to carrier misalignment.

As shown in Table II, Precision, Recall and F1-score exhibit variability due to the highly imbalanced nature of the inspection dataset, characterized by a very low incidence of defective parts. In weeks where no defective parts were observed, Recall and F1-score become undefined because no positive cases are available for evaluation. This behaviour reflects the class imbalance of the monitored production data rather than a degradation in system performance.

From an operational perspective, the most informative indicators are Accuracy and Specificity, which remained consistently high throughout the monitoring period. Both metrics reached 99.98% in Week 49 and 100% in Weeks 50 and 51, indicating reliable classification performance and stable identification of conforming parts.

Importantly, the absence of false negatives is particularly relevant from a quality engineering perspective, as such errors correspond to defective parts incorrectly accepted by the inspection system and therefore represent direct customer risk. These results confirm that the AVI system maintained stable and reliable defect detection capability during real production operation.

D. Integrated Interpretation

The combined results from controlled testing and deployment monitoring provide a comprehensive assessment of inspection reliability.

Repeatability testing confirmed intrinsic decision consistency. Robustness experiments showed that statistically detectable score variations do not translate into operational risk. Real-world monitoring demonstrated stable performance with near-perfect classification rates.

These findings illustrate that good offline model performance can be sustained in production when repeatability and robustness are quantitatively verified.

E. Discussion

The results of this study provide quantitative evidence regarding the validation of an AI-based vision inspection system under real production variability. Three main insights emerge from the analysis.

First, the repeatability assessment demonstrated that the inspection system exhibits full classification consistency under identical operating conditions. The 100% agreement observed across repeated inspections confirms that the system is not affected by short-term internal instability. From a quality engineering perspective, this establishes that the inspection output is deterministic under controlled conditions and that intrinsic algorithmic variability does not compromise decision consistency.

Second, the robustness experiments highlight the importance of distinguishing statistical significance from operational relevance. In the first experimental run, the Block factor showed a statistically significant influence on the Confidence Score ($p = 0.003$). However, the magnitude of the effect remained small relative to the operational threshold, and the effect was not reproducible in subsequent testing. This finding suggests that reliance solely on p-values may lead to overinterpretation of practically negligible effects.

In industrial inspection contexts, decision stability is more critical than minor variations in internal confidence scores. Although measurable score differences were detected, none approached the predefined acceptance threshold. The classification output remained unchanged across all tested factor levels. Therefore, the system demonstrates robustness at the decision level, which is the relevant performance dimension in production environments.

Third, deployment monitoring confirmed stable real-world operation. Over three production weeks, the system achieved accuracy levels above 99.98%, with no observed false negatives and only one false positive attributable to mechanical misalignment. The absence of false negatives is particularly relevant in safety-critical automotive manufacturing, as these represent customer-risk events.

An additional observation concerns the influence of data imbalance on performance metrics. Precision, recall, and F1-score exhibited variability due to the low frequency of defective parts during monitoring. This reinforces the importance of interpreting confusion-matrix-based indicators in the context of class distribution. In highly imbalanced industrial datasets, accuracy and specificity may provide more meaningful insights than aggregated metrics alone.

VI. Conclusions

This study presented a quantitative industrial case study evaluating the repeatability, robustness, and deployment stability of an AI-based vision inspection system in an automotive manufacturing environment. The validation approach combined Attribute Agreement Analysis, General Linear Modelling, and real production monitoring to assess inspection reliability beyond conventional offline predictive metrics.

The results demonstrated full repeatability under controlled conditions, confirming intrinsic classification stability. Robustness experiments revealed that although statistically significant variations in the Confidence Score were detected in one experimental run, these effects were not reproducible and remained practically negligible relative to the operational acceptance threshold. Deployment monitoring over three consecutive production weeks confirmed stable system performance, with accuracy above 99.98%, no observed false negatives, and only one false positive attributed to mechanical misalignment.

These findings highlight the importance of distinguishing between statistical significance and operational relevance in industrial AI applications. Minor variations in internal model confidence do not necessarily translate into decision-level risk. Therefore, validation of AI-based inspection systems should extend beyond offline algorithmic performance metrics and incorporate structured repeatability testing, controlled robustness experiments, and real-world monitoring.

Although this study focuses on a specific automotive production context, the quantitative validation logic demonstrated here may inform similar industrial implementations. Future work may extend this analysis to longer monitoring periods, additional production environments, and adaptive monitoring strategies to further strengthen lifecycle governance of AI-based inspection systems.

References

AIAG (2010), Measurement Systems Analysis (MSA): Reference Manual, 4th edn., Automotive Industry Action Group, Southfield, MI.

Ameri, R., Hsu, C.C. and Band, S.S. (2024), "A systematic review of deep learning approaches for surface defect detection in industrial applications", Engineering Applications of Artificial Intelligence, Vol. 130, pp. 107717.

Bergmann, P., Fauser, M., Sattlegger, D. and Steger, C. (2019), "MVTec AD – A comprehensive real-world dataset for unsupervised anomaly detection", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9592–9600.

Breck, E., Cai, S., Nielsen, E., Salib, M. and Sculley, D. (2017), "The ML test score: A rubric for ML production readiness and technical debt reduction", Proceedings of the IEEE International Conference on Big Data, pp. 1123–1132.

Czimmermann, T., Ciuti, G., Milazzo, M., Chiurazzi, M., Roccella, S., Oddo, C.M. and Dario, P. (2020), "Visual-based defect detection and classification approaches for industrial applications—a survey", Sensors, Vol. 20, No. 5, pp. 1459.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M. and Bouchachia, A. (2014), "A survey on concept drift adaptation", ACM Computing Surveys, Vol. 46, No. 4, pp. 1–37.

Guha, B., Moore, S. and Huyghe, J.-M. (2023), "Application and validation of machine vision inspection for efficient in-process monitoring of complex biomechanical device manufacturing", Journal of Engineering and Applied Science, Vol. 70, pp. 72.

Hendrycks, D. and Dietterich, T. (2019), "Benchmarking neural network robustness to common corruptions and perturbations", International Conference on Learning Representations.

Huang, C., Yang, C., Zheng, X. and Wang, X. (2018), "Surface defect saliency of magnetic tile", IEEE Transactions on Industrial Informatics, Vol. 14, No. 6, pp. 2509–2518.

Hütten, N., Alves Gomes, M., Hölken, F., Andricevic, K., Meyes, R. and Meisen, T. (2024), "Deep learning for automated visual inspection in manufacturing and maintenance: A survey of open-access papers", Applied System Innovation, Vol. 7, No. 1, pp. 11.

Jin, Q. and Chen, L. (2022), "A survey of surface defect detection of industrial products based on a small number of labeled data", arXiv preprint, pp. 2203.05733.

Koh, P.W., Sagawa, S., Marklund, H., Xie, S.M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R. and Liang, P. (2021), "WILDS: A benchmark of in-the-wild distribution shifts", Proceedings of the 38th International Conference on Machine Learning, pp. 5637–5664.

LeCun, Y., Bengio, Y. and Hinton, G. (2015), "Deep learning", Nature, Vol. 521, No. 7553, pp. 436–444.

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J. and Zhang, G. (2019), "Learning under concept drift: A review", IEEE Transactions on Knowledge and Data Engineering, Vol. 31, No. 12, pp. 2346–2363.

Ma, Y., Yin, J., Huang, F. and Li, Q. (2024), "Surface defect inspection of industrial products with object detection deep networks: A systematic review", Artificial Intelligence Review, Vol. 57, No. 12, pp. 333.

Qi, S., Yang, J. and Zhong, Z. (2020), "A review on industrial surface defect detection based on deep learning technology", Proceedings of the 2020 3rd International Conference on Machine Learning and Machine Intelligence, pp. 24–30.

Rawat, W. and Wang, Z. (2017), "Deep convolutional neural networks for image classification: A comprehensive review", Neural Computation, Vol. 29, No. 9, pp. 2352–2449.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M. and Crespo, J.-F. (2015), "Hidden technical debt in machine learning systems", Advances in Neural Information Processing Systems, Vol. 28.

Tao, F., Zhang, M., Liu, Y. and Nee, A.Y.C. (2018), "Digital twin driven smart manufacturing", Journal of Manufacturing Systems, Vol. 48, pp. 157–169.

Weiher, K., Rieck, S., Pankrath, H., Beuss, F., Geist, M., Sender, J. and Fluegge, W. (2023), "Automated visual inspection of manufactured parts using deep convolutional neural networks and transfer learning", Procedia CIRP, Vol. 120, pp. 858–863.

Zhao, Z.-Q., Zheng, P., Xu, S.-T. and Wu, X. (2019), "Object detection with deep learning: A review", IEEE Transactions on Neural Networks and Learning Systems, Vol. 30, No. 11, pp. 3212–3232.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Performance Measurement Systems in Continuous Improvement Programs: A Systematic Literature Review

Luiz Geraldo Goessler [1] and Roberto Antonio Martins [1]

[1] Department of Industrial Engineering, Federal University of Sao Carlos; luiz.goessler@estudante.ufscar.br

***Abstract*—Performance Measurement Systems (PMS) are a critical component of Continuous Improvement (CI) infrastructure, essential to align organizational priorities and foster innovative behaviours. However, while the interdependence of these two fields is acknowledged, current research remains fragmented across multiple research domains, veiling the underlying mechanisms through which PMS supports CI. To address this gap, we developed a Systematic Literature Review (SLR) to comprehensively map and synthesize literature on this research field. We identified nine mechanisms that describe specific PMS roles in CI programs and initiatives, moving beyond the use of KPIs to assess impacts on organizational performance. Additionally, we propose a research agenda to advance theory development on PMS uses for CI and to capture opportunities of emerging business trends such as Industry 4.0.**

Keywords – continuous improvement, performance measurement, systematic literature review

I. Introduction

Organisations around the world pursue CI implementation due to its lower financial investment in comparison to alternative methods to achieve superior quality and productivity results. However, CI still requires a robust infrastructure, commitments toward training, and sustained managerial focus to successfully fulfil stakeholders' expectations (Bessant and Francis, 1999). The PMS is a central component of the CI infrastructure. The implementation of a PMS aligned with the organisational competitive goals compel employees' attention on critical priorities and facilitates the synergy of team efforts throughout the organisation. Furthermore, the PMS influences individual behaviours towards the exploration of innovative ideas that transform daily operational practices to achieve operational excellence (Franco-Santos et al., 2012).

While the relationship between PMS and CI is widely acknowledged (Cunha et al., 2023), research is currently spread across varied industrial contexts and multiple research domains (Alhaddi, 2023). This lack of synthesis veils the underlying mechanisms of how PMS supports CI and refrains theory development on the field. Therefore, the following research question emerges: "What is the Performance Measurement System support to Continuous Improvement programs and initiatives?"

A SLR employs explicit and systematic methods to collate and synthesize findings from studies, addressing a clearly formulated question (Page et al., 2021). The SLR maps the literature comprehensively and synthesizes the main elements of a field, thus facilitating theory advance (Kraus et al., 2020). In the effort to organize such a vast corpus of knowledge as PMS and CI studies, we employed bibliometric methods, which offer the opportunity to analyse large research domains by examining influential ties in terms of key contributions, prolific authors, on-topic journals and chronological evolutions (Marzi et al., 2024)

This paper follows this organisation. Section 2 covers the main theoretical references for CI and PMS. Section 3 details the research method, Section 4 presents the main findings and Section 5 the discussion. Section 6 concludes with contributions, limitations, and recommendations for future research.

II. Theoretical References

A. Continuous Improvement

CI is a company-wide effort of high frequencies and short cycles of change which make a significant contribution to performance improvement when taken cumulatively over time (Bessant et al., 1994). To achieve long-term efficacy in CI, organisations must commit infrastructure and coordination to systematically sustain the enhancement efforts (Anand et al., 2009). Recent global empirical evidence identifies Lean, Six Sigma (SS) and Lean Six Sigma (LSS) as the primary reference frameworks for CI programs (Antony et al., 2024). Since 2010, organisations have increasingly named their CI programs as Operational Excellence (OPEX) (Found et al., 2018), and OPEX teams frequently replace traditional Lean and SS units (Sony, 2019).

Regardless of the framework, CI comprises teams committed to implementing targeted changes over a defined period to achieve desired improvements within a product, process, or service. This conceptualisation aligns closely with the formal definition of project (Arnaiz et al., 2022). Consequently, this research treats such collective, time-bound change endeavours as CI initiatives. Extensive research examines the critical factors (CF) necessary to implement, maintain, and advance the maturity of CI programs and initiatives (Sunder and Prashar, 2020; Sanchez-Ruiz et al., 2022; Arnaiz et al., 2022; Antony et al., 2024). Our analysis of these studies uncovers the following recurrent CFs:

- Commitment and Leadership: Top management activates CI through constant involvement and visible support. Strong leadership builds workforce confidence and develops the commitment to implement CI initiatives.

- Strategic Alignment: CI requires synchronisation with the organisation's overall strategic objectives. Divergence between CI programs and broader corporate aims causes the failure of specific CI initiatives.
- Organisational Context: The design and deployment of CI programs require careful consideration of national culture, the institutional environment, and the specific size and type of industry and market.
- Program Review: The program itself undertakes periodic updates driven by shifts in the internal or external environment, development of new CI frameworks and tools, or analysis of CI impact on organisational performance.
- Organisational Culture and Motivation: Management incites experimentation and learning from errors. Reward and recognition policies are connected to improvement achievements to encourage participation and mitigate resistance to change.
- Resource Allocation: Successful execution depends upon management providing the human, technical, financial, and infrastructural resources required to sustain CI initiatives across the enterprise.
- CI Infrastructure: Organisations define a standardised set of tools and techniques and develop a roadmap to embed them as daily practices. This infrastructure also includes formal measurement systems and supporting information systems to track performance gains and to manage CI initiatives data.
- Management Process: People specifically assigned to identify, select, implement, and review CI initiatives. CI initiatives focus on strategic goals and critical processes and hold clearly defined scopes and objectives.
- Training and Education: A systematic training program equips employees with fundamental CI frameworks, specialized tools, and problem-solving skills.
- Team Composition and Technical Expertise: Successful initiatives leverage interdepartmental teams affected by the improvement opportunity. Customers, suppliers, and other stakeholders are partners in developing solutions.
- Communication: Efficient, two-way interdepartmental communication facilitates the discussion of execution plans, the sharing of best practices, and cooperation across all organisational levels.

The CFs intuitively present an intrinsic link to organisation's performance measurement activities. To understand this connection, it is essential to refine the PMS construct to provide an appropriate theoretical foundation of the elements through which this association operates.

B. Performance Measurement Systems

Performance measurement (PM) is the process for quantifying the efficiency and effectiveness of actions (Neely et al., 2005). The field reached a critical turning point in the late 1990s, when rising criticism of financial short-termism catalysed the implementation of the Balanced Scorecard (BSC). This evolution sparked widespread interest from both academics and practitioners, leading to a significant expansion in the volume of scholarly research (Goshu and Kitaw, 2017). The more significant portion of academic activities in PM field has been consolidated into the management and business, operations management, and supply chain areas (Alhaddi, 2023).

Theoretical development in this field occurred within the domains of organisational and management control theories. According to this developmental path, the PMS constitutes a management control system comprising technical controls involving rational, planned, bureaucratic, and structural elements, and social controls concerning emergent, cultural, and behavioural aspects (Bititci et al., 2018). Fig. 1 shows the social and technical dimensions that form a PMS.

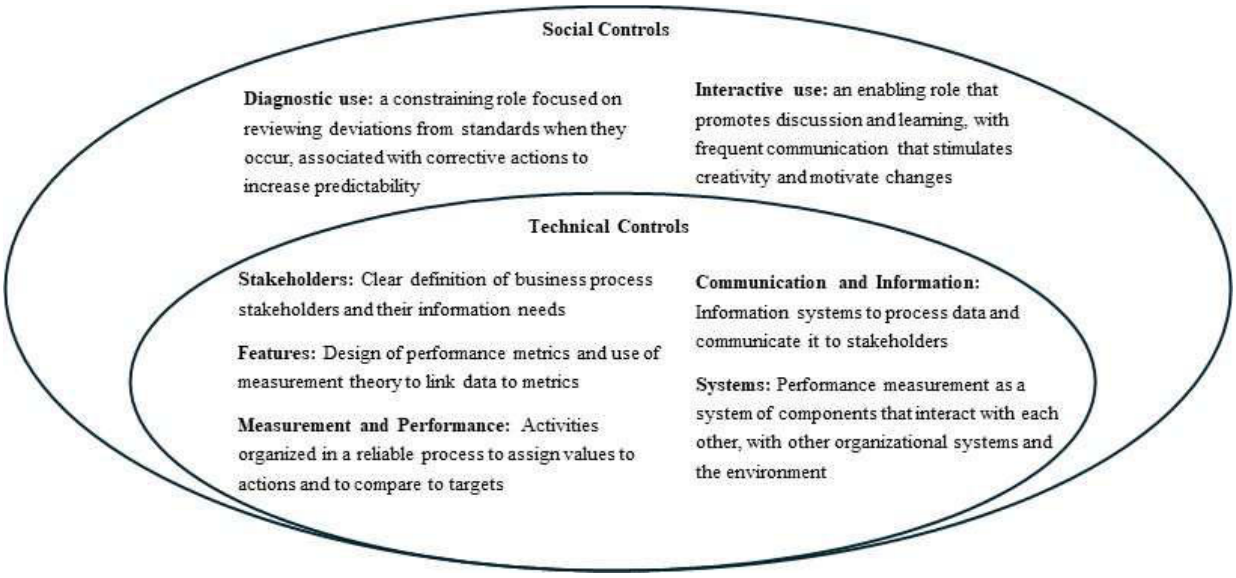


Fig. 1. PMS dimensions and elements. Adapted from Bititci et al. (2018), Choong (2014) and Tessier and Otley (2012)

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

It is worth noting that PMS dimensions are interconnected and mutually influential (Tessier and Otley, 2012). Therefore, a comprehensive understanding of the supportive roles of PMS to CI must consider a systemic view of both social and technical dimensions and elements of PMS as a management control system that facilitates the critical factors for CI.

III. Research Design

Over the past decade, the SLR has become a fundamental methodological instrument in management research. A SLR utilize rigorous rules and procedures to minimise reviewer bias and ensure the transparent report of research decisions. This practice enables other scholars to replicate or extend the work (Sauer and Seuring, 2023).

Fig. 2 delineates the systematic steps within this research. We developed the search string (Step 3) through a combination of keyword groups to represent synonyms for both core domains (CI and PMS). We also utilized a saturation process that concluded when the addition of further synonyms yielded no significant increase in document retrieval, especially regarding the research question (Step 1) and aimed characteristics of primary studies (Step2) (Sauer and Seuring, 2023).

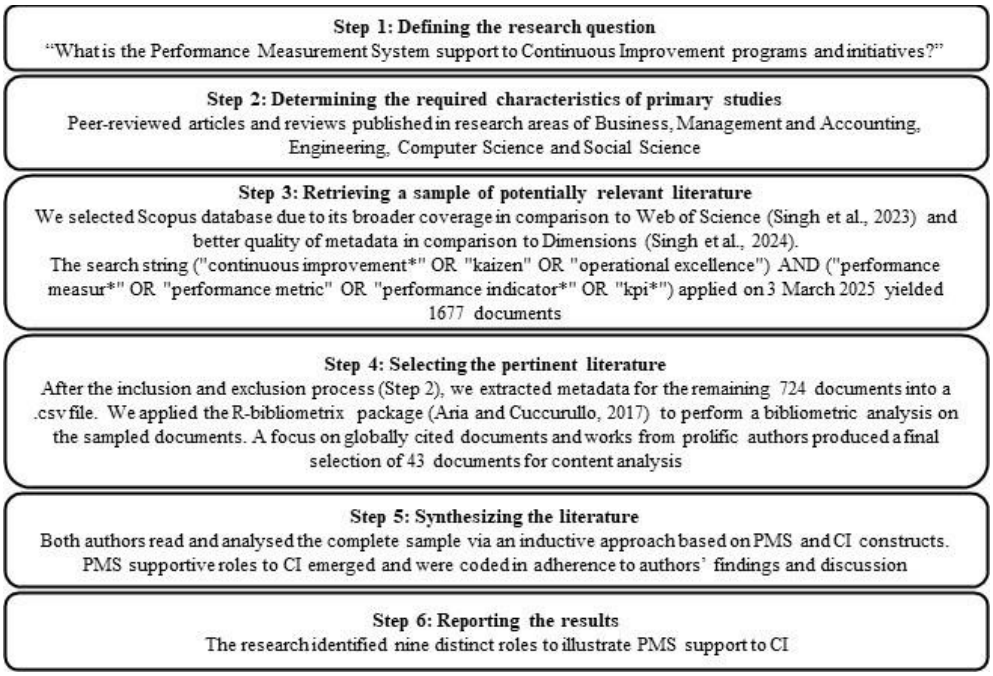


Fig. 2. SLR steps and procedures. Adapted from Sauer and Seuring (2023).

IV. Main Findings

The sample covers documents published from 1988 to 2026 across 414 sources, confirming the diversity of research areas on the field. The studies were produced by 2027 authors, with only 102 single-authored documents. Fig. 3 illustrates the scientific production and mean total citation per year (MeanTCperYear) until 2025. It uncovers an ascendant trend in published research, culminating in a significant surge in 2025. In this year, the volume of published documents nearly tripled compared to the average production of approximately 30 documents per year recorded between 2015 and 2024.

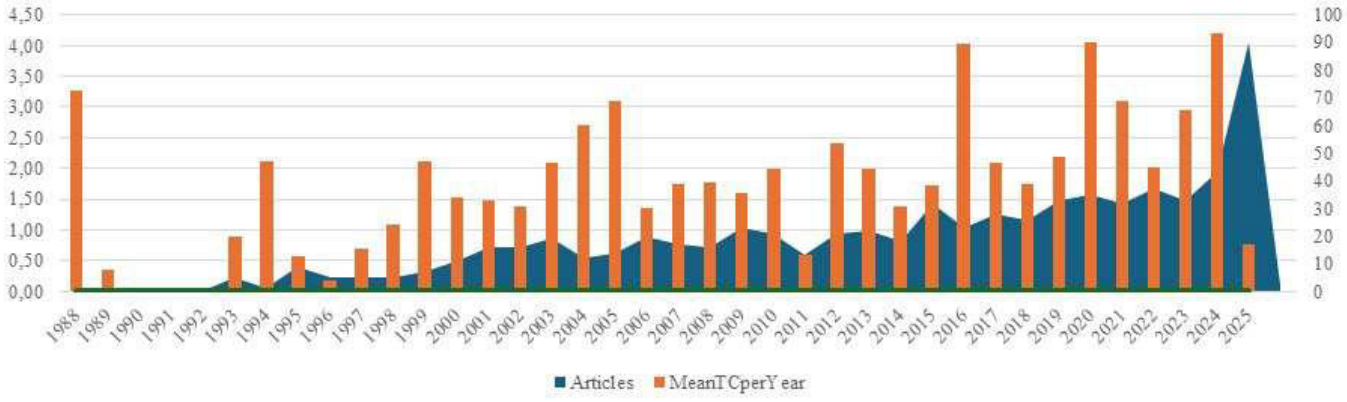


Fig. 3. Production and mean total citation per year

While 2016, 2020, and 2024 exhibit high MeanTCperYear, the overall high production across the timeline frequently yields elevated citation averages, obscuring the identification of specific impactful periods through Figure 2 alone. Consequently, to isolate the most influential contributions and strengthen the analysis, we prioritize the highest-cited global papers identified in the sample, as detailed in Table 1.

TABLE I. TOP TEN HIGH GLOBAL CITED DOCUMENTS IN THE SAMPLE

Rank	Title	Author	Year	Citations
1	The effect of relationship management on project performance in construction	Meng	2012	384
2	Designing a performance measurement system: A case study	Lohman et al.	2004	347
3	Implementation of total productive maintenance: A case study	Chan et al.	2005	301
4	Supply chain integration: A European survey	Bagchi et al.	2005	289
5	A methodology for effective implementation of lean strategies and its performance evaluation in manufacturing organisations	Karim and Arif-Uz-Zaman	2013	279
6	Overall equipment effectiveness as a measure of operational improvement - A practical analysis	Dal et al.	2000	270
7	Evaluation and improvement of manufacturing performance measurement systems - The role of OEE	Jonsson and Lesshammar	1999	228
8	A framework for building a smart port and smart port index	Molavi et al.	2020	219
9	Examination and comparison of the critical factors of total quality management (TQM) across countries	Sila and Ebrahimpour	2003	218
10	AI-Powered Innovation in Digital Transformation: Key Pillars and Industry Impact	Aldoseri et al.	2024	205

Three specific studies listed in Table 1 examine the application of KPIs to evaluate organisational themes without direct reference to PMS or CI theoretical frameworks. Bagchi (2005) analyses supply chain integration through a survey, emphasising the impact of information sharing and inter-organisational collaboration on firm performance. Molavi et al. (2020) proposes a Smart Port Index (SPI) to assess operational activity, environmental impact, energy consumption, and safety protocols, thereby identifying the strengths and weaknesses of port operations. Aldoseri et al. (2024) presents conceptual article advocating real-time process measurement monitoring for anticipation of challenges, identification of opportunities, and mitigation of risks to guide strategic actions for AI-driven digital transformations.

Two other documents examine PMS application in Supply Chain Management (SCM) without explicit alignment with CI programs or initiatives as defined in the theoretical references. Meng (2012) demonstrates, through a survey of the UK construction industry, that enhancing specific relationship dimensions, PMS included, can effectively reduce poor project performance. Similarly, Lohman et al. (2004) presents a case study of Nike's European operations, where the organisation developed a PMS a coordination mechanism based on established PM practices. This framework emphasises the creation of operational-level metrics to foster focus, trigger corrective actions, and facilitate the evaluation and refinement of strategic choices.

Shifting the focus to the specific contributions of PMS to CI, five studies worth particular attention in this research:

- Jonsson and Lesshammar (1999) characterize Overall Equipment Effectiveness (OEE) as a driver for shopfloor continuous improvement rather than a passive control tool. OEE uncovers "hidden" indirect costs and production losses and establishes an objective baseline for autonomous small groups to initiate improvement actions;
- Equally, the research by Dal et al. (2000) presents a case study wherein OEE serves not merely as an operational metric, but as a vital indicator to drive of process improvement activities;
- Sila and Ebrahimpour (2003) reviewed 76 survey studies to identify the critical factors of Total Quality Management (TQM) and categorized then under the 2001 Malcolm Baldrige criteria. PMS supports TQM especially regarding strategic goal setting, feedback to customers on quality improvements and to employees on individual and collective improved performance for recognition and reward purposes;
- Chan et al. (2005) present a case study illustrating the successful implementation of Total Productive Maintenance (TPM). In this context, the PMS facilitates the definition and communication of TPM objectives and targets. Furthermore, PMS enables the analysis, identification, and evaluation of improvement initiatives. It also supports the management of the TPM process through metrics such as the "number of improvements," "number of One-Point Lessons (OPLs)," and "number of suggestions.";
- Karim and Arif-Uz-Zaman (2013) introduced a Continuous Performance Measurement (CPM) metric designed to facilitate lean manufacturing implementation. A case study is developed reporting the CPM implementation at an Australian switchgear manufacturer. The CPM quantifies the process's current state to identify wastes and evaluates performance post CI initiatives implementation.

Fig. 4 illustrates the publication trajectory of the most prolific authors within the sample. The horizontal axis represents timeline by year, while the vertical axis ranks authors according to their total productivity, with the most prolific contributors at the top of the chart. Authors are identified by different colours. Each bubble indicates an author's scholarly output each year, the diameter of the bubble represents the number of published articles.

Antony, J. is the most productive contributor, publishing seven articles, followed by Carpinetti, L.C.R. and Pun, K. with six and five articles, respectively. Sousa, R.M. produced five articles, four of which involved co-authorship with Cunha, F. and

Dinis-Carvalho, J. Another significant collaborative output emerged from Gouvea da Costa, S.E. and Pinheiro de Lima, E., who co-authored four articles. Gupta, M. Parida, A., Katayama H. and Daniels, R.C. complete this scholarly landscape, each contributing four documents.

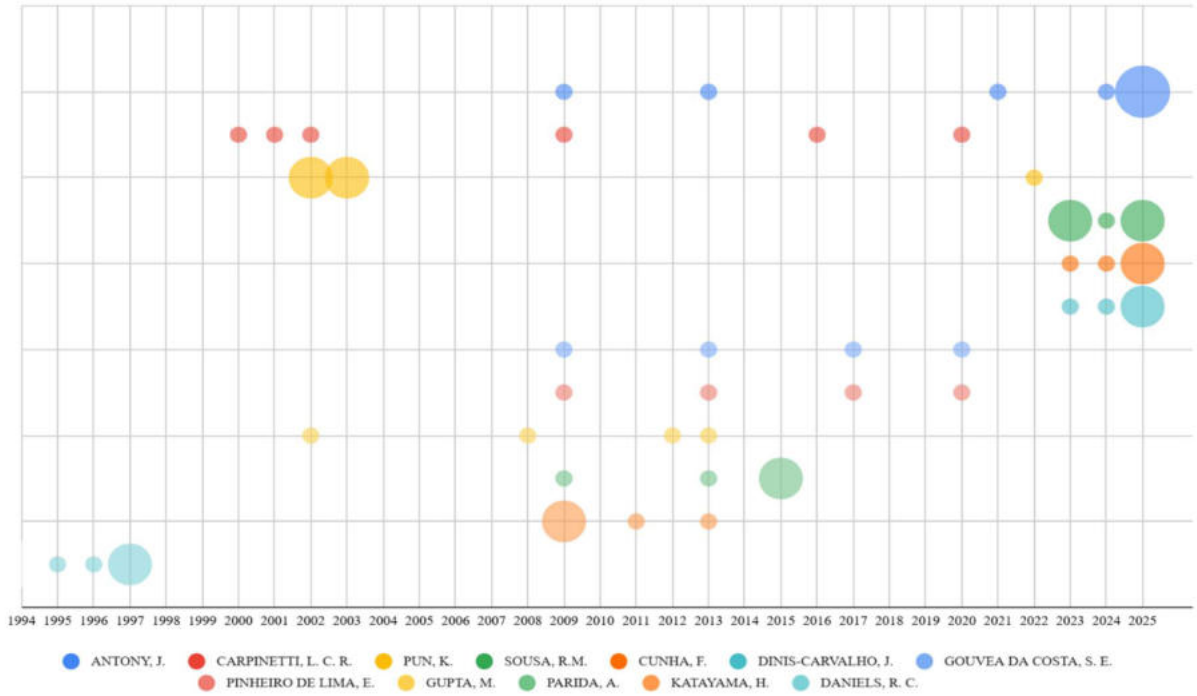


Fig. 4. Most relevant authors production over time

Antony's research identifies critical factors for CI initiatives, with a primary focus on LSS (Kumar et al., 2009, Gijo et al., 2021) and its recently integration into Industry 4.0 (Wankhede et al., 2025). His work emphasizes PMS role to evaluate the tangible impact of CI on organisational performance (Antony et al., 2024, Kumar et al., 2009). Recent studies investigate how LSS 4.0 serves as a strategic framework for enhancing energy efficiency and sustainability using KPIs as cycle time reduction, energy consumption per unit, and carbon footprint (Santos et al., 2025).

In revisiting Bessant´s maturity model for CI, Fryer, Ogden and Antony (2013) present PM as a specific indicator to measure CI organisational maturity. Its roles encompass the link between CI and strategic business goals, the assessment of CI initiative results and achievements, and the report to individuals and teams regarding their performance contribution to strategic goals through CI.

Carpinetti's recent research focuses on the application of fuzzy inference systems to predict how supply chain coordination impacts customer perceived value, specifically through SCOR® model indicators (e.g., agility, reliability, and responsiveness). Previous author´s contributions include the development of a frameworks that align CI efforts with manufacturing competitive criteria using PMS. This framework highlights the use of PMS to identify critical process for CI initiatives and to review progress of improvement actions (Carpinetti et al., 2000). Latter empirical research reveals no clear indication of alignment between competitive priorities and deployment of improvement actions in Brazilian companies' practices (Carpinetti and Martins, 2001).

Pun's earlier research focused on the conceptual development and empirical testing of integrated TQM and PM concepts, drawing primarily from national excellence award criteria. Within this proposed framework, the PMS elements received only superficial consideration, serving mainly to measure organisational results and process performance with no specific relation to CI programs and initiatives (Pun, 2002; Pun et al., 2002, Pun et al., 2003, Lalla et al., 2003). In recent research, Mahabir and Pun (2022) propose a KPI-based Performance Management System (KPMS) designed to revitalise project management office (PMO) operations within engineering-service contractors.

The collaborative work of Cunha, Dinis-Carvalho, and Sousa provides a comprehensive analysis of the barriers to effective PMS use (Cunha et al., 2024; Cunha et al., 2025a) and mechanisms to overcome then (Cunha et al., 2025b). Their research argues that PMS and CI are intrinsic and failing to use a PMS for improvement renders the system useless (Cunha et al., 2023). Gouvea da Costa and Pinheiro de Lima underline the PMS support in the development of organisational CI capability (Pinheiro de Lima et al., 2009, Pinheiro de Lima et al., 2013). Despite these authors' advancements, the extant literature fails to delineate the specific PMS mechanisms that facilitate the critical factor for CI.

Gupta's research on the Theory of Constraints (TOC) presents specialized metrics to drive CI such as throughput, inventory, and operating expenses. His work demonstrates how identifying system constraints and implementing Drum-Buffer-Rope scheduling enhance profitability. TOC metrics are used to measure specific processes and overall organisational performance (Gupta, et al., 2002, Gupta and Kline, 2008; Gupta and Andersen, 2012).

Katayama´s research focus on CI through JIT and presents a decision support systems that utilize Key Activity Indicators (KAI) to measure improvement processes (Hwang and Katayama, 2009) and quantitative models to evaluate how visual management impacts process performance (Murata and Katayama, 2013). Parida's research emphasizes that CI in railway infrastructure depends on Maintenance Performance Indicators (MPIs) for risk mapping and benchmarking to drive targeted technical improvements (Åhrén and Parida, 2009).

Daniels' studies advocates that fundamental shopfloor improvement requires allowing operators to establish their own metrics for CI initiatives aligned with business strategy (Daniels, 1995; Daniels and Burns, 1997). Additionally, the results of the project should be widely publicised to recognize employee's achievement and foster commitment.

Fig. 5 provides the authors' most frequent key words that confirm the representativeness of our sample to our research objectives. Over the 1959 authors' keywords, only 324 appears more than one time in the sample. It is worth noting that among the high cited key words, traditional CI initiatives such as "Lean" (along with "Lean Production" or "Lean Manufacturing"), "Lean Six Sigma", "Total Quality Management" appears together with wide used PMS framework "Balanced Scored Card".

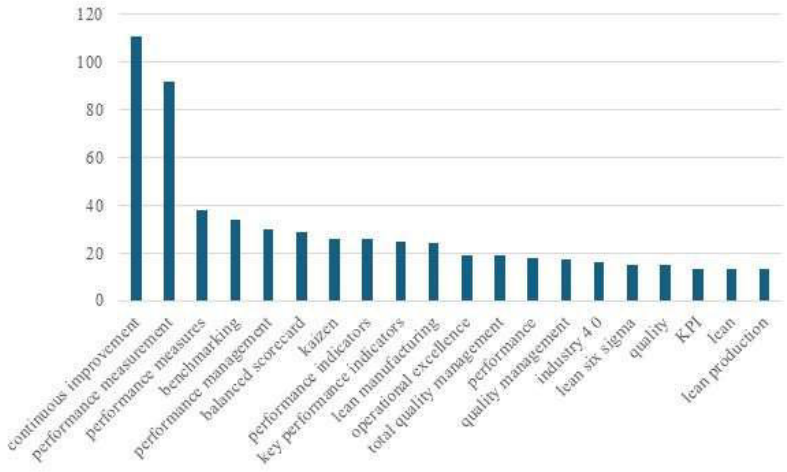


Fig. 5. High cited author´s key words

Additionally, "Industry 4.0" is a high cited key word, indicating a growing interested in research of PMS and CI in this field. Fig. 6 offers the trend topics analysis that corroborates with this growing interest. The analysis of trend topics also uncovers the arise of "Operational Excellence" as a recent key word, confirming that "Continuous Improvement" construct is migrating to Operational Excellence, as advocated by Found et al. (2018).

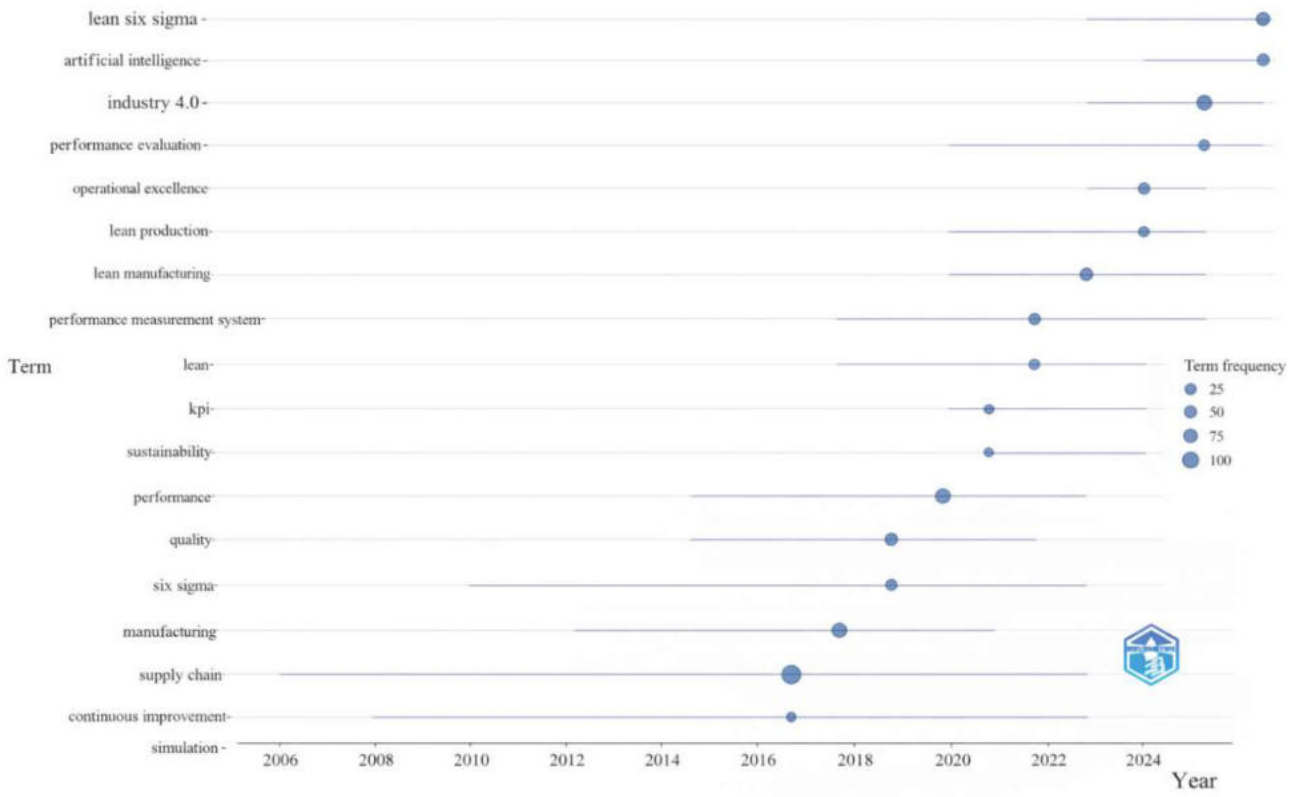


Fig. 6. Trend topics

V. Discussion

Management research generally acknowledges a fundamental interplay between PMS and CI. Nevertheless, our research identified that scientific output in these domains focuses primarily on the use of PMS to evaluate CI program success through KPIs. An exclusive reliance on PMS to measure CI past performance creates a notable concern as this approach relegates PMS solely to a diagnostic control role (Fig.1). Such a role inhibits creativity and leads to unintended consequences that limit innovation (Franco-Santos and Otley, 2018) and refrains other critical PMS contributions to CI. Despite this limitation, the SLR captured relevant mechanisms that depict specific PMS supportive roles in CI programs, as presented in Table 2.

TABLE II. PMS SUPPORTIVE ROLES TO CI

Role	PMS role description	Critical factor to CI	Key references
Strategic Top-Down Alignment	To connect CI programs to organisational competitive goals and deploy CI initiatives in accordance with business vision and strategic directions	Strategic alignment; management process	Carpinetti et al. (2000); Fryer, Ogden and Antony (2013)
Operational Bottom-Up Alignment	To identify opportunities for CI initiatives through process performance gaps assessment	Management process	Jonsson and Lesshammar (1999), Dal et al., (2000), Carpinetti et al. (2000), Karim and Arif-Uz-Zaman, (2013), Fryer, Ogden and Antony (2013)
Operational Engagement	To empower personnel at the shop-floor level allowing autonomous metrics and measurement processes definition for CI initiatives	Management process, organisational culture and motivation	Daniels (1995), Daniels and Burns (1997)
Organisational Communication	To disseminate CI priorities throughout communication of CI program goals in all levels of the organisation	Strategic alignment, communication	Chan et al. (2005), Sila and Ebrahimpour (2003)
Initiative Coordination	To review CI initiative against top-down or bottom-up objectives and measure CI initiatives contribution to systemic gains	Management process, program review	Carpinetti et al. (2000), Fryer, Ogden and Antony (2013), Karim and Arif-Uz-Zaman (2013)
Stakeholder Feedback	To provide performance data on CI achievements to stakeholders	Management process program review	Sila and Ebrahimpour (2003)
Employee Reinforcement	To deliver feedback information to employees and motivate sustained participation in CI initiatives	Organisational culture and motivation	Sila and Ebrahimpour (2003), Fryer, Ogden and Antony (2013)
Incentive Foundations	To establish an objective basis for recognition and reward policies derived from documented process improvement results	Organisational culture and motivation	Sila and Ebrahimpour (2003), Daniels and Burns (1997)
Programmatic Monitoring	To track CI program activities and identify opportunities for enhancing CI management processes, resource allocation, and organisational infrastructure	CI infrastructure, program review	Chan et al. (2005), Karim and Arif-Uz-Zaman (2013)

The nine supportive roles in Table 2 have no parallel in the reviewed literature. They establish an actual typology for PMS applications within CI programs and initiatives. The systematic literature review also reveals an emergent research trend toward the integration of Industry 4.0 technologies in PMS and CI contexts. Wankhede et al. (2025) demonstrate the key

characteristics of this technological convergence. While their study identifies tools with significant potential for I4.0 integration, it lacks precise details on the operationalisation of these connections.

Additionally, our findings suggest a gradual erosion of construct integrity in the field over time. Earlier studies define CI as an organisational collective effort for incremental innovation. In contrast, later research often defines CI as a perceived long-term increase in business or process performance and PMS is approached merely as a tool to assess these outcomes via KPIs.

VI. Conclusion and Contributions

We successfully addressed the primary research question and identified nine specific roles through which PMS contribute to CI. These findings advance beyond the traditional, narrow PMS use to assess CI program success solely through KPIs. Instead, this study elucidates how a PMS functions as an effective management control system for CI programs. Additionally, we propose of future research agenda in three main topics:

- While we uncovered an emerging trend regarding I4.0 technologies integration for PMS uses in CI, future empirical research would provide more guidance on academics and practitioners on the challenges, benefits, and impact of new information technologies adoption in this field.
- Moreover, this study calls for a rescue of theoretical foundations to counteract the observed erosion of construct integrity. Future research should rely on the theoretical underpinnings of CI (Bessant et al., 1994) or dynamic capabilities theory (Bessant and Francis, 1999; Anand et al., 2009). Such theorical foundation will contribute to a programmatic development of PMS roles for CI, especially in contemporary business trends.
- Limitations must be acknowledged, offering some new opportunities for future research. First, the selection of Scopus as the sole database, while justified, may have excluded relevant documents indexed exclusively in other specialized repositories. Second, the bibliometric approach must be further developed, including bibliometric coupling and co-occurrence analysis for selection of documents for content analysis (Marzi et al, 2024). This enhancement will provide to a more comprehensive and deeper understanding of the elements and characteristics of PMS that enable CF for CI.

In summary, this study addressed the need for a systematic synthesis of PMS uses as an intrinsic management control system for CI. We delineated the specific mechanisms through which PMS contributes to CI programs and initiatives. Furthermore, this research proposes new developmental paths for theory development in the field, especially regarding new business trends such as industry 4.0.

Acknowledgment

The authors acknowledge the reviewers for contributions and suggestions that enhanced this research.

References

Åhrén, T. and Parida, A. (2009), "Maintenance performance indicators (MPIs) for benchmarking the railway infrastructure: a case study", Benchmarking: An International Journal, Vol. 16 No. 2, pp. 247-258.

Aldoseri, A., Al-Khalifa, K. N. and Hamouda, A. M. (2024), "AI-powered innovation in digital transformation: key pillars and industry impact", Sustainability, Vol. 16 No. 5, pp. 1790.

Alhaddi, H. (2023), "Evolution of performance measurement research: an update on research development from 2005 to 2020 and future outlook for the field", Operations and Supply Chain Management: An International Journal, pp. 399-412.

Anand, G., Ward, P. T., Tatikonda, M. V. and Schilling, D. A. (2009), "Dynamic capabilities through continuous improvement infrastructure", Journal of Operations Management, Vol. 27 No. 6, pp. 444-461.

Antony, J. et al. (2024), "Global perspectives on operational excellence: unveiling critical failure factors and sustainable pathways", International Journal of Quality & Reliability Management, Vol. 41 No. 10, pp. 2604-2626.

Aria, M. and Cuccurullo, C. (2017), "bibliometrix: An R-tool for comprehensive science mapping analysis", Journal of Informetrics, Vol. 11 No. 4, pp. 959-975.

Arnaiz, F. D., Alvarez, V., Montequin, V. R. and Cousillas, S. M. (2022), "Identifying critical success factors in continuous improvement projects in a steel company", Procedia Computer Science, Vol. 196, pp. 832-839.

Bagchi, P. K., Chun Ha, B., Skjoett-Larsen, T. and Boege Soerensen, L. (2005), "Supply chain integration: a European survey", The International Journal of Logistics Management, Vol. 16 No. 2, pp. 275-294.

Bessant, J., Caffyn, S., Gilbert, J., Harding, R. and Webb, S. (1994), "Rediscovering continuous improvement", Technovation, Vol. 14 No. 1, pp. 17-29.

Bessant, J. and Francis, D. (1999), "Developing strategic continuous improvement capability", International Journal of Operations & Production Management, Vol. 19 No. 11, pp. 1106-1119.

Bititci, U. S., Bourne, M., Cross, J. A., Nudurupati, S. S. and Sang, K. (2018), "Editorial: towards a theoretical foundation for performance measurement and management", International Journal of Management Reviews, Vol. 20 No. 3, pp. 653-660.

Carpinetti, L. C. R., Gerólamo, M. and Dorta, M. (2000), "A conceptual framework for deployment of strategy-related continuous improvements", The TQM Magazine, Vol. 12 No. 5, pp. 340-349.

Carpinetti, L. C. R. and Martins, R. A. (2001), "Continuous improvement strategies and production competitive criteria: some findings in Brazilian industries", Total Quality Management, Vol. 12 No. 3, pp. 281-291.

Chan, F. T. S., Lau, H. C. W., Ip, R. W. L., Chan, H. K. and Kong, S. (2005), "Implementation of total productive maintenance: a case study", International Journal of Production Economics, Vol. 95 No. 1, pp. 71-94.

Choong, K. K. (2014), "The fundamentals of performance measurement systems: A systematic approach to theory and a research agenda", International Journal of Productivity and Performance Management, Vol. 63 No. 7, pp. 879-922.

Cunha, F., Dinis-Carvalho, J. and Sousa, R. M. (2023), "Performance measurement systems in continuous improvement environments: obstacles to their effectiveness", Sustainability, Vol. 15 No. 1, pp. 867.

Cunha, F., Dinis-Carvalho, J. and Sousa, R. M. (2024), "Assessment of performance measurement systems' ability to mitigate or eliminate typical barriers compromising organisational sustainability", Sustainability, Vol. 16 No. 5, pp. 2173.

Cunha, F., Dinis-Carvalho, J. and Sousa, R. M. (2025a), "Analysis of barriers for performance measurement system effectiveness in a company: perceptions across hierarchical levels", International Journal of Lean Six Sigma, Vol. 16 No. 2, pp. 542-559.

Cunha, F., Dinis-Carvalho, J. and Sousa, R. M. (2025b), "Proposal of a performance measurement system model: Overcoming effectiveness barriers", Production Engineering Archives, Vol. 31 No. 2, pp. 247-265.

Dal, B., Tugwell, P. and Greatbanks, R. (2000), "Overall equipment effectiveness as a measure of operational improvement – a practical analysis", International Journal of Operations & Production Management, Vol. 20 No. 12, pp. 1488-1502.

Daniels, R. C. (1995), "Performance measurement at the sharp end driving continuous improvement on the shop floor", Engineering Management Journal, Vol. 5 No. 5, pp. 211-214.

Daniels, R. C. and Burns, N. D. (1997), "A framework for proactive performance measurement system introduction", International Journal of Operations & Production Management, Vol. 17 No. 1, pp. 100-116.

Found, P., Lahy, A., Williams, S., Hu, Q. and Mason, R. (2018), "Towards a theory of operational excellence", Total Quality Management & Business Excellence, Vol. 29 No. 9-10, pp. 1012-1024.

Franco-Santos, M., Lucianetti, L. and Bourne, M. (2012), "Contemporary performance measurement systems: a review of their consequences and a framework for research", Management Accounting Research, Vol. 23, pp. 79-119.

Franco-Santos, M. and Otley, D. (2018), "Reviewing and theorizing the unintended consequences of performance management systems", International Journal of Management Reviews, Vol. 20 No. 3, pp. 696-730.

Fryer, K., Ogden, S. and Anthony, J. (2013), "Bessant's continuous improvement model: revisiting and revising", International Journal of Public Sector Management, Vol. 26 No. 6, pp. 481-494.

Gijo, E. V., Bhat, S., Antony, J. and Park, S. H. (2021), "Ten commandments for successful implementation of design for six sigma", The TQM Journal, Vol. 33 No. 8, pp. 1666-1682.

Goshu, Y. Y. and Kitaw, D. (2017), "Performance measurement and its recent challenge: a literature review", International Journal of Business Performance Management, Vol. 18 No. 4, pp. 381-402.

Gupta, M., Hyun-Jeung, H. and Hokey, M. (2002), "TOC-based performance measures and five focusing steps in job-shop manufacturing environment", International Journal of Production Research, Vol. 40 No. 4, pp. 907-930.

Gupta, M. and Kline, J. (2008), "Managing a community mental health agency: a theory of constraints based framework", Total Quality Management & Business Excellence, Vol. 19 No. 3, pp. 281-294.

Gupta, M. and Andersen, S. (2012), "Revisiting local TOC measures in an internal supply chain: a note", International Journal of Production Research, Vol. 50 No. 19, pp. 5363-5371.

Hwang, R. and Katayama, H. (2009), "A multi-decision genetic approach for workload balancing of mixed-model u-shaped assembly line systems", International Journal of Production Research, Vol. 47 No. 14, pp. 3797-3822.

Jonsson, P. and Lesshammar, M. (1999), "Evaluation and improvement of manufacturing performance measurement systems – the role of OEE", International Journal of Operations & Production Management, Vol. 19 No. 1, pp. 55-78.

Karim, A. and Arif-Uz-Zaman, K. (2013), "A methodology for effective implementation of lean strategies and its performance evaluation in manufacturing organizations", Business Process Management Journal, Vol. 19 No. 1, pp. 169-196.

Kraus, S., Breier, M. and Dasí-Rodríguez, S. (2020), "The art of crafting a systematic literature review in entrepreneurship research", International Entrepreneurship Journal, Vol. 16 pp. 1023-1042.

Kumar, M., Antony, J. and Douglas, A. (2009), "Does size matter for six sigma implementation? Findings from the survey in UK SMEs", The TQM Journal, Vol. 21 No. 6, pp. 623-635.

Lalla, T. R. M., Lewis, W. G., Pun, K., Chin, K. S. and Lau, H. C. W. (2003), "Manufacturing strategy, total quality management and performance measurement: an integrated model", International Journal of Manufacturing Technology and Management, Vol. 5 No. 5/6, pp. 414.

Lohman, C., Fortuin, L. and Wouters, M. (2004), "Designing a performance measurement system: a case study", European Journal of Operational Research, Vol. 156 No. 2, pp. 267-286.

Mahabir, R. J. and Pun, K. (2022), "Revitalising project management office operations in an engineering-service contractor organisation: a key performance indicator-based performance management approach", Business Process Management Journal, Vol. 28 No. 4, pp. 936-959.

Marzi, G., Balzano, M., Caputo, A. and Pellegrini, M. M. (2024), "Guidelines for bibliometric – systematic literature reviews: 10 steps to combine analysis, synthesis and theory development", International Journal of Management Reviews, Vol. 27 No. 1, pp. 81-103.

Meng, X. (2012), "The effect of relationship management on project performance in construction", International Journal of Project Management, Vol. 30 No. 2, pp. 188-198.

Molavi, A., Lim, G. J. and Race, B. (2020), "A framework for building a smart port and smart port index", International Journal of Sustainable Transportation, Vol. 14 No. 9, pp. 686-700.

Murata, K. and Katayama, H. (2013), "A study of the performance evaluation of the visual management case-base: development of an integrated model by quantification theory category III and AHP", International Journal of Production Research, Vol. 51 No. 2, pp. 380-394.

Neely, M. and Platts, L. (2005), "Performance measurement system design: a literature review and research agenda", International Journal of Operations & Production Management, Vol. 25 No. 12, pp. 1228-1263.

Page, M. J. et al. (2021), "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews", Systematic Reviews, Vol. 10 No. 1, pp. 89.

Pinheiro De Lima, E., Gouvea Da Costa, S. E. and Angelis, J. J. (2009), "Strategic performance measurement systems: a discussion about their roles", Measuring Business Excellence, Vol. 13 No. 3, pp. 39-48.

Pinheiro De Lima, E., Gouvea Da Costa, S. E., Angelis, J. J. and Munik, J. (2013), "Performance measurement systems: a consensual analysis of their roles", International Journal of Production Economics, Vol. 146 No. 2, pp. 524-542.

Pun, K. (2002), "Development of an integrated total quality management and performance measurement system for self-assessment: a method", Total Quality Management, Vol. 13 No. 6, pp. 759-777.

Pun, K., Chin, K., Chow, M. and Lau, H. C. W. (2002), "An effectiveness centered approach to maintenance management: a case study", Journal of Quality in Maintenance Engineering, Vol. 8 No. 4, pp. 346-368.

Pun, K. and Lau, C. W. (2003), "Integration of total quality management and performance measurement in government departments: an empirical study in Hong Kong", International Journal of Business Performance Management, Vol. 5 No. 4, pp. 316.

Sanchez-Ruiz, L., Gomez-Lopez, R. and Blanco Rojo, B. (2022), "Key facilitators to continuous improvement: a Spanish insight", Business Process Management Journal, Vol. 28 No. 4, pp. 1071-1092.

Santos, P. J. et al (2025), "Powering efficiency: A Lean Six Sigma 4.0 agenda and roadmap for sustainable energy enhancement", Production Planning & Control, pp. 1-23.

Sauer, P. C., Seuring, S. (2023), "How to conduct systematic literature reviews in management research: a guide in 6 steps and 14 decisions", Review of Managerial Science, Vol 17 no 5, pp. 1899–1933.

Sila, I. and Ebrahimpour, M. (2003), "Examination and comparison of the critical factors of total quality management (TQM) across countries", International Journal of Production Research, Vol. 41 No. 2, pp. 235-268.

Singh, P., Singh, V. K. and Piryani, R. (2023), "Scholarly article retrieval from Web of Science, Scopus and Dimensions: a comparative analysis of retrieval quality", Journal of Information Science, 01655515231191351.

Singh, P., Singh, V. K. and Kanaujia, A. (2024), "Exploring the publication metadata fields in Web of Science, Scopus and Dimensions: possibilities and ease of doing scientometric analysis", Journal of Scientometric Research, Vol. 13 No. 3, pp. 715-731.

Sony, M. (2019), "Implementing sustainable operational excellence in organizations: an integrative viewpoint", Production & Manufacturing Research, Vol. 7 No. 1, pp. 67-87.

Sunder M, V. and Prashar, A. (2020), "Empirical examination of critical failure factors of continuous improvement deployments: stage-wise results and a contingency theory perspective", International Journal of Production Research, Vol. 58 No. 16, pp. 4894-4915.

Tessier, S. and Otley, D. (2012), "A conceptual development of Simons' levers of control framework", Management Accounting Research, Vol. 23 No. 3, pp. 171-185.

Wankhede, V. A., Vinodh, S. and Antony, J. (2025), "Empirical analysis of key operational characteristics of lean six sigma and Industry 4.0 integration", The TQM Journal, Vol. 37 No. 2, pp. 611-634.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Between Compliance and Excellence: A Diagnosis of Quality Culture in Portuguese Higher Education through External Evaluation Reports

Sílvia Santos [1], Patrícia Moura e Sá [2] and Leonor Pais [3]

1 University of Coimbra, Faculty of Psychology and Education Sciences, Coimbra, Portugal; silvia.santos@uc.pt
2 University of Coimbra, Faculty of Economics, Centre for Social Studies of the University of Coimbra (CES), Coimbra, Portugal; pmourasa@fe.uc.pt
3 University of Coimbra, Faculty of Psychology and Education Sciences, Center for Research in Neuropsychology and Cognitive and Behavioral Intervention (CINEICC), Coimbra, Portugal; leonorpais@fpce.uc.pt

Abstract - Purpose: ***The purpose of this study is twofold and interconnected: (i) to analyse how the discourse within external evaluation reports distinguishes between the structural existence of a Quality Management System (QMS) and the internalisation of a mature Quality Culture (QC); and (ii) to determine the extent to which these specific qualitative assessments of QMS performance influence and associate with the final accreditation outcomes and justifications issued by the A3ES Board of Directors. Design/Methodology/Approach: A qualitative research design was employed, involving a systematic document analysis of 97 institutional evaluation reports from the most recent evaluation cycle. Data were processed through inductive content analysis, following Bardin's framework. Findings: The analysis identifies a four-stage Quality Culture Maturity Model (QCMM): Compliance-Driven, Structural, Aligned, and Integrated. Results reveal a significant association between QMS maturity and accreditation results: 100% of non-accredited institutions were rated as having "Insufficient" QMS, while 93.7% of fully accredited HEIs demonstrated "Good" or "Very Good" performance in their QMS. Key systemic weaknesses include "decoupling" between formal procedures and daily operations, and a frequent failure to "close the loop" in improvement cycles. Practical Implications: The findings offer actionable guidance for Higher Education Institutions (HEIs) seeking to move beyond regulatory compliance towards the cultivation of genuine quality excellence. First, institutions should prioritise the systematic closure of the improvement cycle by transitioning from passive data collection to evidence-informed decision-making — a practice consistently associated with higher quality ratings in the evaluation reports analysed. Second, the development of integrated information systems emerges as a structural prerequisite for what may be termed quality intelligence: the institutional capacity to synthesise data across organisational units and transform fragmented metrics into transversal knowledge assets. Third, at the cultural and leadership level, quality must be reframed as a collective responsibility, with senior leadership positioned to present external evaluation not as a compliance burden but as a strategic instrument for institutional learning and improvement. Finally, Quality Management Systems should be deliberately expanded to encompass research activities and internationalisation processes, ensuring that quality frameworks are appropriate to the full scope of the contemporary HEI mission. Originality/Value: By utilising a full-sector regulatory dataset from the most recent evaluation cycle, this research extends existing maturity models into a developmental continuum that differentiates "possessing" a system from "living" a quality culture within higher education institutions.***

Keywords – Quality Culture, Higher Education, Institutional Accreditation, Quality Management Systems.

I. Introduction

In the last two decades, Higher Education Institutions (HEIs) across Europe have faced increasing pressure to demonstrate efficiency and transparency. Quality Management (QM) in Higher Education (HE) has emerged as a pivotal domain for the strategic development and modernisation of these institutions, serving as a fundamental instrument for accountability, continuous improvement, and the pursuit of academic excellence (Hildesheim & Sonntag, 2020; Rosa & Amaral, 2007).

At the European level, the 2015 revision of the Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG) represented a watershed moment, providing a shared framework for quality assurance underpinned by institutional autonomy, public responsibility, and a developmental ethos (ENQA, 2015).

The ESG were designed to be applied to all HE, delivering the framework for both internal and external quality assurance systems, contributing to shift the focus of Quality Assurance (QA) from a purely bureaucratic exercise to the promotion of a pervasive "Quality Culture." (EUA, 2006; ENQA, 2015).

"Quality Culture" is a complex construct that blends structural elements (tools, procedures, and responsibilities) with cultural elements (values, beliefs, and commitment) (EUA, 2006). Empirical evidence highlights that a mature Quality Culture (QC) is an indispensable prerequisite for the effectiveness of QM systems (Sursock, 2011; Hildesheim & Sonntag, 2020).

Recent research developed by Nygren-Landgärds et al. (2024) further elucidates this concept by identifying four consistent associated descriptive themes: 1) Structured and systematic approaches to quality; 2) Integrated and well-known processes across the institution; 3) A foundation of trust, commitment, and participation; 4) An active dynamic driven by personal responsibility. Ultimately, these authors argued that quality culture plays a central role in institutional

development, serving as a fundamental instrument for the strategic modernisation and accountability of HEI by moving beyond a purely bureaucratic exercise.

The ESG establish that "Institutions should undergo external quality assurance" to "verify the effectiveness of institutions' internal quality assurance" and "act as a catalyst for improvement" - external quality assurance may take different forms focusing not only on the evaluation of programmes, but also at institutional level (ENQA, 2015).

Reflecting broader European trends, the Agency for Evaluation and Accreditation of Higher Education (A3ES) has transitioned towards a multilevel model where institutional evaluation complements the assessment of individual programmes. This approach acknowledges that sustainable quality at the programme level is a direct function of the effectiveness of the institution's global Quality Management System (QMS) (A3ES, 2016 & 2022; Santos, 2011).

For HEIs, achieving full institutional accreditation is no longer just about meeting minimum technical standards; it requires demonstrating that quality processes are embedded in the institution's daily operations and strategic vision. In this context, the primary challenge for Portuguese HEIs lies in moving beyond "compliance culture" - where the QMS exists primarily to satisfy the regulator - towards a "consolidated quality culture." In the latter, quality is internalised by all stakeholders, leading to a state where improvement becomes an autonomous and shared responsibility (A3ES, 2022; Santos, 2011).

Under the current Portuguese legal framework (Law no. 38/2007), institutional evaluation is mandatory every six years, resulting in differentiated accreditation outcomes (A3ES, 2022). These outcomes - ranging from full accreditation to conditional accreditation or non-accreditation - carry significant regulatory, developmental, and reputational weight. Despite the wealth of institutional evaluation reports generated by A3ES since the first cycle began in 2016, there remains limited research on how the regulator's discourse defines and identifies a "consolidated" QMS. Specifically, it remains unclear how external evaluators distinguish between an institution that merely "possesses" a system and one that "lives" a quality culture.

This paper addresses this gap by employing a qualitative analysis of the reports produced by external evaluation panels during the most recent institutional evaluation cycle in Portugal. By cross analysing the evaluators' performance assessments of QMS with the final decisions of the A3ES Board of Directors, this study seeks to decode the specific linguistic markers and justifications that characterise a mature quality culture addressing the variances between regulatory requirements and institutional reality by exploring two primary research questions (RQ):

RQ1: how does the discourse within external evaluation reports distinguish between the structural existence of a QMS and the internalisation of a mature Quality Culture?

RQ2: to what extent do the specific qualitative assessments of QMS performance influence and associate with the final accreditation outcomes and justifications issued by the A3ES Board of Directors?

The findings offer concrete guidance for quality managers and institutional leaders navigating A3ES requirements - and, more broadly, for any HEI operating within an ESG-aligned quality assurance framework.

II. METHODOLOGY

This study adopts a qualitative research design, employing systematic document analysis of publicly available institutional evaluation reports and decisions as its primary method, to understand how QMS performance relates to overall institutional evaluation outcomes.

The primary data were sourced from the Agency for Evaluation and Accreditation of Higher Education. A comprehensive census of all publicly available institutional evaluation reports produced under the 2022 A3ES Manual (n=97), whose results were published in 2024, was conducted. These documents represent the official qualitative discourse of external peer review panels regarding the maturity of internal quality assurance systems. Additionally, the decisions of the A3ES Board of Directors were extracted, compiled, and analysed, together with their respective justifications for full accreditation, conditional accreditation, or non-accreditation of HEIs.

Content analysis targeted key components of these institutional evaluation reports and Board of Directors decisions: the rationale for overall evaluation results, the evaluation of QMS, and proposed improvement recommendations - the unit of analysis was the evaluators' narrative assessment of the 'Quality Management System' criterion within each institutional report, supplemented by the global justification section and Board of Directors decisions. Information was systematically extracted from the reports / decisions and organised into a structured database comprising:

- Institutional Metadata: Identification of the HEI, including designation, nature, typology, and geographic location.
- Quality Management Narratives: The full text of the assessments regarding the "Quality Management System" criterion.
- Accreditation Outcomes: The final global result of the institutional evaluation process (e.g. "Accredited" vs "Accredited with conditions" vs "Non-Accredited").

The research focuses on the thematic interpretation of institutional evaluation reports to understand how QMS performance relates to overall accreditation outcomes, employing a systematic and inductive content analysis approach. Following the procedural framework proposed by Bardin (2011), the analysis was conducted in three distinct phases: pre-analysis, material exploration, and the treatment of results through inference and interpretation. To ensure

systematic data organisation, a structured Excel grid was developed, allowing for the rigorous categorisation of textual units.

To support coding rigour, MAXQDA AI Assist served as an organisational and consistency-checking tool, with categories iteratively reviewed across the full corpus of 97 reports. Gemini Pro was additionally used during the pre-analysis phase to summarise and compare sections of the institutional evaluation reports. Trustworthiness was further strengthened by grounding the categorisation framework in the ESG and EUA quality benchmarks, ensuring alignment with established international standards.

To complement the content analysis, a comparative mapping was performed to examine the relationship between QMS evaluations and global accreditation results. Specifically, the analysis sought to determine whether outcomes of full institutional accreditation consistently occur in instances where the QMS assessment is more positive, aligned with a mature "Quality Culture". This allowed for an exploration of the importance attributed to quality management maturity within the broader evaluators decision-making process.

III. RESULTS AND DISCUSSION

This section presents the findings from the content analysis of A3ES institutional evaluation reports and Board of Directors decisions, exploring the relationship between QMS assessment, Quality Culture and global assessment results.

A. Institutional Evaluation – global data

According to the reports available on the A3ES website, the most recent institutional evaluation cycle, whose results were published in 2024, comprises all 97 HEIs operating in Portugal. This group includes university (27), polytechnic (61), and mixed education sectors (9), with a predominant share of private institutions (63%). Geographically, approximately 38% of these HEIs are located in the Lisbon and Tagus Valley region, while 36% are situated in the North. Overall, 32.99% of HEIs received full accreditation, 60.82% were conditionally accredited, and 6.19% were not accredited.

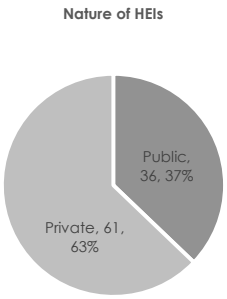


Fig. 1: Number and percentage of HEIs evaluated, by nature [n=97]

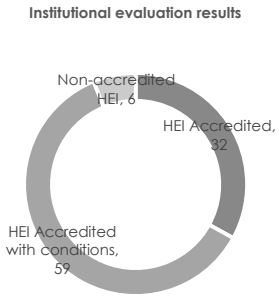


Fig. 2: Number of HEIs by institutional evaluation result [n=97]

B. Association Between Most Frequently Used Words/Expressions and Institutional Accreditation Results

A content analysis was conducted on the justifications provided by the A3ES Board of Directors, to substantiate the overall result of the institutional evaluation process, identifying the most frequently used words and expressions within this rationale, as presented in table 1.

Table 1: Most frequently used words and expressions by the A3ES Board of Directors

Most frequently used words and expressions	Rationale	*Frequency (n = 97 reports)*		
		Accredited [n = 32]	**Accredited with conditions** [n = 59]	**Non-accredited** [n = 6]
Ensure the quality of its performance	Rooted in the national legal framework (Law no. 38/2007), this establishes quality management as a mandatory institutional duty.	32	59	6
International good practices & ENQA standards	Used to benchmark the institutions against European higher education guidelines.	32	59	6
Mandatory disclosure	Frequently cited to emphasise transparency, specifically regarding the public availability of documents related to evaluation processes.	32	59	0
Quality culture	A central concept that the Accreditation Agency expects institutions to embed internally.	32	50	0
Progressively consolidated	Highlights that quality is viewed as a continuous, dynamic process rather than a static goal.	32	59	6
Internal quality management systems - QMS	Presented as a vehicle for achieving and maintaining the quality culture.	32	59	6
Trust-responsibility binomial	A core expression framing the relationship between the higher education institution and society.	32	49	0

While the foundational text presented by A3ES is nearly identical for most HEI, specific expressions branch out depending on the final accreditation result:

- **Fully Accredited & Accredited with Conditions:** The narrative emphasises that the quality culture must be "progressively consolidated through the adoption and continuous improvement of internal quality management systems." For institutions that are *Accredited with conditions*, A3ES explicitly adds an expression reiterating "the need for a careful analysis of all the Recommendations for improvement suggested by the CAE [External Evaluation Commission]." In cases with severe conditions, explicit mandates are written into the justification, such as "Demonstrate the full implementation of the Internal Quality Assurance System.
- **Not Accredited:** The language shifts significantly. The standard compliance text is replaced by a broader list of foundational requirements that the institution has failed to meet. Expressions uniquely associated with non-accreditation include "clarification and consensus of its strategy", "affirmation of its educational offer with a high success rate", and "placing students at the centre". The ultimate justification for non-accreditation is the failure to confirm the "recognition of public interest". This expression is exclusively used for non-accredited institutions to signal that they lack the necessary trust and responsibility demanded by society.

Regarding the justifications for each type of result, in a more detailed way:

Full Accreditation - Justifications found in the reports for full accreditation (typically for a 6-year period) are characterised by a discourse of systemic consolidation and congruence:

- Structural Alignment: The primary motive is the high level of agreement between the A3ES Board of Directors and the External Assessment Commission. The justifications emphasise that the institution has effectively internalised the "obligations to guarantee quality performance."
- Quality Culture Maturity: Beyond mere compliance, these justifications highlight the effectiveness of communication within the QMS. A recurring theme is the transition from formal systems to lived practices, where quality information is understood by all stakeholders—from top management to students.
- Strategic Continuous Improvement: Even in cases of success, A3ES Board of Directors reiterates specific areas for "careful analysis," such as the formalisation of research strategies or the expansion of quality systems to new organic units, signalling that full accreditation is not an endpoint but a stage in a continuous improvement cycle.

Accreditation with Conditions - Accreditation with conditions (usually for 1 or 3 years) reflects a state of transitional compliance or specific transparency deficits.

- Transparency and Public Accountability: A distinctive feature in this category is the explicit reference to the "mandatory disclosure of documentation on the institutional website." A3ES often uses this outcome to signal institutions that fail to maintain a transparent public record of their evaluation processes and results.
- Divergent Expert Opinions: The justifications frequently reveal a "discordance" (either favourable or unfavourable) between the A3ES Board of Directors and the external evaluators recommendation. This

suggests that conditions are often imposed when the Board identifies risks or gaps that the initial external evaluators might have appraised differently, particularly regarding the sustainability of the quality system.

- Supervised Development: The motive presented is the "necessity of a careful analysis of all improvement recommendations." This indicates that the institution possesses the foundational elements of quality management but lacks the maturity to operate without direct A3ES oversight or "follow-up" conditions.

Non-Accreditation - Justifications for non-accreditation focus on foundational failures and the absence of institutional coherence.

- Failure in Strategic Definition: The discourse shifts from "maintenance" to "requirement." A3ES emphasises that quality must be "progressively consolidated through the clarification and consensus-building of strategy." The motive for denial is often the lack of a clear, shared institutional vision.
- Fragmented Systems: The justifications highlight deficiencies in the "adoption of internal quality management systems" and a failure to place "students at the centre of teaching and research processes."
- Social Responsibility Gap: Non-accreditation is framed as a failure to meet the "trust and responsibility demanded by society." The justifications suggest that the institution has not provided sufficient evidence of scientific development or internationalisation, thereby failing to justify its "public interest recognition."

C. QMS / Quality Culture performance

The performance assessment of HEIs concerning "Quality Management" yielded the following results (Figure 3): insufficient (n=15), sufficient (n=43), good (n=33), and very good (n=6).

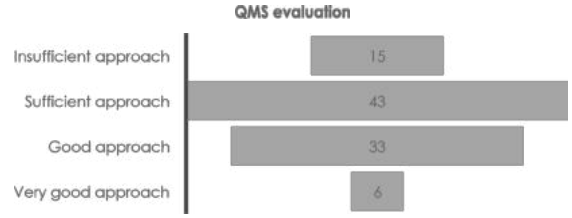


Fig. 3: Number of HEIs by QMS performance evaluation [n=97]

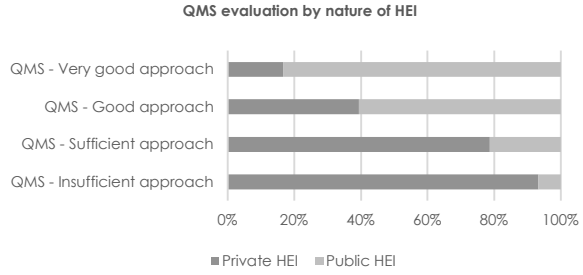


Fig. 4: QMS performance evaluation by nature of HEIs [n=97]

As shown in Figure 4, private institutions are disproportionately represented at lower QMS performance tiers, a pattern that warrants further study.

D. Recurring Weaknesses in QMS

Institutional evaluation reports consistently identify several systemic vulnerabilities in the QMS of HEIs. The primary shortcomings are summarised as follows:

1. **Incipient Quality Culture:** Quality processes are often perceived as a top-down bureaucratic burden rather than being embedded in the academic community's values. This results in a lack of alignment between quality policies and the institution's overarching educational project.
2. **Low Maturity and Incomplete Implementation:** Many systems remain in developmental or informal stages. A significant "decoupling" is often observed, where formal procedures documented in Quality Manuals are not effectively applied in daily operations, hindering system functionality.
3. **Limited Scope (Excessive Focus on Teaching):** QMSs are frequently restricted to teaching and learning activities. They often fail to integrate other core institutional pillars, such as scientific research, internationalisation, innovation, and knowledge transfer to society.

4. **Deficient Information Systems and Lack of Interoperability:** Data fragmentation across non-integrated platforms remains a critical barrier. The absence of robust, automated information systems hampers the collection of key performance indicators (KPIs), diverting time from critical analysis and evidence-based strategic decision-making.
5. **Incomplete Improvement Cycles and Absence of Meta-evaluation:** Institutions frequently fail to "close the loop" (Plan-Do-Check-Act). There is a notable lack of critical reflection, follow-up, and meta-evaluation of the QMS itself, leading to poor monitoring of data-driven improvement plans.
6. **Low Stakeholder Engagement:** Participation remains weak among key groups, specifically students, non-academic staff, alumni, and external partners. Their involvement in monitoring structures and consultative processes is often superficial or non-existent.
7. **Lack of Transparency and Poor Communication:** Many HEIs fail to meet public disclosure requirements. Deficiencies in publishing quality instruments and results (e.g., annual course reports or pedagogical surveys) on institutional websites represent a significant barrier to accountability.
8. **Resource Scarcity and Lack of Professionalisation:** There is a shortage of dedicated human resources exclusively assigned to quality management. Furthermore, there is a pressing need to professionalise quality offices to enhance their capacity for effective system monitoring.

E. Main fragilities in QMSs assessed as "insufficient"

Table 2 presents the frequency of the weaknesses identified by the External Evaluation Committees across the 15 HEIs in which the QMSs were assessed as "insufficient".

Table 2: Main fragilities in QMSs assessed as "insufficient" [n = 15]

Fragility	Frequency [n = 15]
Incipient Implementation & Lack of Formalisation	15
Deficient Monitoring, Evaluation, & Meta-evaluation	15
Low Stakeholder Participation & Engagement	14
Inadequate or Non-integrated Information Systems	12
Limited Scope (Focus on Teaching vs. Research/Society)	10
Lack of Public Transparency & Disclosure	9
Human Resource Constraints & Professionalisation	8
Governance & Structural Misalignment	8
Weak Quality Culture & Internalisation	7

The data reveals a critical state of "Incipient Implementation" and "Deficient Monitoring" across all 15 HEIs. This suggests that while Quality Policies are often formally defined, they lack the necessary operational mechanisms to achieve functional maturity. Also, a fundamental failure in the Plan-Do-Check-Act (PDCA) cycle is evident. There is a systemic absence of meta-evaluation reports and a failure to close the continuous improvement cycle.

F. A Four-Stage Quality Culture Maturity Framework

Through the content analysis of the reports produced by the External Evaluation Panels a four-stage Quality Culture Maturity Model (QCMM) in HEIs was conceptualised. This model stems from an integration of international standards and European strategic frameworks. Primarily, the model draws its foundational logic from the European University Association (EUA) and its seminal "Examining Quality Culture" project (EUA, 2006; Sursock 2011). The EUA posits that a mature quality culture is the result of an interplay between a "structural/managerial" element (standardised processes) and a "cultural/psychological" element (shared values and commitment).

The model also integrates the guidelines set forth by ENQA in the ESG 2015, particularly regarding the imperative of "Information Management" (Standard 1.7). This ensures that the progression through maturity levels is intrinsically linked to the institution's ability to transform raw data into actionable knowledge.

Complementing this sector-specific approach, the model adopts the architectural principles of ISO 9004:2018, which focuses on the "quality of an organisation" and its "sustained success". By aligning ISO's focus on self-assessment and organisational learning with the EUA's qualitative benchmarks, the four levels (see Table 3) reflect a trajectory from mere compliance-based management (Level 1) to a culture where quality is institutionally owned (Level 4), no longer a peripheral administrative task but a core strategic competency.

Table 3: Contributions to the development of the four-stage Quality Culture Maturity Model (QCMM)

Source	Contribution to the four-stage Quality Culture Maturity Model (QCMM)
EUA (Quality Culture Project)	The distinction between "Staff Commitment" and "Structural Systems"
ENQA (ESG 2015)	The standardisation of internal and external quality loops and the requirement for robust information systems.
ISO 9004	The maturity-level logic (from "Initial" to "Best-in-class") and the focus on the organisation's long-term sustainability.

Table 4 synthesises the main findings that emerged from the various reports analysed. Each HEI was placed into one quadrant depending on the dominant evidence.

Table 4: Quality culture stage and empirical evidence [n=97]

QMS Evaluation Tier	Institutional Evaluation possible outcome	Number of HEIs	Quality culture stage	Empirical evidence	Organisational insight	Cultural paradigm
Insufficient	Non-accreditation risk	15	The Compliance-Driven Stage	The semantic analysis is dominated by terms indicating systemic deficits: ***lack***, ***absence***, and ***low***. Evaluators frequently noted a "lack of mandatory institutional information," an "absence of student and external representation," and "fragile monitoring of the teaching-learning process."	There is a pronounced disconnect between data collection and utilisation. Indicators are heavily focused on basic outputs rather than efficacy or impact, leading to a failure in closing the continuous improvement loop.	Defensive / Disengaged
Sufficient	Conditional Accreditation zone	43	The Structural Stage	Recurring themes include ***structures*** and ***incipient***. The data explicitly points to "structures that are created and functioning reasonably," yet highlights a critical "lack of global integration of the quality policy into the educational project."	Institutions at this level suffer from structural silos. The QMS operates in parallel to the institution's core pedagogical and research missions rather than being embedded within them. Information systems are often described as rudimentary.	Siloed / Procedural
Good	Full Accreditation zone	33	The Aligned Stage	The vocabulary shifts positively toward ***continuous***, ***aligned***, and ***involvement***. Evaluators highlight a "consolidated culture of quality and continuous improvement" that is "aligned with the educational project."	At this tier, quality management ceases to be the sole responsibility of a single department. However, the evaluators frequently mention the need to further develop Information Systems to ensure that data flows seamlessly across different academic and administrative units.	Proactive / Aligned
Very Good	Full Accreditation zone	6	The Integrated Stage	The discourse centres on ***academic community***, ***comprehensive***, and ***effective***. QMS structures are noted as having a "positive impact on the organisation and the satisfaction of the academic community," with highly effective feedback mechanisms.	Quality culture is embedded in how the institution operates. Data regarding teaching, research, and societal cooperation is systematically gathered, analysed, and transformed into actionable institutional intelligence.	Internalised / Embedded

The **Compliance-Driven Stage**: Insufficient Maturity - At the lowest tier, into which 15 HEIs fall, the QMS is characterised by a defensive and disengaged culture. The semantic landscape of these evaluations is dominated by terms such as "lack", "absence", and "low", particularly regarding mandatory institutional information and the representation of external stakeholders. These institutions suffer from a fundamental failure in "closing the loop" of continuous improvement. Data collection is treated as an isolated bureaucratic requirement rather than a catalyst for efficacy or impact. This stage reflects a failure to meet the basic requirements of the ESG, where monitoring is fragile and institutional intelligence is virtually non-existent.

The **Structural Stage**: Sufficient Maturity - The majority of evaluated HEIs (43) are at the Structural Stage, where the QMS is often described as "incipient". While these institutions have successfully established the necessary formal

structures, they struggle with structural silos. The quality policy remains an adjunct to the core pedagogical and research missions rather than an integrated component of the educational project. The "Siloed / Procedural" paradigm indicates that while the machinery of quality exists, it lacks global integration. Information systems at this level are frequently rudimentary. In these cases, quality is performed at the institution rather than by the institution, mirroring a transition phase where procedural adherence precedes cultural internalisation.

The **Aligned Stage**: Good Maturity - A significant cohort of 33 HEIs has reached the Aligned Stage, where the organisational discourse shifts toward "involvement" and "continuous improvement". At this tier, quality management is no longer confined to a single department; instead, it is "aligned with the educational project" and involves a proactive academic community. The primary challenge at this level of maturity is the technical and flow-based integration of data. Evaluators consistently identify a need to enhance Information Systems to ensure that knowledge flows seamlessly across different academic and administrative units. While the "Proactive / Aligned" paradigm is evident, the full potential of a QMS is restricted by these lingering infrastructural bottlenecks.

The **Integrated Stage**: Very Good Maturity - The "Integrated Stage," achieved by only 6 HEIs, represents the benchmark for quality culture within the EHEA. Here, the QMS is not just a system but a core element of organisational behaviour, deeply "internalised and embedded". The discourse focuses on the "academic community" as a whole, with feedback mechanisms that demonstrate high levels of effectiveness and satisfaction. In these institutions, data concerning the pillars of HE - teaching, research, and societal cooperation - is systematically transformed into "actionable institutional intelligence". This stage exemplifies the complete maturation of the PDCA (Plan-Do-Check-Act) cycle, where the "Act" phase is driven by robust evidence and a shared cultural commitment to excellence.

Figure 5 illustrates the Quality Culture Maturity Model (QCMM), delineating the evolutionary trajectory from basic regulatory compliance to a fully integrated stage characterised by strategic alignment and deep-rooted academic engagement.

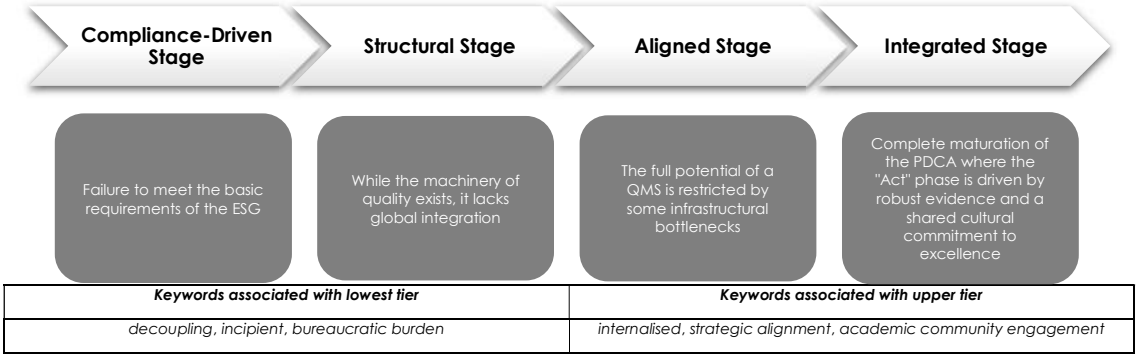


Fig. 5: QMS performance stages & keywords

G. Quality Culture maturity & Institutional evaluation results

An analysis of the cross-tabulated data reveals a strong and consistent association between QMS performance evaluation and institutional accreditation outcomes (Figure 6).

- **Non-Accredited Institutions:** 100% of the non-accredited institutions (6 out of 6) received an "Insufficient" rating for their QMS. An inadequate quality system acts as a predictor of institutional failure.
- **Accredited with Conditions:** These institutions predominantly fall into the middle of the spectrum. The vast majority (41 out of 59) received a "Sufficient" rating, whilst 9 were "Insufficient" and 9 were "Good". A QMS that is merely "Sufficient" or "Insufficient" almost invariably triggers conditions for improvement, directly linking A3ES's conditional demands to the operational immaturity of the institution's quality mechanisms.
- **Accredited Institutions (Unconditional):** 93.7% of fully accredited institutions (30 out of 32) achieved a "Good" (24) or "Very Good" (6) evaluation for their QMS. Zero fully accredited institutions had an insufficient QMS, proving that a mature QMS is a fundamental prerequisite for unconditional accreditation.

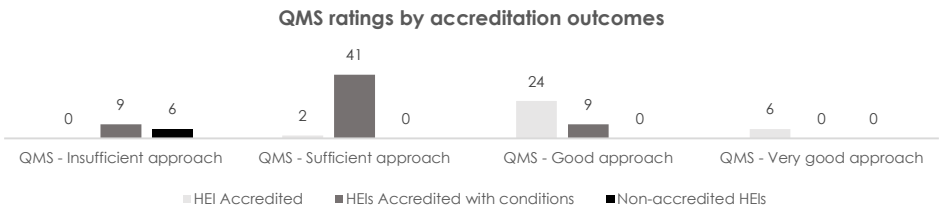


Fig. 6: QMS ratings broken down by accreditation outcome [n = 97]

IV. CONCLUSION AND PRATICAL IMPLICATIONS

This study examined how QMS evaluations relate to institutional accreditation outcomes across the Portuguese HE sector. Systematic analysis of A3ES evaluation reports and Board decisions makes clear that full accreditation is no longer a matter of procedural compliance alone. Instead, success is deeply rooted in the maturity of an institution's Quality Culture.

Academic discourse reflects a conceptual plurality mirrored in the diverse Quality Management methodologies adopted by HEIs, ranging from prescriptive normative models - such as ISO 9001 or the EFQM Excellence Model - to participatory approaches grounded in organisational learning (Bendermacher et al., 2017; Manatos, Sarrico & Rosa, 2017). Synthesising these perspectives with the empirical data, this study proposes a four-stage Quality Culture Maturity Model (QCMM): Compliance-Driven, Structural, Aligned, and Integrated. The four maturity stages identified resonate with the descriptive dimensions of Nygren-Landgärds et al. (2024) - particularly the progression from 'structured approaches' to 'integrated processes' and 'trust-based participation'. Notably, the present study extends this framework by grounding the stages in regulatory discourse, offering an operationalised, evaluator-validated continuum. While Hildesheim & Sonntag's (2020) Quality Culture Inventory measures quality culture through staff perceptions, the present model complements this approach by deriving its stages from external regulatory discourse, offering a systemic-level lens rather than an individual-level one.

The analysis reveals significant heterogeneity across the Portuguese HE sector. While most institutions have attained the *Structural* or *Aligned* stages, only a distinct minority has achieved a fully *Integrated* quality culture. This four-tier taxonomy illustrates that the primary obstacle in contemporary quality management is no longer the establishment of formal procedures, but rather the internalisation of these processes into the institutional fabric. In lower maturity tiers, a "defensive" posture and "structural silos" prevent the QMS from driving continuous improvement. As institutions progress, the focus shifts from administrative adherence to the effective flow of information. However, even at the "Good" level, technical bottlenecks in Information Systems often impede the smooth transition of data into actionable institutional intelligence. It is only at the *Integrated* stage that the QMS fulfils its mission, enabling a comprehensive understanding and improvement of teaching, research, and societal impact.

In response to RQ1, the content analysis demonstrates that evaluators discourse distinguishes structural QMS existence from cultural internalisation through a consistent set of lexical markers — 'decoupling', 'incipient', and 'absence' signalling lower maturity, and 'consolidated', 'aligned', and 'academic community' signalling higher maturity.

In response to RQ2, the cross-tabulated data establish a strong and consistent association: QMS maturity ratings are near-deterministic predictors of accreditation outcomes, with all six non-accredited institutions rated Insufficient and 93.7% of fully accredited institutions rated Good or Very Good.

These findings corroborate and extend the observations of Bendermacher et al. (2017). Drawing on a full-sector regulatory dataset (the second institutional evaluation cycle conducted under the 2022 A3ES Manual, whose reports were published in 2024 (n=97)) rather than isolated case studies, this research demonstrates that a successful QMS cannot rely exclusively on "hard" structural frameworks. The data confirms a direct and consistent association between QMS maturity and accreditation outcomes.

Based on the empirical evidence, the following strategic roadmap is proposed for HEIs seeking to navigate the regulator's qualitative demands and foster authentic excellence:

1. **Prioritise "Closing the Loop":** HEIs must shift from mere data collection to data utilisation. This is the most consistent differentiator between "Sufficient" and "Good" ratings. Ensuring that feedback mechanisms have a demonstrable impact on the academic community is essential.
2. **Invest in Integrated Information Systems (IS):** Robust IS are a precondition for quality intelligence. To move beyond siloed paradigms, institutions must develop platforms that bridge the gap between academic and administrative units, ensuring data becomes transversal.
3. **Cultural Internalisation and Leadership:** Quality must be fostered as a shared responsibility rather than a departmental burden. Leadership should encourage proactive organisational behaviour where external evaluation is perceived as a strategic opportunity for growth rather than a bureaucratic hurdle.
4. **Expansion of Scope:** Beyond teaching, the QMS must be expanded to include research and internationalisation, reflecting the holistic nature of institutional mission and impact.

In conclusion, as the HE system evolves, the pressure for transparency and accountability will intensify. The most resilient institutions will be those that convert external regulatory requirements into authentic institutional intelligence, shared across the entire academic community.

Whilst this research provides a systematic exploration of the relationship between QMS and institutional accreditation outcomes, certain limitations must be acknowledged to contextualise the findings accurately:

- Nature of Secondary Data - The study relies exclusively on secondary data derived from the Accreditation Agency institutional evaluation reports and Board of Directors decisions. These documents represent a "formalised reality" as perceived and articulated by external review panels / A3ES. Consequently, the findings

reflect a regulatory and evaluative lens, which may not fully capture the informal complexities, internal tensions, or the "lived" organisational culture within the HEIs.

- Subjectivity and Linguistic Interpretation - A significant challenge in the qualitative analysis of these reports is the subjectivity inherent in peer review. Different evaluation panels may employ varying terminologies or criteria when describing a "consolidated quality culture". While the content analysis sought to standardise these expressions, the interpretative nature of the source material means that the linguistic markers identified are dependent on the specific discourse of the evaluators involved.
- Coding and categorisation exercise – The content analysis and coding were conducted by a single researcher. To mitigate this risk, MAXQDA AI Assist was employed as a consistency-checking tool, and an iterative review process was applied across the full corpus. Nonetheless, the absence of a second independent human coder remains a limitation that future replications of this study should address.
- Contextual Generalisability - The scope of this study is bounded by the Portuguese HE regulatory framework. While the findings offer valuable insights into quality maturity, the specific relationship between QMS assessment and global accreditation outcomes may be influenced by national legislation and Accreditation Agency specific guidelines. Therefore, generalising these results to other national contexts with different quality assurance traditions should be approached with caution.

V. Future Research

Building upon the limitations and findings of this study, several avenues for future inquiry are possible, including: 1) validation of the four-stage maturity model across other HE quality assurance systems; 2) capturing the "lived experience" of staff through qualitative interviews or focus groups in order to reveal whether the "consolidated culture" described by external panels aligns with internal perceptions; 3) Longitudinal analysis of HEIs across multiple evaluation cycles to track the evolution of quality culture language.

References

Agência de Avaliação e Acreditação do Ensino Superior. (2016). Referenciais para os sistemas internos de garantia da qualidade nas instituições de ensino superior. A3ES. https://www.a3es.pt/sites/default/files/Referenciais_SIGQ_PT.pdf (accessed February 2026)

Agência de Avaliação e Acreditação do Ensino Superior. (2022). Manual de avaliação institucional do ensino superior 2022. A3ES. https://www.a3es.pt/sites/default/files/Manual_Avaliacao_Institucional_2022.pdf (accessed February 2026)

Bardin, L. (2011). Análise de conteúdo (L. A. Reto & A. Pinheiro, Trads.; 5.ª ed.). Edições 70

Bendermacher, G. W. G., Egbrink, M. G. A., Wolfhagen, I. H. A. P., & Dolmans, D. H. J. M. (2017). Unravelling quality culture in higher education: A realist review. Higher Education, 73(1), 39–60. https://doi.org/10.1007/s10734-015-9979-2

European Association for Quality Assurance in Higher Education (ENQA). (2015). Standards and guidelines for quality assurance in the European Higher Education Area (ESG) (3.ª ed.). https://doi.org/10.12657/9788394204204

European University Association. (2006). Quality culture in European universities: A bottom-up approach. https://eua.eu/resources/publications/655:quality-culture-in-european-universities-a-bottom-up-approach.html (accessed February 2026)

Hildesheim, C., & Sonntag, K. (2020). The Quality Culture Inventory: A comprehensive approach towards measuring quality culture in higher education. Studies in Higher Education, 45(4), 892–908. https://doi.org/10.1080/03075079.2019.1672641

International Organization for Standardization. (2018). Quality management — Quality of an organization — Guidance to achieve sustained success (ISO Standard No. 9004:2018). https://www.iso.org/standard/70397.html

Lei n.º 38/2007 de 16 de agosto. (2007). Diário da República: I série, n.º 157. Assembleia da República. https://diariodarepublica.pt/dr/detalhe/lei/38-2007-639553 (accessed February 2026)

Manatos, M. J., Sarrico, C. S., & Rosa, M. J. (2017). The integration of quality management in higher education institutions: A systematic literature review. Total Quality Management & Business Excellence, 28(1-2), 159–175. https://doi.org/10.1080/14783363.2015.1050178

Nygren-Landgärds, C., Mårtensson, L. B., Pyykkö, R., Bjørnestad, J. O., & von Schoultz, R. (2024). Quality culture at Nordic universities. European Journal of Higher Education, 14(1), 40–59. https://doi.org/10.1080/21568235.2022.2116066

Rosa, M. J., & Amaral, A. (2007). A self-assessment of higher education institutions. In Quality assurance in higher education (pp. 129–156). Springer. https://doi.org/10.1007/978-1-4020-6012-0_6

Santos, S. M. (2011). Análise comparativa dos processos europeus para a avaliação e certificação de sistemas internos de garantia da qualidade. Agência de Avaliação e Acreditação do Ensino Superior. https://www.a3es.pt/sites/default/files/A3ES_Estudo1_Garantia_Qualidade.pdf (accessed February 2026)

Sursock, A. (2011). Examining quality culture: Part II. Processes and tools – Participation, ownership and bureaucracy. European University Association. https://eua.eu/resources/publications/386:examining-quality-culture-part-ii.html (accessed February 2026)

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Standardisation as a strategic motor for interoperability in BIM

Pedro Oliveira [1]

1 CTCV – Centro Tecnológico da Cerâmica e Vidro; pedro.oliveira@ctcv.pt

The construction sector has a significant impact on the community, both in Portugal and worldwide. In addition to being an operational area with a considerable degree of standardisation, it is involved in structural supply chains and in design and construction processes comprising a complex combination of different materials, specialisms and functionalities.

The collaborative processes associated with the BIM (Building Information Modelling) methodology are currently undergoing global expansion, particularly in Portugal. This implementation generates new possibilities for more efficient and consistent working practices but simultaneously requires significant adjustments to traditionally established workflows.

The uptake of BIM in Portugal, although growing, is still far from the maturity seen in other sectors or European countries. Some pioneering companies and large construction groups have already digitised their operations, incorporating BIM into flagship projects, particularly in public works and international markets.

Decree-Law No. 10/2024 mandates the use of BIM in all public projects from 2030 onwards, and this milestone is expected to drive the adoption of the methodology across the board, particularly in terms of the creation and integration of national parametric digital libraries, the training of technical staff, and regulatory alignment.

Despite the benefits associated with the implementation of the BIM methodology, it requires a systemic approach that combines technical standardisation, legislative adaptation and the capacity-building of actors in the value chain, ensuring the technical, economic and environmental sustainability of the built environment.

This paper presents a bibliographic review of the main standards associated with the BIM methodology, aiming to frame the technical content applicable to project development and subsequent implementation. It systematically describes these standards while identifying the potential difficulties related to their adoption within the national business context, as well as the key challenges faced by stakeholders across the construction value chain.

Keywords – Construction, Interoperability, Methodology, Standardisation.

I. INTRODUCTION

The architecture, engineering and construction (AEC) industry is rapidly adopting digital technologies and processes, driving structural changes in conventional working methods and representing a strategic opportunity for innovation and differentiation within the sector.

The BIM (Building Information Modelling) methodology is becoming established internationally, with its implementation in Portugal progressing in line with global trends. It is a restructuring and evolutionary process, expected to be developed gradually across all phases of the production cycle, from the design phase through to final use.

The benefits of this methodology include greater efficiency, transparency and coordination in project development, enabling the various stakeholders in the process (architects, engineers, contractors and material suppliers) to collaborate and centralise all relevant information from the construction chain on a single digital platform or system. [1]

Thus, the digital transformation of the process, including the entire operationalisation of construction work, is crucial to enabling the various stakeholders to contribute to a properly organised and structured digital model.

The use of this methodology enhances interoperability, reduces data loss and optimises process efficiency, resulting in time and cost savings throughout the construction lifecycle. [1]

However, the complexity of the construction chain, coupled with communication difficulties between the various stakeholders and the use of different working systems, leads to incorrect and inconsistent interpretations, errors and inefficiencies, compromising effective decision-making. [1]

Standardisation is fundamental, as it establishes a uniform framework for data and working methods, ensuring consistency and compatibility across multiple sources, as well as compliance with predefined requirements, thereby improving the reliability and accuracy of the information essential to successful construction projects.

Information associated with the life cycle of construction projects must integrate technical, operational, sustainability, maintenance and performance data, as required by applicable regulatory frameworks, such as the CE marking in the EU, as set out in the Declaration of Performance (DoP) or the Environmental Product Declaration (EPD). [4]

Whilst the adoption of BIM methodology brings benefits, it also has implications for all legislation and standards applicable to construction, resulting from the increasing digitalisation of construction processes and systems, the need

for systems that ensure their interoperability, growing market demands in terms of sustainability, and the new responsibilities assigned to stakeholders throughout the construction process lifecycle. [5] [6]

II. The Digital Revolution and BIM Culture

Building Information Modelling involves the design, construction, operation and maintenance of buildings or infrastructure through shared parametric digital models. BIM is more than just a simple "3D design"; it integrates detailed, multidimensional information, including geometry, material properties, costs, schedules (4D), energy and environmental analyses (5D and 6D) and maintenance data (7D). Its core principles include:

· Data centralisation: All stakeholders work from a single digital model, minimising errors and rework, and promoting more informed and realistic decision-making.

· Multi-stakeholder collaboration: It facilitates the simultaneous contribution of diverse inputs (architects, engineers, builders, manufacturers, owners, etc.) within a shared virtual environment, anticipating potential issues and simplifying technical changes.

· Interoperability: The use of standardised formats (e.g. IFC – Industry Foundation Classes) ensures that information can be transferred between software and phases without loss or error.

· Comprehensive lifecycle management: The model tracks the entire process in real time, from the preliminary design phase through to the end of the asset's life, recording all relevant changes.

· Analysis and Automation: It enables simulations, automated cost estimates, environmental impact forecasting, advanced logistics management, predictive maintenance analysis, and more. [1]

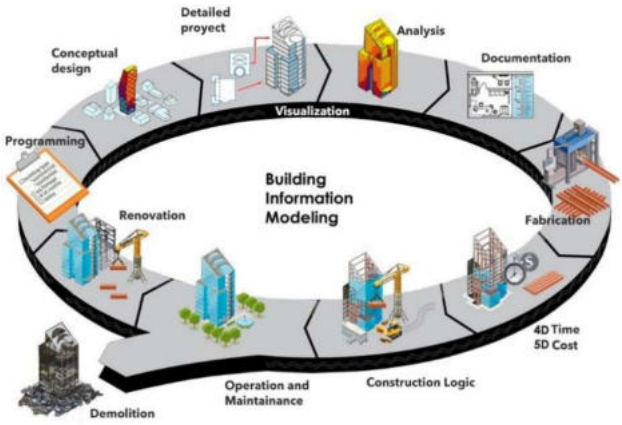


Figure 1 – BIM life cycle (Source: IpsaBIM)

The digital revolution is no longer a future trend, but a present-day necessity. The construction sector is undergoing a transition from the era of two-dimensional technical drawing (CAD) to three-dimensional and multi-dimensional digital representations (BIM), capable of integrating multiple technical aspects into a single digital model. This evolution is largely supported by European policies, with the European Union's promoting the digitalisation of the sector as an essential driver for competitiveness and sustainability.

BIM goes beyond 3D, enabling the virtual simulation of all phases of an asset's life cycle (from design to demolition), the analysis of performance, the making of informed decisions in real time, the creation of digital component libraries, accurate budgeting and planning, and the facilitation of communication and collaboration among all stakeholders. Digitalisation therefore implies not only a technological transformation, but also a cultural one, requiring new skills, mindsets open to collaboration and sectoral integration, as well as the abandonment of fragmented and repetitive practices. [1] [23]

III. Legislative and Regulatory Framework in the European Union

Legislation and policies promoting Building Information Modelling (BIM) in the European Union have come to play a pivotal role in the economic transformation of the construction sector, simultaneously driving down costs, improving the efficiency of public spending and strengthening the international competitiveness of businesses. Rather than merely encouraging the voluntary adoption of digital tools, these regulatory and strategic instruments frame BIM as information management infrastructure serving a more efficient, transparent and sustainable construction sector. In this context, the economic impacts stem both from improvements in internal design and construction processes and from the reconfiguration of relationships between public bodies and private companies in the procurement of public works. [5] [6] [23] [24]

The harmonisation of BIM requirements in public works across the European Union is one of the central pillars of the construction sector's digitalisation, but its implementation reveals significant variation among Member States. Building on the common framework provided by Directive 2014/24/EU on public procurement and the guidelines of the EU BIM Task Group, each country has set its own pace, scope and level of detail for requirements, producing a regulatory patchwork that directly influences the market's BIM maturity and the competitiveness of companies. [6] [7] [24]

At European level, the legislation does not establish a 'single package' of minimum BIM requirements, but creates the conditions for Member States to require the use of building information modelling tools in public contracts. This framework is complemented by manuals and strategic guides that recommend the definition of requirements in terms of information management, aligned with standards such as EN ISO 19650, with an emphasis on processes, responsibilities, interoperable formats and quality criteria. Thus, the European focus is more on the 'how' of structuring information requirements than on a fixed list of deliverables or levels of detail. [21]

A comparative analysis of European countries reveals significant differences across four dimensions: degree of mandatory compliance, type of requirements, level of detail and regulatory support. Countries with greater digital maturity, such as the Nordic countries, the United Kingdom (although no longer part of the EU, but frequently taken as a reference) and some Member States with robust national plans, have adopted comprehensive BIM mandates for public projects, linking them to technical guides, data specifications and support platforms. In these contexts, the BIM requirement is accompanied by explicit requirements regarding BIM implementation plans, common data environments, open formats, models by discipline and by phase, as well as minimum information attributes to support design, construction and operational decisions. [20] [23]

In other countries, adoption is more gradual and sector-specific, limiting mandatory requirements to works above certain value thresholds, pilot projects or strategic sectors, such as transport or healthcare infrastructure. In such cases, BIM requirements are often formulated in general terms ("BIM design", "delivery of a three-dimensional model"), with less alignment with international standards and little clarification of information levels, verification criteria or performance requirements. This approach creates significant scope for interpretation, requiring detailed negotiation between project owners and designers to translate the legal mandate into operational practices. [1] [20] [24] [25]

The type of requirements also varies between countries: some structure the requirements primarily around processes and information governance, whilst others attempt to standardise the informational content. The former focus on ensuring the existence of a common data environment, the appointment of BIM coordinators, the preparation of implementation plans and the definition of approval workflows, leaving greater technical freedom to the design team. The latter seek to standardise object libraries, classification systems, minimum attributes per discipline and coding rules, moving towards a more prescriptive view of BIM modelling. In both cases, reference to EN ISO 19650 is appearing with increasing frequency as the structural backbone of information management processes. [1] [2] [21]

The level of detail in the requirements is another distinguishing feature. In more advanced contexts, public works specifications clearly state which models must be delivered at each stage (preliminary study, detailed design, construction and as-built), which data fields are mandatory (e.g. classification codes, performance properties, maintenance information), which file formats are accepted and what verification procedures apply (clash detection, attribute validation, version control). This precision reduces contractual ambiguities, improves the predictability of implementation costs and facilitates the evaluation of tenders. In contrast, in countries and organisations where BIM is still in its infancy, a reduced or inconsistent level of detail tends to prevail, which can lead to models that formally meet the requirements but do not generate effective gains in quality, cost or time. [1] [2] [14]

In this scenario, technical standardisation plays a mediating role between regulatory diversity and the need for a more cohesive digital single market. Convergence around the EN ISO 19650 family for information management, the use of interoperable formats and common principles for defining information requirements helps to reduce technical barriers and facilitate the participation of companies in cross-border tenders. This convergence is, however, incomplete: many Member States continue to develop national guidelines and specifications with varying levels of compatibility between them, which maintains a certain degree of fragmentation. [8] [9] [22]

In terms of implications for practical implementation, a comparison of BIM requirements across EU countries shows that regulatory maturity is not limited to the presence or absence of a legal mandate. The effectiveness of legislation depends on the BIM requirement being accompanied by: i) clear and coherent regulatory frameworks, ii) procurement guidelines and requirement templates usable by public bodies, iii) capacity-building tools for buyers and designers, and iv) mechanisms for monitoring and progressively adjusting requirements. Countries that have combined these elements are able to transform BIM into a lever for efficiency, transparency and innovation; those that merely 'mention BIM' in the law, without structuring the requirements in an operational manner, face greater risks of formalism, market resistance and results falling short of expectations. [25]

In this regard, the current trend points towards a progressive European convergence, not so much towards a single set of minimum requirements, but towards a common framework of principles and standards that each country adapts in accordance with its legal and institutional context. The consolidation of this convergence, supported by cooperation initiatives between Member States and networks such as the European BIM groups, will be decisive in increasing regulatory predictability, reducing compliance costs and fully realising the benefits of the BIM methodology in public works across Europe. [25]

A. National legislation applicable to the BIM context

The existence of multiple national legislative initiatives for the adoption of the BIM methodology, such as Ministerial Order No. 255/2023 of 7 August, which regulates the mandatory content of construction drawings and the processes for preparing public works projects, repealing Ministerial Order No. 701-H/2008, it is evident that its methodological integration remains limited. However, a critical analysis of the legislation shows that, despite mentioning the term BIM for the first time, it lacks specific guidelines on technical requirements, multidisciplinary collaboration or interoperability, remaining tied to traditional practices and omitting fundamental content and concepts of the BIM ecosystem. [5]

Law No. 50/2023 grants the Government the power to reformulate and simplify planning permissions, establishing digitalisation as a central pillar of this process. Among its provisions, a key feature is the requirement to submit architectural and specialist designs modelled digitally and parametrically, coordinated using the Building Information Modelling (BIM) methodology, with pilot projects permitted in specific municipalities or cases. This provision was consolidated by Decree-Law No. 10/2024 of 8 January, which establishes the effective obligation, from 1 January 2030, for architectural designs to be modelled in accordance with BIM. The lack of clear criteria regarding the levels of detail (LOD) required in phases such as design, execution or maintenance reveals a weakness, given the absence of legal guidelines to standardise expectations among all stakeholders. This gap highlights the need for alignment with international models, in which the definition of progressive stages of BIM maturity has been instrumental in the standardisation and scalability of the methodology. [6] [7]

This legal framework presents a twofold challenge: the mandatory adoption of BIM in the national industry and the creation of an automated platform for validating BIM models.

B. Relevant standardisation in the use of the BIM methodology

The regulatory and legal framework has been a catalyst for change in BIM-related work processes. This framework enables the establishment of clear guidelines on data interaction and sharing amongst the parties involved, clarifies the legal requirements that facilitate the cohesive and consistent integration of BIM, and ensures interoperability between different platforms and disciplines.

CEN/TC 442 is the European technical committee responsible for the development and management of standards associated with BIM methodology. Its remit includes the development of a coherent set of standards, specifications and reports that establish methodologies for the definition, exchange, monitoring and secure management of data, semantics and processes related to construction assets throughout their lifecycle.

At national level, CT197 acts as the 'mirror committee' for CEN/TC442 and ISO/TC59, and is coordinated by the Sectoral Standardisation Body of the Instituto Superior Técnico (ONS/IST). Its mission focuses on the development of technical standards for classification systems, data standardisation, information modelling, parameterisation of BIM objects , and processes associated with the life cycle of construction projects, ensuring harmonisation with European and international standards.

The international progress of BIM is intrinsically linked to interoperability protocols, ensuring integration between digital tools and workflows. The BuildingSMART organisation stands out as a pioneer in promoting open standards, such as Industry Foundation Classes (IFC), the standardisation of which culminated in the EN ISO 16739 standard, establishing technical guidelines for platform-neutral data exchange. [17] [18]

The main standards relevant to the establishment of technical content and requirements associated with information sharing to be adopted using BIM methodology are identified and summarised below.

B.1) ISO 19650-1

ISO 19650-1 (Organisation and digitisation of information about buildings and civil engineering works, including building information modelling (BIM) Information management using building information modelling – Part 1: Concepts and principles) establishes the theoretical and methodological foundations for information management in construction projects, harmonising international practices and promoting the efficient digitisation of the AEC (Architecture, Engineering and Construction) sector, serving as a conceptual basis for the implementation of BIM. [8]

This standard defines a unified framework for the organisation, exchange and control of data between stakeholders (designers, contractors, operators), promoting interoperability through a Common Data Environment (CDE), ensuring traceability, version control and secure information sharing. [8]

Furthermore, it establishes key processes such as the definition of requirements, collaborative model production, data classification (aligned with ISO 12006-2) and information security management, throughout the entire life cycle of a built asset, from strategic planning to demolition or reuse.

B.2) ISO 19650-2

ISO 19650-2 (Organisation and digitisation of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling — Part 2: Delivery phase of the assets) sets out the requirements and processes for the collaborative management of data during the design and construction phases of projects, using the BIM methodology. This standard aims to ensure the efficient production, validation and delivery of technical information, aligning procedures with the operation and maintenance phases. [9]

It defines concepts such as the PIM, which is the digital model developed during the design phase and which integrates geometry, non-geometric attributes such as costs and deadlines, and technical documentation, being used

for conflict detection and planning; and the AIM, which corresponds to the operational model for asset management, containing essential data such as maintenance manuals or equipment specifications, as well as ensuring interoperability through open formats such as IFC, guaranteeing compatibility between different software and preventing data loss. [18]

ISO 19650-2 specifies practical processes for the asset handover phase, focusing on the production, validation and handover of information during project development to guide operational execution, reduce rework and ensure compliance with client requirements. [9]

B.3) ISO 16739-1

ISO 16739-1 (Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries – Part 1: Data schema) defines the open international standard, IFC, which is fundamental to ensuring interoperability in projects using the BIM methodology. [11]

This standard establishes a neutral, non-proprietary conceptual data schema, describing entities, attributes and relationships between objects that are essential for BIM projects, enabling the exchange of information between different software applications. This facilitates collaboration between multidisciplinary teams and data traceability through specific definitions for different workflows. [11]

ISO 16739 is widely adopted internationally and is a requirement in public procurement in several European countries, which reinforces its importance in the digitization of the construction sector. [21]

B.4) ISO 29481

ISO 29481 (Building information models – Information delivery manual) establishes methodologies for standardizing information exchange processes in projects using the BIM methodology. [12]

Part 1 of ISO 29481-1 defines the methodology and format for creating Information Delivery Manuals (IDM), including the requirements and data flows between the various stakeholders in the value chain, incorporating process maps and lists of mandatory data for each phase. [12]

Part 2 of ISO 29481-2 specifies an Interaction Framework that harmonizes communication between digital tools and the responsibilities of the various stakeholders, defining protocols for collaborative workflows and ensuring consistency in data exchange between heterogeneous software. [13]

The standard promotes interoperability with other key standards, such as ISO 19650 and ISO 16739, and includes automatic data validation protocols to ensure compliance with specific requirements.

B.5) ISO 12006

ISO 12006 is a standard that establishes guidelines for the organization and classification of information in the construction sector, focusing on the harmonization of categorization systems for construction elements, processes and resources. Part 2 of this standard, ISO 12006-2 (Building construction — Organization of information about construction works — Part 2: Framework for classification), is the most widely applied, defining a generic model for classification systems adaptable to local contexts, whilst maintaining international compatibility to facilitate the exchange of information in global projects. [10]

This part proposes a faceted and hierarchical structure, combining multiple tables that categorize elements by characteristics such as function, form, material or discipline. [10]

The system organises construction resources into components (e.g. beams, doors), agents (e.g. architects, engineers) and processes (e.g. planning, execution), covering the entire life cycle of the project, from design to demolition.

In Portugal, this standard serves as the basis for initiatives such as the National CICS (Construction Information Classification System), which aims to standardize data management in civil construction projects, including those using BIM methodologies.

B.6) ISO 7817-1

ISO 7817-1 (Building information modelling — Level of information need - Part 1: Concepts and principles) is a standard that defines the concepts and principles for specifying the level of detail of information, establishing guidelines to determine the quantity and type of data required at each stage of an asset's life cycle, ensuring that the information is sufficient, relevant and efficient, whilst avoiding redundancies or gaps. [14]

The standard introduces the concept of Level of Information Need (LOIN), which replaces terms such as Level of Detail (LOD) or Level of Information (LOI), consolidating them into an integrated approach. LOIN defines the scope and detail of geometric, alphanumeric and documentary information, tailored to the specific needs of each project phase (planning, design, construction, operation). [14]

This framework avoids ambiguity and enables the various stakeholders to define clear criteria, such as the intended use of the information (e.g. structural analysis, budgeting), the delivery date (life cycle stage) and the entities involved in the production and use of the data. [14]

B.7) ISO 23386 and ISO 23387

The standards ISO 23386 (Building information modelling and other digital processes used in construction – Methodology to describe, author and maintain properties in interconnected data dictionaries) and ISO 23387 (Building information modelling (BIM) – Data templates for construction objects used in the life cycle of built assets – Concepts and principles) are fundamental to the standardization of information in projects using BIM methodology, ensuring interoperability and consistency in digital workflows. [15]

ISO 23386 defines interconnected data dictionaries to eliminate linguistic ambiguities and establish a universal digital language for properties (e.g. 'thermal resistance', 'dimensions') and their attributes (units, definitions, and unique identifiers, GUIDs), ensuring interoperability between BIM software, avoiding semantic conflicts and loss of information. [15]

ISO 23387 specifies the structure for Product Data Templates (PDTs), which standardize the technical properties of materials, systems or building components. [16]

PDTs are developed taking into account various data resources, amongst other factors, such as IFC interoperability requirements, Construction Products Regulations (CPR), Environmental Product Declarations (EPD) and other standardization initiatives.

Through PropertySets (e.g. IfcMaterialProperties), data is linked to technical documents and certifications, ensuring consistency in the digital data provided by manufacturers and enabling automated model validation, thereby ensuring compliance with design requirements. [19]

IV. Challenges of BIM Standardization in the Construction Sector

The standardization of BIM methodology in Portugal faces several challenges that hinder its full and effective implementation. Firstly, there is a technical complexity associated with the absence of detailed specific national standards, which leads to inconsistencies in the practical application of the methodology. Despite initiatives such as SECClasS, which is the Portuguese classification system, fragmentation still persists between international standards, such as ISO 19650 and ISO 12006-2, and local adaptations, hindering the harmonization of processes.

Another major challenge is the lack of adequate training and skills development. The shortage of qualified BIM professionals, combined with resistance to change on the part of many companies, particularly small and medium-sized enterprises, limits the widespread adoption of this methodology. Many of these companies lack the financial and human resources to invest in training or digital tools, which exacerbates the situation.

The high costs associated with implementing BIM also represent a significant barrier. The necessary investment in software, hardware and training, as well as the need to purchase additional modules and carry out frequent updates, makes the transition to digital processes costly, particularly for small businesses.

Interoperability between different BIM tools, such as Revit and ArchiCAD, and integration with management platforms or standards such as ISO 16739 (IFC) remain problematic. These difficulties often result in data loss and rework, whilst the coexistence of distinct classification systems, such as the UK's Uniclass and Portugal's SECClasS, requires complex mappings, particularly in international projects. [17] [18]

In the legislative sphere, although Decree-Law No. 10/2024 mandates the use of BIM in architectural projects for planning permission, the legislation has not yet defined mandatory levels of detail for construction drawings, which creates ambiguities in its application, whilst maintaining a strong emphasis on traditional processes and the redefinition of the roles of the various parties involved in a project's life-cycle. [7] [23]

Furthermore, cultural and organizational resistance is a significant obstacle. The construction industry in Portugal is highly fragmented, and the lack of cooperation between the various stakeholders – architects, engineers, contractors – hinders the adoption of collaborative processes. Many companies continue to prefer conventional methods, considering them sufficiently effective.

Data management also presents challenges, notably the absence of clear protocols for the continuous updating of BIM models and the lack of feedback mechanisms to evaluate successful implementations, which limits process improvement. The integration of emerging technologies, such as the Internet of Things (IoT) and digital twins, also requires frequent revisions of current standards.

The management of large volumes of information, including data relating to safety, inspections and maintenance, requires the adoption of effective security measures to protect sensitive data, particularly in critical infrastructure projects.

Finally, Portugal relies heavily on European initiatives for BIM standardization, but the slow adaptation of European directives, such as ISO 19650, to the national context delays the consolidation of local standards.

In short, the standardization of BIM methodology in Portugal faces technical, cultural, financial and legislative challenges. Although Decree-Law No. 10/2024 represents a significant step forward, the lack of operational details and the under funding of training initiatives keep the sector in a phase of slow transition. Overcoming these barriers will require close collaboration between the Portuguese Institute of Quality, universities and the private sector, as well as specific government incentives to support small and medium-sized enterprises.

Results, Conclusions and Future Studies

This paper contributes to the current body of knowledge by presenting a structured bibliographic review of the main standards associated with the BIM methodology, clarifying their scope and relevance for the definition of technical content in project development and implementation processes. Rather than proposing new frameworks or solutions, the study systematizes existing normative references, including international standards such as ISO 19650, ISO 12006-2, and ISO 16739, alongside national initiatives such as SECClasS, thereby supporting a clearer understanding of their interrelations and applicability.

The analysis highlights that, within the Portuguese context, the primary challenges are not only technical but also organizational and economic, particularly regarding the capacity of the business fabric—largely composed of small and medium-sized enterprises—to adapt to these requirements. The paper identifies key barriers, including the fragmentation between standards, limited interoperability, insufficient training, and the financial burden associated with BIM adoption, without extending into prescriptive or solution-oriented approaches.

In this sense, the main contribution of this work lies in consolidating dispersed knowledge and critically identifying the difficulties associated with the implementation of BIM-related standards in the national context. This provides a foundation for future research and policy development, supporting more targeted strategies for standard harmonization, capacity building, and gradual integration of BIM practices within the construction sector.

REFERENCES

[1] Eastman, C., Sacks, R., Lee, G. & Teicholz, Paul. (2018). BIM Handbook: A Guide to Building Information Modelling for Owners, Designers, Engineers, Contractors, and Facility Managers (3rd ed.). New Jersey: Wiley.

[2] D. Holzer. (2015). The BIM Manager's Handbook. United Kingdom, Wiley.

[3] Lima, R., Couto, P., Falcão, M., Salvado, F., Resendo, R. & Parece, S. (2021). Analysis of concepts, standards and classification systems for construction information. Lisbon: SECCLasS.

[4] Regulation (EU) 2024/3110 of the European Parliament and of the Council of 27 November 2024.

[5] Ordinance No. 255/2023 of 7 August.

[6] Law No. 50/2023 of 28 August.

[7] Decree-Law No. 10/2024 of 8 January.

[8] ISO 19650-1: 2018, "Organization of information about construction works — Information management using building information modelling — Part 1: Concepts and principles", International Organization for Standardization (ISO), 2018.

[9] ISO 19650-2: 2018, "Organization of information about construction works — Information management using building information modelling — Part 2: Delivery phase of the assets", International Organization for Standardization (ISO), 2018.

[10] ISO 12006-2:2015, "Building construction — Organization of information about construction works — Part 2: Framework for classification", International Organization for Standardization (ISO), 2015.

[11] ISO 16739:2024, "Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries — Part 1: Data schema", International Organization for Standardization (ISO), 2024.

[12] ISO 29481-1:2016, "Building information models — Information delivery manual - Part 1: Methodology and format", International Organization for Standardization (ISO), 2016.

[13] ISO 29481-2: 2012, "Building information models — Information delivery manual - Part 2: Interaction framework", International Organization for Standardization (ISO), 2012.

[14] ISO 7817-1:2024, "Building information modelling — Level of information need — Part 1: Concepts and principles", International Organization for Standardization (ISO), 2024.

[15] ISO 23386:2020, "Building information modelling and other digital processes used in construction — Methodology to describe, author and maintain properties in interconnected data dictionaries", International Organization for Standardization (ISO), 2020.

[16] ISO 23387: 2020, "Building information modelling (BIM) — Data templates for construction objects used in the life cycle of built assets — Concepts and principles", International Organization for Standardization (ISO), 2020.

[17] "buildingSMART Data Dictionary (bSDD)". Retrieved from https://www.buildingsmart.org/users/services/buildingsmart-data-dictionary.

[18] "Industry Foundation Classes (IFC)". Retrieved from https://technical.buildingsmart.org/standards/ifc.

[19] "IfcElementQuantity". Retrieved from https://standards.buildingsmart.org/IFC/DEV/IFC4_2/FINAL/HTML.

[20] "Europe's Digital Progress Report 2017". Retrieved from https://digital-strategy.ec.europa.eu/pt/node/4161.

[21] "Report on the workshop held with the industry", Digital Transition Plan for the Construction Industry – Plan4Digital. Retrieved from https://revconstruction.pt/wp-content/uploads/2023/06/E4.4.1-Relatorio-dos-workshops-desenvolvidos-com-a-industria.pdf.

[22] "Report on BIM and the Digitization of Construction and Infrastructure". Retrieved from BIM and the Digitalisation of Construction and Infrastructure - COTEC Portugal.

[23] Succar, B. & Poirier, E. (2020). Lifecycle information transformation and exchange for delivering and managing digital and physical assets. Automation in Construction, Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0926580518309415?via%3Dihub.

[24] "Fostering the policies to incorporate BIM in Public Procurement across the EU". Retrieved from https://eubim.eu/wp-content/uploads/2025/12/Position_Paper_of_the_EU-BIM-Task-Group-on-BIM-for-Public-Procurement.pdf

[25] "Advancing BIM in Public Procurement: Key Insights from the BIM Community Survey". Retrieved from https://public-buyers-community.ec.europa.eu/communities/bim-and-public-procurement/resources/advancing-bim-public-procurement-key-insights-bim

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Using Design Methods for the Development of a New Framework for Root Cause Analysis

Matthew Barsalou [1], André Carvalho [2] and Rogério Puga-Leal [2]

1 NOVA School of Science and Technology (NOVA FCT), Universidade NOVA de Lisboa;

2 UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal

* Corresponding author: matthew.a.barsalou@gmail.com

Abstract— This paper describes the development, evaluation, and publication of a new framework for root cause analysis using design methods. The RCA framework is intended to provide manufacturing organizations with an approach to root cause analysis that is both easy to use, and includes all necessary actions for correctly and economically identifying the root cause of a failure.

The root cause analysis framework was developed according to the design science research methodology to ensure a reproducible and scientific approach to the development of the root cause analysis framework. In addition, design methods, consisting of exploration of the design space, generation of the root cause analysis framework, evaluation of the root cause analysis framework, communication of the root cause analysis framework, and communicating the root cause analysis framework to both academics and practitioners, was used to provide clear major steps in the development of the root cause analysis framework. *The root cause analysis framework can be applied in manufacturing organizations for investigating failures in manufactured products with limited to no training. The root cause analysis framework concentrates on finding root causes, and does not include other necessary actions, such as preventing the delivery of suspect parts while the investigation is in progress, or correcting the problem once the root cause is identified. The scope of the root cause analysis framework was limited to investigating the cause of failures in manufactured products. Additional studies would need to be performed to adapt the root cause analysis for investigating the failures in service industries.*

Keywords – Root Cause Analysis; Investigation; Problem Solving; Quality Tools

I. Introduction

Quality failures happen in manufacturing organizations (Hoerl et al. 2021) and can negatively impact the satisfaction of customers (Sreedharan et al. 2018), being detrimental to the competitiveness of the organization (Oliveira et al. 2022). When failures happen, a root cause analysis (RCA) is performed to identify the cause of failures (Gangidi 2019) and the ability to properly perform an RCA is an essential skill (Mohaghegh and Furlan 2020).

There are many approaches for performing an RCA. For example, there is the 8D report (Biben and Dhounchak 2017) and the A3 report, (Chakravorty 2019). There are also proprietary methods, such as Shainin methods (Panahi et al. 2021), which require training from the Shainin company and can't be fully described elsewhere because the methods used are proprietary (Tanco et al., 2008). However, a commonality amongst approaches to RCA is that there are many steps in the approach that are unrelated to actually finding the root cause. For example, only one of the eight steps in an 8D report is related to finding the root cause (Okes, 2019). There are also many quality tools available for various intended uses (Starzyńska et al., 2018). For example, Tague (2023) describes over one hundred (100) quality tools. The sheer number of quality tools available can be a complication when those performing an RCA don't know which quality tool to use, or don't know how to use all of the quality tools. A new framework for RCA is needed to address the limitations of the current approaches to RCA.

This paper describes the development of a new, simple framework for RCA using design science research as a scientific methodology, and design methods for the development of the RCA framework. There were four major steps in the development of the RCA framework consisting of exploration of tools and approaches for RCA, generation of the RCA framework, evaluation of the RCA framework, and communicating the RCA framework to both academic researchers and practitioners in industry. Combining design science research and design methods provided a scientifically developed framework that was developed using an established approach to developing artifacts, such as frameworks.

II. Literature Review

Much of the literature on RCA is in the form of case studies, such as investigating a door panel alignment problem (Uluskan and Oda 2020) or a product generating an unwanted noise (Suarez-Barraza et al. 2021). Literature also describes both approaches to RCA and quality tools for RCA (Antony et. al., 2021; McDermott et al., 2023a). For example, although intended for improving process performance by reducing variation by using phases consisting of Define, Measure, Analyze, and Control (DMAIC), Six Sigma (Barsalou and Perkin, 2015) can be used for RCA. A common

approach to RCA is the 8D report, which has eight steps starting with forming a team and ending with congratulating the team (Borosani et al., 2017). The A3 report (Lenort et al., 2017), which has steps such as identifying the background, RCA, and target condition, can also be used for RCA. Approaches to RCA often include steps defining the problem, investigating to find the cause of the problem, identifying corrective actions, and implementing corrective actions to prevent a reoccurrence of the problem (Mohaghegh and Furlan, 2020).

There are also proprietary approaches to RCA, such as Apollo Root Cause (Mulenga et al., 2018) and TapRoot (Paradies, 2025). A well-known proprietary approach to RCA is Shainin Methods (Kumar et al.. 2022). There are disadvantages to proprietary methods. For example, costly training may only be available from the owners of the methods and the methods are not evaluated in research journals (Tanco et al., 2008), with the available literature consisting of case studies (Kumar et al., 2022).

Quality tools are often used during an RCA and a commonly used quality tool is the Ishikawa diagram (Suárez-Barraza and Rodríguez-Gonzále 2019). Hypotheses (VDA. 2019) explain the failure under investigation are listed in the Ishikawa diagram and investigated. Another quality tool used during an RCA is five whys, which consists of asking why five times (Berstene, 2018); however, five whys is often used without investigation.

There are weaknesses to approaches to RCA. For example, there may only be recommended tools in place of a process (De Feo and Caldwell, 2017; Okes, 2019), no hypotheses prioritization (Furterer and Wood, 2021), many steps unrelated to finding the root cause (Berstene, 2018), or the approach may be proprietary (Panahi et al. (2021)

III. Methodology

The methodology used for the development of the RCA framework was design science research. Design science research was selected because design science research is used for ensuring scientific rigor in situations in which requirements and where human creativity is needed (Cruz and da Cruz, 2020) for the development of artifacts (Romme and Holmström, 2023) that solve problems in the real world (Chou et al, 2024). The artifacts can consist of models, methods, and frameworks, as well as physical objects (Weigand and Johannesson, 2023).

A. Design Science Research

There are multiple approaches that can be used in design science research (Agung Premananda et al., 2022). For example, Anotney et al. (2024) list the steps as identify the problem and mediatization, define the objective of a solution, design and development, demonstration, evaluation, and communication, while Bozkurt et al. (2025) describe the steps as understanding of the problem space, design and development of an artifact, demonstration and evaluation, and communication. Design science methods do not emphasize the creation of the artifact (Daalhuizen and Cash, 2021); therefore, the steps of design methods were used for developing the RCA framework.

B. Design Methods

Design methods is an approach that provides methods that can be both repeated and measured and is used across fields (Löwgren and Stolterman, 1999). Design methods the designers of artifacts with guidance for developing the artifact (Dickson and Stolterman, 2016). However, the guidance is not prescriptive and there is flexibility in design rules Eisenmann, et al. (2021), which is necessary due to designers of artifacts needing to adapt to changing situations (Daalhuizen et al., 2019). Design methods consist of four steps, which are exploration, generation, evaluation, and communication. Table 1 displays the alignment of design science research, design methods, and the major actions taken to develop the RCA framework.

TABLE 1. COMPARISON OF DESIGN SCIENCE RESEARCH, DESIGN METHODS, AND DEVELOPMENT OF THE RCA FRAMEWORK (SOURCE: BARSALOU, 2026).

Design science research	Design methods	Development of RCA Framework
Identification of real world problem; Identify objective/Understand the problem space	Exploration	Literature explored and surveys conducted
Design and development	Generation	Framework developed based on literature and surveys
Demonstration/Evaluation	Evaluation	Analysis of a database, experiments, surveys, and a case study
Communication	Communication	Publication of the framework in a practitioner journal, online quality magazine with link to framework documents, and a webinars

1) Exploration

The first step in design methods is exploration, which is where the problem is explored to understand the problem that is being solved and to determine what the objectives should be (Cross, 2018). Actions taken during exploration include performing a literature review (Shminan et al., 2022) and conducting surveys with questionnaires (Agung Premananda et al., 2022).

For exploration, a literature review was conducted to understand the state of the art in RCA. This included papers on approaches to RCA and quality tools for RCA. Once sufficient information was gained, studies were conducted. Studies were also conducted to identify quality tools to include in the RCA framework. The first study (Barsalou and Perkin, 2024) used an international manufacturing organization's quality leader email distribution list to send out a survey that asked respondents which quality tools they used during an RCA. The second study (Barsalou, 2023a) identified the classic seven quality tools from the literature and then used the same email distribution list to ask respondents to select

from a list the classic seven quality tools they used during an RCA. Preliminary criteria to use for the prioritization of cost and effort was identified by surveying quality professionals in organizations on a list of ISO 9001 certified organizations in a specific region (Barsalou, et al. 2024a). Respondents were asked to list what they would consider to be extensive, moderate, and low for both costs in monetary units and hours of labour. Criteria for rating the strength of evidence during an RCA was determined by conducting a study. Types of evidence (Barsalou, 2023a) during an RCA were identified from the literature and the previously described quality leader email distribution list was used to ask respondents to rate the strength of the types of evidence as high, medium, and low. The use of five whys is described in the literature as a brainstorming tool (Murugaiah et al. 2010) and as a method that requires investigation (de Mast, 2013). Therefore, a study was conducted (Barsalou and Starzyńska, 2023a) to understand how five whys is applied in the literature and the study found that five whys is not consistently used with investigation.

2) Generation

The next step in design methods is generation, where an artifact is developed to solve the problem that is being addressed (Lavrsen et al., 2022). When using design methods, there must be justifications must be given for the design choices made (Delport et al., 2024). Designing is an iterative process, and the first design created may be modified based on the results of the evaluation of the design (Chou et al. 2024).

During the generation step, a process for the RCA framework was established, starting with a team list for listing those involved with the investigation. An is/is-not is then used to capture details that are transferred to a problem description. If preliminary data is available, it is then graphed to potentially gain insights. An Ishikawa diagram is created for listing hypothesis and the hypotheses are transferred to an action item tracker and prioritized in consideration of cost in hours of labor and time to investigate and strength of evidence. Top hypotheses are then investigated and transferred to an evidence matrix to compare the hypotheses to observations. If the hypothesis is supported, it is transferred to a dig deeper worksheet to ensure proper use of five whys to identify the underlying cause of the failure. If the root cause is not identified, the next hypothesis is then investigated and the process repeats.

3) Evaluation

The third step is using design methods is the evaluation of the artifact under real-world conditions while ensuring scientific rigor Lavrsen et al., 2022). The evaluation can consist of using the artifact to address simulated problems (Agung Premananda et al. 2022). The evaluation can also consist of application of the artifact in a case study to evaluate the artifact (Antony et al., 2024), and the multiple evaluations per performed to evaluate the RCA framework.

The first assessment was of the use of an is/is-not, which was assessed in two studies. The first study (Barsalou and Perkin, 2022) assessed the time to find a root cause when using a problem description based on the categories in an is/is-not, versus time to find a root cause without a problem description based on the categories in an is/is-not. The study used over 200 entries in a problem solving database and found that better defined problems were solved in less time than less well-define problems. The effectiveness of an is/is-not for generating better defined problem descriptions was assessed (Barsalou and Perkin, 2022) by asking untrained students and people in industry trained in writing problem descriptions to write problem descriptions with and without an is/is-not analysis. The study found that use of an is/is-not resulted in better detailed problem descriptions when used by people without training in wiring problem descriptions.

The effectiveness of EDA was then explored (Barsalou et al., 2024b) by conducting a study with university students with some statistical knowledge. The students were randomly assigned to either a group with data in a run chart, or a group with data in a table. Both groups were tasked with finding salient features and they were timed. The group with data in a graph identified salient features in less time than those with data in a table.

A different type of study was performed to evaluate the effectiveness of a matrix listing hypotheses and observations in a matrix in identifying root causes (Barsalou et al., 2023b) in a study that randomly assigned engineers a matrix with observations and evidence, or a list of hypotheses and observations. The study found that using a matrix led to finding the root cause more often than not using a matrix.

The use of an Ishikawa diagram action item tracker verses not using an Ishikawa diagram action item tracker to develop testable hypotheses was also evaluated (Barsalou et al., 2023b). Engineers were presented with a technical drawing and given a failure scenario. The engineers were tasked with identifying explanatory hypotheses and randomly assigned to a group with an action item tracker, or a group without an action item tracker. The group with the action item tracker identified more testable hypotheses.

The criteria for cost and effort prioritization were evaluated by presenting quality professionals with the criteria and asking if they strongly agreed, agreed, disagreed, or strongly disagreed. Strongly agreed and agreed were combined and strongly disagreed and disagreed were combined and the combined agree and disagree were assessed statistically. The quality professionals agreed with the criteria, but only if cost or effort values could be changed for local conditions.

Next, the effectiveness of performing an RCA with a process-based approach was assessed (Barsalou et al., 2023a) by using five years' worth of data from a problem- solving database. The problems in the first two years were solved without a structured approach to problem solving and the remaining three years used a structured approach to problem solving. A statistical analysis found a difference in the time to solve problems, with problems that were solved with a structured approach being solved in less time than problems solved without a structured approach.

The final evaluation of the RCA framework consisted of a case study. The RCA framework was applied to a real-world problem in a manufacturing organization and actions taken were documented for analysis purposes. The case study demonstrated the effectiveness of the RCA framework, and user feedback led to optimizations such as including

a place to track non-investigated related actions in the Ishikawa diagram action item tracker and formatting changes in the Ishikawa diagram to better fit text in the Ishikawa diagram.

4) Communication

The last step in design methods is communication, which is where the artifact is communicated (Delport et al., 2024). When communicating the artifact, the target audience must be considered. Communicating the artifact only in academic journals risks failing to reach practitioners who would use the artifact, so publication can take place in both academic and practitioner journals (Gray, 2022).

There were 12 studies performed for the creation and evaluation of the RCA framework and the studies have resulted in 10 publications in academic journals (Barsalou and Starzyńska, 2023a; Barsalou and Starzyńska, 2023b; Barsalou and Perkin, 2022; Barsalou and Perkin, 2024; Barsalou et al., 2023a; Barsalou et al., 2023b; Barsalou et al., 2024a; Barsalou et al., 2024b; Barsalou, 2023a; Barsalou, 2024b). There is also one additional academic journal publication planed, which would bring the total to 11 academic journal publications.

There is a disadvantage in publishing only in academic journals because academic journals publish research for a target audience consisting of researchers. Therefore, a practitioner journal publication (Barsalou, 2025) has also been published to reach quality practitioners who may benefit from applying the RCA framework when confronted with quality failures. To further reach quality practitioners, three practical papers will be submitted to a non-peer revised online publication.

In addition to publications, a webinar has been conducted to communicate the RCA framework to quality practitioners. An additional webinar is intended, and this one should be recorded with a link posted on the internet to make the RCA framework readably available.

IV. Results and Discussion

The RCA framework was developed using design science research, together with the steps based on design methods. During the exploration phase, literature was reviewed and studies were conducted. Based on the literature review and studies, the RCA framework was developed during the generation phase. Evaluation is where parts of the RCA framework were evaluated individually, and the complete RCA framework was evaluated through the use of a case study. Finally, the development of the RCA framework was communicated to an academic audience and the RCA framework itself was communicated to a practitioner audience. The RCA framework is shown in Fig. 1.

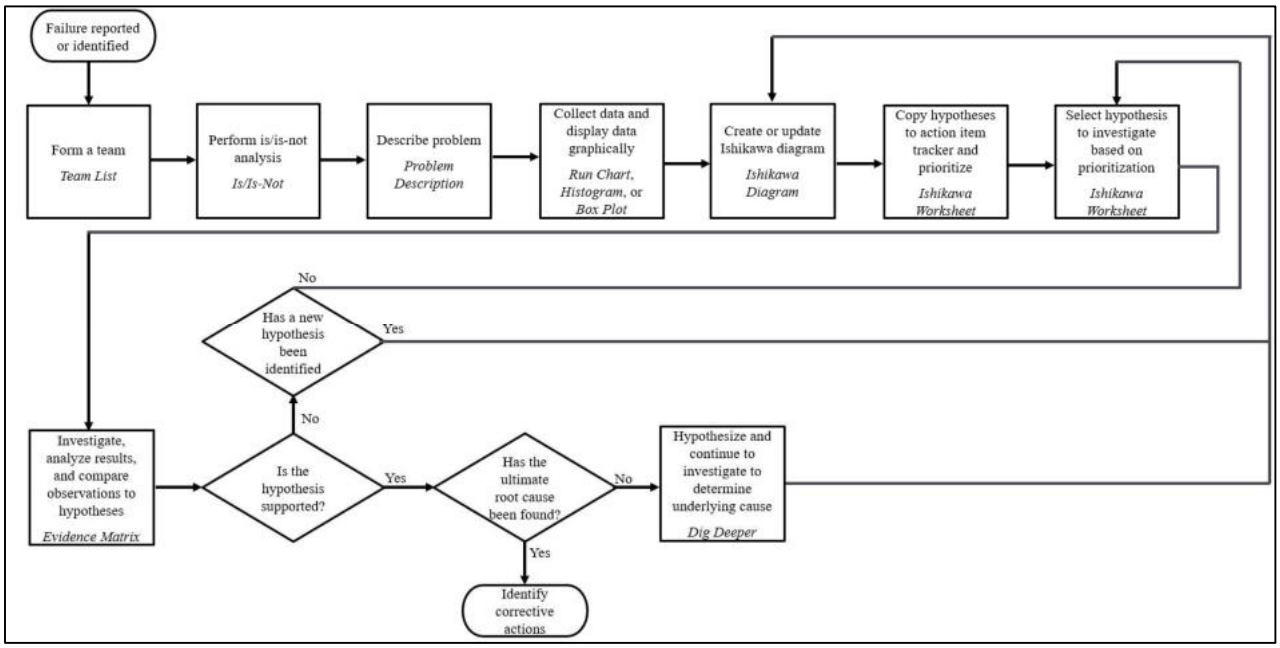


Fig. 1. Framework for RCA (Source: Barsalou, 2025)

The RCA framework is based on templates in spreadsheets. The first template is for listing the team, and the second one is for the is/is-not. The is/is-not is intended to serve as a check sheet to ensure key details are captured for writing the problem statement, with missing information being a visual indicator that more information is needed to achieve a complete problem description. Next, templates are available for graphing data, with the specific graphs based on the type of data used. An Ishikawa diagram is then created to list explanatory hypotheses. These hypotheses are transferred to an action item tracker and prioritized using prioritization criteria that consider combinations of cost and effort to investigate, and the strength of the evidence supporting a hypothesis. The results of the investigation are entered into a matrix listing evidence together with observations that support the hypotheses. Identified causes are

then listed in a dig deeper table that requires explaining how the cause was identified to ensure an investigation is performed and not just brainstorming.

The development of the RCA framework was an extensive endeavour. Using design science research ensured a scientific approach was used to develop the RCA framework and use of the steps in design science research broke the development process down into more manageable steps with clear objectives for each step. A problem was suspected; current approaches to RCA had weaknesses, such as a lack of empirical investigation or the use of a proprietary approach.

Using design science research, the problem space was explored to better understand the problem. The RCA framework was then developed based on an exploration of relevant literature and the use of surveys. Next, individual parts of the RCA framework were evaluated using a combination of experiments and surveys. The suitability of the RCA framework as a whole was assessed using a case study approach, with the RCA framework applied for solving a quality problem in a manufacturing organization. Developing an RCA framework in a haphazard manner would have missed out on opportunities for modification based on evaluation results. For example, the dig deeper worksheet contained redundant requirements that could have led to user confusion, and this weakness was eliminated after it was identified during the case study. The final step was communication and multiple methods were identified to communicate the RCA framework to potential users, such as publication in a practitioner journal and planned publication in an online quality magazine, presentation at a conference, and a webinar.

Without design science research, the RCA framework would have been created without a scientific approach, potentially resulting in a sub-optimal solution or a completely inadequate solution. Instead, the RCA framework provides a scientifically developed and evaluated RCA framework for investigating failures in manufacturing organizations. The RCA framework can be applied with no or limited training and can be integrated into other problem-solving processes.

The RCA framework does not include containment of failed parts. This was a deliberate decision choice to concentrate on finding the root cause. Actions to take for containment of failed parts are well-established and already imbedded in methods such as 8D reports and A3 reports. However, the finding of causes is just one part of 8D reports and A3 reports and while both approaches may specify methods to use, concrete steps to take are often lacking. The root cause analysis phase of 8D reports and A3 reports often consists of a list of tools to use, without descriptions of how to perform a proper investigation. Table 2 contrasts the RCA framework with 8D reports and A3 reports.

TABLE 2. COMPARISON OF THE RCA FARMEWORK, 8D REPRTS, AND A3 REPORTS

RCA framework	**8D reports**	**A3 reports**
Form team	Team	Team
Perform is/is not analysis	Problem Description	Background
Describe problem		Current condition
No counterpart	Immediate Containment Action	
Collect and graph data	Root Cause	Root cause analysis
Create Ishikawa diagram		
Move hypotheses to track and prioritize		
Investigate and compare observations to hypotheses		
No counterpart	Planned improvement actions	Target condition
No counterpart	Implemented improvement actions	Implementaiton plan
No counterpart	Lessons learned	Fallow up
No counterpart	Congratulate the team	No counterpart

V. CONCLUSION

This paper describes the development of a new framework for root cause analysis. The objective was to provide manufacturing organizations with a simple and easy to use approach to RCA that concentrates on finding the root cause of failure's both quickly and economically. However, there are limitations to the RCA framework.

The RCA framework is intended for identifying the cause of failures in manufactured products in manufacturing organizations. In its current version, the RCA framework is not intended for investigating failures in the delivery of services. This is a weakness of the RCA framework, and an opportunity for further research. Especially important would be identifying the correct quality tools for investigating the failure of services, and adapting the prioritization matrix to cover costs in hours of labor, effort to investigate, and types of evidence for service failures. Another opportunity for future research would be determining the necessary modifications of the RCA framework for investigating project failures.

REFERENCES

Agung Premananda, I- G., A. Tjahyanto, and A. Mukhlason. (2022), "Design Science Research Methodology and Its Application to Developing a New Timetabling Algorithm," *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom),* Malang, Indonesia, 2022. 433-438. DOI: 10.1109/CyberneticsCom55287.2022.9865661.

Antony, J., O. McDermott, M. Sony, M. M. Fernandes, and R. V. C. Ribeiro. (2021), "A Study on the Ishikawa's Original Basic Tools of Quality Control in South American Companies: Results from a Pilot Survey and Directions for Further Research." *The TQM Journal*. 33(8): 1770-1786. DOI: 10.1108/TQM-01-2021-0004.

Antony, J., M. Sony, B. Lameijer, S. Bhat, R. Jayaraman, and L. Gutierrez. (2024), "Towards a Design Science Research (DSR) Methodology for Operational Excellence (OPEX) Initiatives," *The TQM Journal*. 36 (8): 2383–2397. DOI: 10.1108/TQM-01-2023-0017.

Barosani, S., N. Bhalwankar, V. Deshmukh, S. Kokane, and P. R. Kulkarni. (2017), "A Review on 8D Problem Solving Process." *International Research Journal of Engineering and Technology*. 4 (4): 529-535.

Barsalou, M. and R. Perkin. 2015, "Structured, Yet Flexible, Approach." *Six Sigma Forum Magazine*. 14 (4): 22-25

Barsalou, M. and R. Perkin. (2022), "An Empirical Assessment of Problem Statement Creation with Is/Is-Not," *Proceedings on Engineering Sciences*. 4(3): 407-416. DOI: 10.24874/PES04.04.003.

Barsalou, M. (2023a), "Determining Which of the Classic Seven Quality Tools are in the Quality Practitioner's RCA Tool Kit," *Cogent Engineering*. 10(1). DOI: 10.1080/23311916.2023.2199516.

Barsalou, M. (2023b). "Criteria for the Prioritization of Hypotheses in Root Cause Analysis," Quality and Reliability Engineering International. 39(1): 132-142. DOI: 10.1002/qre.3224.

Barsalou, M. and B. Starzyńska. (2023a). "Five Whys: A Possible Path to Failure," *International Journal of Six Sigma and Competitive Advantage*. 15(1): 19-32. DOI: 10.1504/IJSSCA.2024.139599.

Barsalou, M. and B. Starzyńska. (2023b), "A New Method for Formulating a Strong Hypothesis in RCA," *International Journal for Quality Research*. 17(1): 77-96. DOI: 10.24874/IJQR17.01-06.

Barsalou, M., M. Grabowska, and R. Perkin. (2023a), "Inquiry into the Effectiveness of 8D-Based Problem-Solving," *Quality Innovation Prosperity*. 27(2): 61-76. DOI: 10.24874/IJQR17.01-06.

Barsalou, M., P. Saraiva, and R. Henriques. (2023), "Exploring Exploratory Data Analysis: An Empirical Test of Run Chart Utility," *Management Systems in Production Engineering*. 31(4): 442-448. DOI: 10.2478/mspe-2023-0050.

Barsalou, M. and R. Perkin. (2024), "Inquiry into the use of Quality Tools for Root Cause Analysis," *Proceedings on Engineering Sciences*. 6(1): 127-138. DOI: 10.24874/PES06.01.015.

Barsalou, M., B. Starzyńska, M. Konrad, A. Carvalho, and R. Puga-Leal. (2024a), "Identification and Assessment of Cost and Effort Criteria for Prioritization During Root Cause Analysis," *Quality Innovation Prosperity*. Vol. 28 No. 2, pp. 60-75. DOI: 10.12776/qip.v28i2.1988.

Barsalou, M., A. Carvalho, R. Puga-Leal. (2024b), "Investigation into the use of Is/Is-Not for Writing Problem Descriptions for Root Cause Analysis," *Total Quality Management and Business Excellence*. Vol. 35 No. 15-16, pp. 1833-1846. DOI: 10.1080/14783363.2024.2406873.

Barsalou, M. (2025), "Rooting Out Failure Quality Progress. Vol. 58 No. 12, pp. 62-69.

Barsalou, M. (2026), Framework for Root Cause Analysis. [Doctoral dissertation, Nova University Lisbon]. Ahead-of-print.

Berstene, T. G. (2018), *The Root Cause Analysis Memory Jogger: A Process Based Approach to Root Cause Analysis*. Methuen, MA: GOAL/QOC.

Biban, L. K. and D. Dhounchak. (2017), "Relevance of 8D Methodology," *Journal of Emerging Technologies and Innovative Research*. Vol. 4 No. 10, pp. 415-149.

Bozkurt, Y., A. Rossmann, Z. Pervez, and N. Ramzan. (2025). "Development and Evaluation of an Urban Data Governance Reference Model Based on Design Science Research," *Government Information Quarterly*. Vol. 42, No. 2. DOI: 10.1016/j.giq.2025.102025.

Chakravorty, S. S. (2019). "Process Improvement: Using Toyota's A3 Reports," *Quality Management Journal*. Vol. 16 No. 4, pp. 7-26. DOI: 10.1080/10686967.2009.11918247.

Chou, T.-H., Y.-S. Chen , C.-W. Pan, and H.-H. Chang. (2024), "A Smart Scale for Efficient Inventory Management Based on Design Science Research Principles," *Contemporary Economics*. Vol. 18 No. 2, pp. 153-170. DOI: 10.5709/ce.1897-9254.531%0A.

Cross, N. (2018), "Developing Design as a Discipline," *Journal of Engineering Design*. Vol. 29 No. 12, pp. 691–708. DOI: 10.1080/09544828.2018.1537481.

Cruz, E. F. and A. M. R. da Cruz (2020), "Design Science Research for IS/IT Projects: Focus on Digital Transformation," *2020 15th Iberian Conference on Information Systems and Technologies (CISTI) 24 – 27 June 2020, Seville, Spain*. DOI: 10.23919/CISTI49556.2020.9140972.

Daalhuizen, J. and P. Cash. (2021), "Method Content Theory: Towards a new Understanding of Methods in Design," *Design Studies*. No. 75, pp.1-34. DOI: 10.1016/j.destud.2021.101018.

De Feo, J. and K Caldwell. (2017), *Root Cause Analysis Methods in Juran's Quality Control Handbook: The Complete Guide to Performance Excellence* (7th ed.). De Feo, J. A. (ed.). New York: McGraw-Hill.

Delport, P. M. J., R. V. Solms, and M. Gerber. (2024), "Methodological Guidelines for Design Science Research," *Procedia Computer Science*. No. 237, pp. 195-203. DOI: https://doi.org/10.1016/j.procs.2024.05.096.

de Mast, J. (2013), "Diagnostic Quality Problem Solving: A Conceptual Framework and Six Strategies." *Quality Management Journal*. Vol. 20 No. 4, pp. 21-36. DOI: 10.1080/10686967.2013.11918362.

Dickson, G. and E. Stolterman. (2016), "Why Design Method Development is not always carried out as User-Centered Design," in *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems, ACM, 2016*. DOI: 10.1145/2858036.2858324.

Furterer, S. L. and D. C. Wood. (ed.). (2021), *The Certified Manager of Quality/Organizational Excellence Handbook*. (5th ed.). Milwaukee, WI: ASQ Quality Press.

Gangidi, P. (2019), "A Systematic Approach to Root Cause Analysis Using 3 × 5 Why's Technique," *International Journal of Lean Six Sigma*. Vol. 10 No. 1, pp. 295-310. DOI: 10.1108/IJLSS-10-2017-0114.

Gray, C. M. (2022), "Languaging Design Methods," *Design Studies*. No. 78. Article No. 101076. DOI: 1016/j.destud.2021.101076.

Hoerl, R., W. Jensen, and J. de Mast. (2021), "Understanding and Addressing Complexity in Problem Solving," *Quality Engineering*. Vol. 33 No. 4, pp. 612-626. DOI: 10.1080/08982112.2021.1952230.

Kumar, A., K. Singh, K. Singh, K. Thakur, H. Singh, and J. Singh. (2022), "Product Performance Optimization in an Automotive Manufacturing Plant Using Shainin DOE Tools: A Case Study," *Materials Today: Proceedings*. Vol. 50 No. 5, pp. 1679-1682. DOI: 10.1016/j.matpr.2021.09.150.

Lavrsen, J. C., J. Daalhuizen, S. Dømler, and K. Fisker. (2022), "Towards a Lifecycle of Design Methods," in D. Lockton, S. Lenzi, P. Hekkert, A. Oak, J. Sádaba, and P. Lloyd. (eds.), *DRS2022: Bilbao, 25 June - 3 July, Bilbao, Spain*. DOI: 10.21606/drs.2022.542.

Löwgren J. and E. Stolterman. (1999). "Methodology and Tools: Design Methodology and Design Practice," *Interactions*. Vol. 6 No. 1, pp. 13-20. DOI: 10.1145/291224.291233.

McDermott, O., J. Antony, M. Sony, A. Rosa, M. Hickey, and T. A. Grant. (2023), "A Study on Ishikawa's Original Basic Tools of Quality ´Control in Healthcare," *The TQM Journal*. Vol. 35 No. 7, pp. 1686-1705. DOI: 10.1108/TQM-06-2022-0187 .

Mohaghegh, M. and A. Furlan. (2020), "Systematic Problem-Solving and its Antecedents: A Synthesis of the Literature." *Management Research Review*. Vol. 43 No. 9, pp. 1033-1062. DOI: 10.1108/MRR-06-2019-0284

Mulenga, K., X. Zhao, M. Xie, and C. Chikamba. (2018), "Investigating the Root Causes of Major Failures of Critical Components – With a Case Study of Asbestos Cement Pipes," *Engineering Failure Analysis*. No. 84, pp. 121-130. 10.30572/2018/KJE/160219.

Murugaiah, U., S. J. Benjamin, M. S. Marathamuthu, and S. Muthaiyah. (2010), "Scrap Loss Reduction Using the 5-Whys Analysis," *International Journal of Quality and Reliability Management*. Vol. 27 No. 5, pp. 527-540. DOI: 10.1108/02656711011043517.

Okes, D. (2019), *Root Cause Analysis: The Core of Problem Solving and Corrective Action.* (2nd ed.) Milwaukee, WI: Quality Press.

Oliveira, E. e, V. L. Miguéis, and J.L. Borges. (2022), "On the Influence of Overlap in Automatic Root Cause Analysis in Manufacturing." *International Journal of Production Research*. Vol. 60 No. 21, pp. 6491-6507. DOI: 10.1080/00207543.2021.1992680.

Paradies, M. (2025), "2024 Global TapRoot® Summit Keynote Speakers." www.taproot.com. Retrieved 29 November 2025 from https://www.taproot.com/2024-summit-keynote-speakers.

Panahi, M., J. De Mast, and S. H. Steiner. (2021), "Identifying Dominant Causes using Leveraged Study Designs," *Quality Engineering*. Vol. 33 No. 4, pp. 581-593. DOI: 10.1080/08982112.2021.1938118.

Romme, A. G. L. and J. Holmström. (2023), "From Theories to Tools: Calling for Research on Technological Innovation Informed by Design Science," *Technovation*. No. 121. DOI: /10.1016/j.technovation.2023.102692.

Shminan, A. S., S. N. Ain Romly, M. Aren, L. J. Choi, and W. N. Wan Hashim. (2022), "Applying Design Science Research Methodology for Development of a Mobile-Based Digital Quail Farming Guide," *2022 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), Karachi, Pakistan, 2022*. pp. 1-7. DOI: 10.1109/MAJICC56935.2022.9994157.

Sreedharan, R. V., S. Rajasekar, S. S. Kannan, P. Arunprasad, and R. Trehan. (2018), "Defect Reduction in an Electrical Parts Manufacturer: A Case Study," *The TQM Journal*. Vol. 30 No. 6, pp. 650-678. DOI: 10.1108/TQM-03-2018-0031.

Starzyńska, B., Górski, F., Buń, P. (2018), "Virtual Reality System for Learning and Selection of Quality Management Tools," *Advances in Intelligent Systems and Computing*. No. 747: 264-274. DOI: 10.1007/978-3-319-77700-9.

Suárez-Barraza, M.F. and Rodríguez-González, F.G. (2019), "Cornerstone Root Causes Through the Analysis of the Ishikawa Diagram, is it Possible to Find Them? A First Research Approach," *International Journal of Quality and Service Sciences*. Vol. 11 No. 2, pp. 302-316. DOI: 10.1108/IJQSS-12-2017-0113.

Suarez-Barraza, M. F., J. A.Miguel-Davila, and M. F. Morales-Contreras. (2021), "Application of Kaizen-Kata Methodology to Improve Operational Problem Processes. A Case Study in a Service Organization," *International Journal of Quality and Service Sciences*. Vol. 13 No. 1, pp. 29-44. DOI: 10.1108/IJQSS-07-2020-0113.

Tague, N. R. (2023), *The Quality Toolbox*. (3rd ed.), The ASQ Quality Press, Milwaukee, WI.

Tanco, M., E. Viles,and L. Pozueta. (2008), "Are All Designs of Experiments Approaches Suitable for Your Company?" Proceedings of the World Congress on Engineering 2008. Vol II WCE 2008. London, U.K.

Uluskan, M. and E. P. Oda. (2020), "A Thorough Six Sigma DMAIC Application for Household Appliance Manufacturing systems: Oven Door-Panel Alignment Case," *The TQM Journal*. 32(6): 1683-1714. DOI: 10.1108/TQM-06-2019-0171.

Verband der Automobilindustrie e. V. (VDA). (2018), *Quality Management in the Automotive Industry: 8D – Problem Solving in 8 Disciplines: Method, Process, Report*. Berlin, Germany: Verband der Automobilindustrie e. V. (VDA) / Qualität Management Center (QMC).

Weigand, H. and P. Johannesson. (2023), "How to Identify your Design Science Research Artefact," *2023 IEEE 25th Conference on Business Informatics (CBI), Prague, Czech Republic, 2023*. pp. 1-10. DOI: 10.1109/CBI58679.2023.10187511.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Effectiveness of Training Programmes for Software Engineers in Digital Transformation Contexts

Georgiana-Alexandra Voicu[1] , Irina Severin[2]

[1] National University of Science and Technology POLITEHNICA Bucharest, Romania; georgiana.voicu2108@stud.fiir.upb.ro
[2] National University of Science and Technology POLITEHNICA Bucharest, Romania; irina.severin@upb.ro

***Abstract*—The digitalisation of processes in the automotive industry has become a strategic priority, increasing the need for effective quality related training that helps engineers apply quality tools, standards, and methods in software development environments. This study investigates the extent to which training provided by quality engineers supports organisational digitalisation needs using an assessment framework based on the European Foundation for Quality Management (EFQM) Model 2025. The analysis focuses on a self assessment of EFQM Criterion 5: Driving Performance and Transformation, which examines how organisations manage performance, innovation, technology adoption, and organisational transformation. The assessment is supported by aggregated input from key stakeholder groups, a radar chart used to visualise the assessment results, and an Ishikawa root cause analysis to explore underlying factors influencing training effectiveness. The results indicate that although relatively high maturity levels are achieved in areas such as innovation, technology adoption, organisational transformation, and resource management, lower maturity is observed in performance management and data driven evaluation. These findings reveal a gap between technological capabilities and the ability to translate training activities into measurable digitalisation outcomes. The Ishikawa analysis shows that this gap is mainly related to organisational methods, measurement practices, people related factors, and management decisions rather than to limitations in technology or infrastructure. Overall, the study concludes that training provided by quality engineers does not fully support digitalisation needs, highlighting the importance of aligning training design, delivery, and evaluation with digital transformation objectives.**

Keywords – training, automotive, EFQM, digitalisation

I. Introduction

The digitalisation of processes in the automotive industry has become a strategic necessity in response to rapidly evolving technologies and increasingly complex market demands. Modern vehicles are no longer purely mechanical systems; they integrate advanced software architectures, embedded systems, data driven functionalities, and interconnected digital platforms. As a result, the nature of engineering work has fundamentally changed. Development engineers are now required to operate within agile environments, interact with complex software tools, and make decisions based on real time data and performance indicators. This transformation places significant pressure on organisations to continuously adapt their internal capabilities, including the way they design and deliver training.

In the digital age, quality is no longer limited to traditional business processes but increasingly extends to everyday organisational practices and technology driven environments (Krivokuca, 2019). In the context of Industry 4.0, the training of engineers must evolve to reflect the technologies shaping modern production and development environments (Garad & Rozana, 2024), (Dias, et al., 2023). One relevant example is the digital twin – a virtual representation of a physical system that enables real time monitoring, simulation, and optimisation – whose integration into training programmes can improve engineers' understanding of defect detection, process control, and data driven decision making (Mendi, 2022). At the same time, the continuous development of engineers is essential for maintaining high standards of performance and safety in rapidly changing industrial contexts (Tafakur*, et al., 2023), and organisations investing in human capital and learning environments tend to achieve stronger performance outcomes (Bhargavi., 2024).

In this context, the role of quality management is also evolving. Quality engineers are no longer responsible only for compliance verification and process documentation; they increasingly act as facilitators of performance improvement and risk management in digitally enabled environments. Their training activities influence how development engineers understand standards, apply quality methods, and integrate them into software development processes. Consequently, the effectiveness of quality related training has a direct impact on project outcomes, product reliability, and organisational competitiveness.

Learning Management Systems (LMS) have emerged as key tools supporting this transition. Beyond simply hosting training materials, modern LMS platforms provide structured learning paths, progress tracking, analytics dashboards, and adaptive content delivery. These capabilities enable organisations to personalise training, monitor competency development, and identify potential skill gaps early (Hernández-Chávez, et al., 2021). Such digital learning environments are increasingly recognised as essential for supporting continuous skill development and organisational performance in digitalised industries (Supiani, 2024).

However, in many organisations, quality training programmes remain predominantly theoretical and compliance oriented. They often focus on explaining standards, procedures, and documentation requirements without sufficiently connecting training content to real software development challenges, digital tools, or data driven practices. As digital transformation initiatives accelerate, this misalignment becomes more visible. Development engineers may complete mandatory quality training sessions, still struggle to apply digital quality methods effectively within complex software projects (Hernández-Chávez, et al., 2021).

This situation raises important questions regarding the actual contribution of quality engineering training to organisational digitalisation. While technological infrastructure and digital tools are increasingly available, the translation of training activities into measurable digital competencies and performance improvements is not always evident. In particular, limited attention has been given to understanding how training delivered by quality engineers to development engineers involved in software development supports broader digital transformation objectives.

Therefore, this study seeks to investigate the effectiveness of such training initiatives within a real organisational context. By applying an European Foundation for Quality Management (EFQM) based assessment combined with a structured root cause analysis, the research aims to identify the underlying factors that limit training effectiveness and to provide practical recommendations for aligning training design, delivery, and evaluation with digitalisation needs and organisational performance goals.

II. Literature Review

The rapid pace of digital transformation in industrial sectors has significantly altered the skills required of quality management teams. Traditional training programs, which rely heavily on in person sessions and static materials, are increasingly insufficient to address the needs of modern, digitally enabled workflows. Robertson and Lapiņa emphasize that conventional approaches to quality management training fail to equip employees with the competencies necessary for operating in digitalised environments. Their study highlights that organisations must increasingly adopt digital tools, including LMS, in order to maintain workforce performance and organisational efficiency (Robertsone & Lapina, 2021).

Hernández-Chávez et al. further explore the role of digital technologies in engineering education by developing a virtual reality automotive laboratory designed to support training activities. Their research demonstrates that immersive digital environments can significantly enhance the acquisition of practical competencies, allowing learners to interact with realistic industrial scenarios. The authors show that integrating digital tools into training enables more engaging and effective learning experiences, while also facilitating continuous monitoring of learner progress (Hernández-Chávez, et al., 2021).

Ralea et al. investigate the broader digital transformation of quality management and stress the importance of structured and adaptive training programmes. Their findings indicate that organisations that conduct initial competency assessments and continuously monitor employee progress through digital tools achieve more effective training outcomes. The study also highlights that early identification of skill gaps allows organisations to provide targeted learning interventions, improving both individual and team performance while aligning quality management practices with digital workflows (Ralea, et al., 2019).

Supiani evaluates the effectiveness of Learning Management Systems using the Leadership, Planning, Organizing, Monitoring and Relationships (LPOMR) model and shows that LMS based training success depends not only on technological infrastructure but also on instructional design and feedback mechanisms. The study emphasises that well designed LMS environments promote active learning, autonomy, and engagement, particularly when training activities are supported by structured feedback and data driven evaluation (Supiani, 2024).

Similarly, Ponsignon, Kleinhans and Bressolles demonstrate that quality management functions can act as key enablers of organisational digital transformation. Based on qualitative interviews, the authors show that continuous training and engagement of quality teams are essential for the successful adoption of digital tools and processes. Their research highlights the importance of aligning training initiatives with organisational transformation strategies (Ponsignon, et al., 2019).

Krivokuca also contributes to this discussion by emphasising the role of quality culture in the digital age. The study suggests that the development of a digital quality culture requires continuous learning and the integration of new technologies into organisational practices. This reinforces the idea that training plays a critical role in enabling digital transformation (Krivokuca, 2019).

Mendi focuses on the implementation of digital twin technology in automotive production and demonstrates how digital simulations can support process optimisation and decision making. The study highlights the growing need for engineers to develop digital competencies that allow them to work with advanced technologies in modern production environments (Mendi, 2022).

Tafakur* examines the development of online training platforms in the automotive industry and shows that digital training environments must be aligned with the real needs of trainees. The research emphasises the importance of designing training programmes that reflect actual work practices and technological requirements (Tafakur*, et al., 2023).

In addition, Bhargavi. S investigates the impact of training and development on employee performance in the automotive industry and confirms that investment in human capital remains a key driver of organisational performance (Bhargavi., 2024).

Recent literature on employee training emphasises both the importance and the complexity of evaluating training effectiveness in organisational contexts. Systematic reviews highlight key approaches and best practices that enhance training outcomes and align development programmes with organisational goals (Zalukhu, et al., 2025). Empirical studies demonstrate that training transfer, or the application of learning in the workplace, critically influences employee performance (Maheswari & Nalini, 2025). In addition, research on specific training modalities, such as on the job training, underlines their positive effects on task proficiency and adaptability (Naing & Fang, 2026).

Overall, the literature consistently shows that digital tools and LMS platforms play an increasingly important role in supporting employee learning and organisational transformation. However, most existing studies focus either on technology adoption or on general employee engagement. Limited attention has been given to the effectiveness of training delivered specifically by quality engineers to development engineers involved in software development. In particular, the relationship between training effectiveness, organisational performance, and digital transformation outcomes remains insufficiently explored. To address this

gap, the present study applies an EFQM based assessment combined with root cause analysis to evaluate how quality engineering training supports organisational digitalisation in a real industrial context.

III. Materials and Methods

This study adopts a qualitative and exploratory research design aimed at diagnosing the factors that limit the effectiveness of training provided by quality engineers in supporting digitalisation needs within an automotive software company. The research follows a case based approach, focusing on organisational practices related to training, performance management, and digital transformation within a real industrial context. This approach was considered appropriate because the objective of the study is not to test a predefined hypothesis, but rather to understand organisational practices, identify existing gaps, and explore underlying causes affecting training effectiveness.

The EFQM model was selected as the main analytical framework guiding the research. The EFQM model provides a holistic perspective on organisational performance and transformation and is widely used to assess organisational maturity, continuous improvement, and strategic alignment. Within this study, the analysis focused specifically on EFQM Criterion 5 – Driving Performance and Transformation, which is particularly relevant for examining the relationship between training, organisational change, and digitalisation in software driven automotive environments. The assessment included the five subcriteria of Criterion 5: Drives Performance and Manages Risk (5.1), Transforms the Organisation for the Future (5.2), Drives Innovation and Technology (5.3), Leverages Data-Driven Insights and Knowledge (5.4), and Defines and Implements the Overall Experience (5.5). The EFQM framework was selected due to its suitability for identifying gaps between technological maturity and human capability development, especially in the context of digital transformation (EFQM, 2024), (Hariri, et al., 2024).

Data collection relied on structured input from key organisational stakeholders involved in training and digitalisation activities. The stakeholder groups included quality engineers, project managers, software developers, and training managers, as they represent complementary perspectives on training design, implementation, and outcomes in software development projects. Rather than analysing individual opinions separately, stakeholder input was consolidated and triangulated using objective organisational evidence. This evidence included training documentation, Learning Management System (LMS) reports, internal performance dashboards, and process descriptions. The use of both stakeholder input and documentary evidence was intended to enhance the reliability of the assessment and reduce subjectivity.

The collected information was analysed using the EFQM Results – Approach – Deployment – Assessment & Refinement (RADAR) logic, which provides a structured method for assessing organisational maturity. The RADAR logic supports the evaluation of approaches, deployment, assessment, and refinement processes. Based on this logic, aggregated maturity scores were derived for each sub-criterion of EFQM Criterion 5. The scoring process aimed to reflect overall organisational maturity rather than individual perceptions. Final scores were expressed on a percentage scale from 0 to 100, enabling comparison across the assessed sub-criteria and supporting the identification of strengths and improvement areas.

To facilitate interpretation and communication of the results, the aggregated EFQM scores were visualised using a radar chart. This visual representation provides an integrated overview of maturity levels across performance management, organisational transformation, innovation and technology, data driven insights, and overall experience. The radar chart supports the identification of imbalances between technological capabilities and training effectiveness.

Following the EFQM assessment, a root cause analysis was conducted using an Ishikawa (Fishbone) diagram. The Ishikawa method was applied as a diagnostic tool to systematically structure and categorise the factors contributing to the limited effectiveness of training. Identified causes were grouped into six main categories: People, Methods, Measurement, Technology, Management, and Environment. This step enabled a deeper exploration of the relationships between organisational practices, training design, and digitalisation outcomes within the automotive software context.

IV. Results

In this study, the self assessment of EFQM Criterion 5 (sub-criteria 5.1–5.5) was conducted from the perspective of four key stakeholder groups within an automotive software company. These stakeholders provide complementary insights into the implementation, impact, and alignment of digitalisation training programmes:

- Quality Engineers – responsible for planning, delivering, and monitoring training programmes and ensuring alignment with organisational objectives.
- Project Managers – evaluate the impact of training on project execution, efficiency, and performance outcomes.
- Software Developers – provide feedback on the applicability and effectiveness of the training in daily work and skill development.
- Managers – oversee strategic alignment, resource allocation, and overall training governance.

By aggregating the perspectives of these stakeholder groups, the study captures a comprehensive view of organisational performance, transformation, innovation, data utilisation, and resource management. This approach ensures that the assessment reflects both operational realities and strategic objectives while maintaining an impersonal and generalised perspective suitable for academic reporting. The scores presented therefore reflect the combined perspectives of all stakeholder groups, providing an aggregated assessment of organisational maturity for each sub-criterion.

Each sub-criterion of EFQM Criterion 5 is analysed individually using consolidated input from all stakeholder groups, beginning with sub-criterion 5.1 (EFQM, 2024).

A. 5.1 Drives Performance & Manages Risk

Based on the consolidated perspectives of key stakeholder groups, digitalisation related training activities are systematically monitored using a Learning Management System (LMS). Attendance, course completion, assessment results, and post training evaluations are recorded and reviewed through periodic reports. Risk management practices are embedded in training processes, addressing potential issues such as insufficient digital competencies, trainer availability, and technical disruptions. Lean and agile principles are applied in training planning, and continuity is ensured through remote delivery options and predefined contingency measures.

The assessment is supported by objective evidence, including LMS reports (attendance and completion rates), risk registers, training schedules, agile planning documents, sprint retrospectives, and continuity plans.

The analysis indicates that a key strength lies in the structured organisation and monitoring of training activities. Established risk awareness and agile practices enable continuous operational adjustments and ensure the continuity of training delivery even in dynamic project environments.

However, several opportunities for improvement were identified. Training performance indicators are not directly linked to digitalisation outcomes or project performance, and risk management activities mainly address operational continuity rather than long term capability development. In addition, predictive or scenario based performance evaluation is limited.

Based on this assessment, the criterion is rated C (65), reflecting a moderate level of maturity: performance monitoring and risk management practices are in place, but the impact of training on digitalisation outcomes is not sufficiently measured.

B. 5.2 Transforms the Organisation for the Future

Stakeholder input indicates that digitalisation training initiatives are aligned with organisational strategy and sustainable value creation objectives. Transformation roadmaps, ecosystem scanning, and trend analysis are used to anticipate future challenges. Organisational structures and processes are periodically reviewed and supported by structured change management approaches and stakeholder engagement.

This assessment is supported by objective evidence, including transformation roadmaps, change management documentation, portfolio and programme reviews, and agile retrospectives.

One important strength is the clear alignment between training initiatives and strategic transformation goals. Stakeholder involvement in transformation related activities is consistent, and agile methods support organisational adaptability.

However, opportunities for improvement were also identified. The long term impact of training on future digital roles is not clearly measured, and digital training priorities are not fully reflected in operational capabilities. Furthermore, foresight and future skills analysis are rarely integrated into training design.

Based on this assessment, the criterion is rated B (75), reflecting a good level of maturity, while highlighting the need to strengthen the translation of strategy into digital skill development.

C. 5.3 Drives Innovation & Technology

The assessment shows that innovation and digital technologies are actively leveraged to support both current operations and future transformation. Digital tools, automation, and emerging technologies are integrated into organisational practices, supported by a culture that encourages creativity and experimentation. Stakeholders are engaged throughout the innovation lifecycle.

This assessment is based on objective evidence such as innovation project documentation, technology adoption reports, digitalisation initiative records, and stakeholder engagement logs.

A major strength is the strong technological environment supporting innovation, together with the high availability of digital tools and platforms. The organisational culture promotes experimentation and cross functional collaboration.

However, opportunities for improvement remain. The impact of innovation focused training on the development of applied digital competencies is not systematically evaluated. Innovation related training is not consistently linked to quality engineering practices, and emerging technologies could be more effectively embedded in structured training programmes.

Based on this assessment, the criterion is rated B (80), reflecting a good to high level of maturity, as innovation and technology practices are well established, while training effectiveness lags behind technological maturity.

D. 5.4 Leverages Data-Driven Insights & Knowledge

Training related and operational data are collected and used to support performance management and transformation initiatives. Dashboards and reports provide descriptive insights, while governance mechanisms ensure data quality, ethical use, and cybersecurity. Knowledge sharing platforms support organisational learning and collaboration.

The evaluation draws on evidence including performance dashboards, LMS reports, internal knowledge management systems, and data governance policies.

Strengths include the availability of structured training and performance data and established governance mechanisms ensuring data quality and security. Internal platforms facilitate knowledge sharing across the organisation.

However, several improvement areas were identified. Advanced analytics are only limitedly used to evaluate training effectiveness, and available data is not fully exploited to anticipate future digital skill requirements. Knowledge management practices are also not consistently connected to training outcomes.

Based on this assessment, the criterion is rated C (65), reflecting a moderate level of maturity, as data systems exist but their analytical use remains limited.

E. 5.5 Defines & Implements the Overall Experience

Stakeholders confirm that financial, tangible, and intangible resources supporting training are managed responsibly. Digital learning platforms enhance accessibility and continuity, while sustainability principles are partially integrated through reduced travel and efficient resource use.

The assessment is supported by objective evidence such as resource management reports, financial tracking records, sustainability audit results, and asset lifecycle documentation.

Strengths include stable infrastructure supporting training delivery and responsible management of organisational assets. Digital platforms improve accessibility and participation across roles.

However, opportunities for improvement were identified. Training programmes are not sufficiently personalised to address different levels of digital competence, and circular economy principles are not explicitly integrated into training design. Resource allocation tends to prioritise compliance and operational requirements rather than advanced digital capability development.

Based on this assessment, the criterion is rated B (75), reflecting a good level of maturity, with opportunities to strengthen the strategic use of resources for digital skill development.

Error! Reference source not found.. summarises the aggregated EFQM Criterion 5 scores for all assessed sub-criteria.

TABLE 1. OVERVIEW OF EFQM CRITERION 5 ASSESSMENT RESULTS

Criterion	Stakeholders	Score	Maturity Level
5.1 Drives Performance & Manages Risk	QAE, PM, Developer, Manager	65	Moderate
5.2 Transforms the Organisation for the Future	QAE, PM, Developer, Manager	75	Good
5.3 Drives Innovation & Technology	QAE, PM, Developer, Manager	80	Good to High
5.4 Leverages Data-Driven Insights & Knowledge	QAE, PM, Developer, Manager	65	Moderate
5.5 Manages Assets & Resources	QAE, PM, Developer, Manager	75	Good

Based on the consolidated EFQM assessment results, the aggregated scores for Criterion 5 (sub-criteria 5.1–5.5) were visualised using a radar chart in order to provide a clearer and more intuitive overview of organisational maturity. This visual representation helps to highlight the relative strengths and weaker areas across performance management, organisational transformation, innovation and technology, data driven insights, and overall experience, making it easier to identify gaps related to training effectiveness.

Fig. 1 illustrates the aggregated EFQM scores for Criterion 5 and reveals noticeable differences between the assessed areas. The highest maturity level was observed for innovation and technology (5.3 – score 80), followed by organisational transformation and overall experience (5.2 and 5.5 – score 75). These results suggest that the organisation benefits from a strong technological environment and from the availability of resources and structures that support transformation and change initiatives.

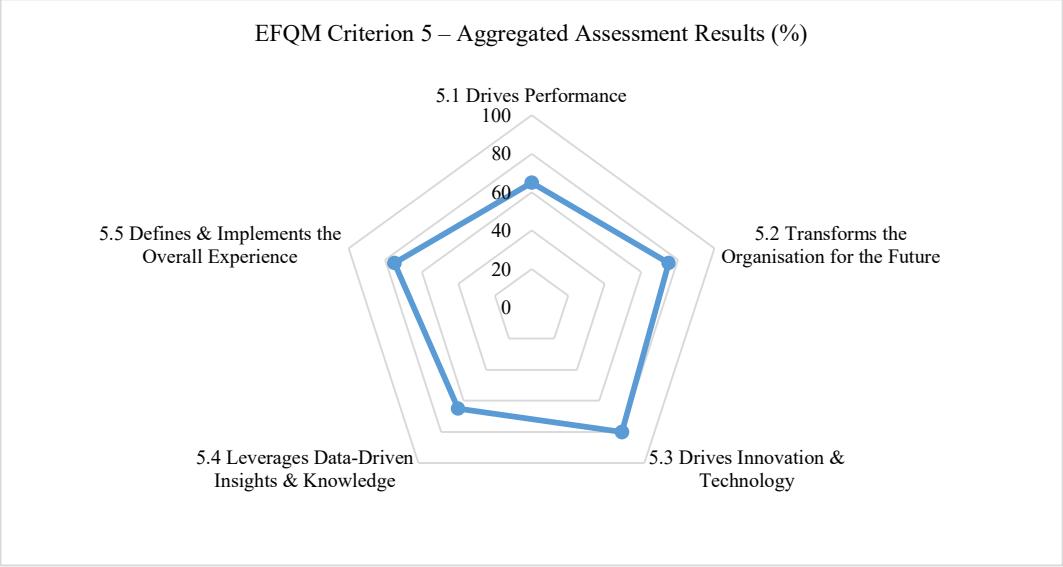


Fig. 1 EFQM - RADAR chart

In contrast, lower scores were recorded for performance management (5.1 – 65) and data driven insights and knowledge (5.4 – 65). These results suggest that, although training activities are well structured and supported by adequate infrastructure,

their outcomes are mainly monitored at an administrative level, focusing on participation and completion rather than on the development of measurable digital competencies.

The observed imbalance between higher innovation related scores and lower performance and data related scores points to a gap between the organisation's technological maturity and its ability to translate training efforts into effective digitalisation outcomes. While quality engineers play a key role in the design and delivery of training programmes, the current training approach appears to provide limited support for the acquisition and practical application of advanced digital skills required for digital transformation.

Consequently, these findings indicate that training provided by quality engineers is not fully effective in addressing the organisation's digitalisation needs. To further investigate the underlying causes of this limitation, a root cause analysis was conducted using an Ishikawa cause and effect diagram.

Based on the EFQM assessment results and the interpretation of the radar chart, a root cause analysis was conducted to further investigate the factors contributing to the limited effectiveness of training provided by quality engineers in addressing digitalisation needs. The analysis was informed by consolidated stakeholder input and objective evidence identified during the assessment process.

The identified factors were systematically structured using the Ishikawa (Fishbone) framework and grouped into six main categories: People, Methods, Materials, Environment, Equipment/Technology, and Management. This categorisation enabled a structured and comprehensive examination of the underlying causes influencing training effectiveness (Glevitzky, et al., 2025).

Fig. 2 presents the Ishikawa diagram summarising the identified causes. Each main branch represents one causal category, while the sub branches illustrate specific contributing factors. The diagram provides a visual synthesis of the analysis and highlights key areas where deficiencies contribute to the gap between training delivery and digitalisation requirements.

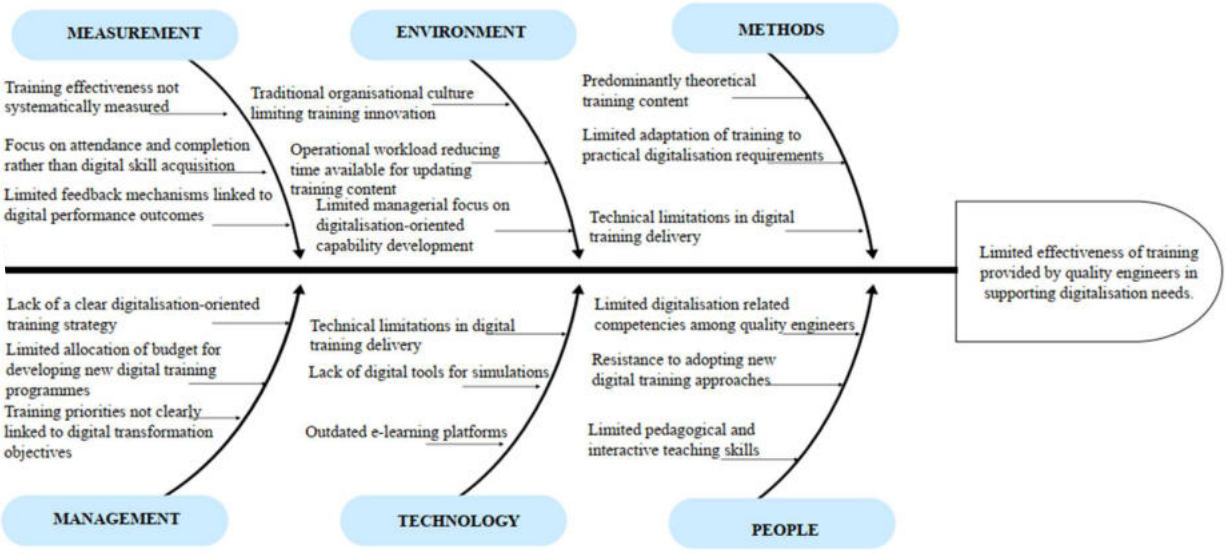


Fig. 2 Ishikawa diagram

The Ishikawa diagram highlights that the limited effectiveness of training provided by quality engineers in supporting digitalisation needs is not caused by a single factor, but by a combination of interrelated organisational issues. Rather than being driven by technological limitations, the results suggest that the main challenges are associated with organisational practices, training design, and the way training outcomes are evaluated and integrated into everyday work.

Measurement and methods related causes appear to play a central role. The analysis indicates that training effectiveness is mainly assessed through administrative indicators such as attendance, completion rates, or participant satisfaction, while the development and practical application of digital competencies are rarely measured. As a result, organisations have limited visibility regarding the real impact of training on software development activities, project performance, or digital transformation objectives. At the same time, training content is often predominantly theoretical and strongly oriented towards compliance and standards, with limited integration of real project based learning or hands on digital practices. This reduces the transfer of knowledge from training sessions to real development environments and helps explain the lower maturity levels previously identified for performance management and data driven insights.

People and management related factors also contribute significantly to the identified gap. The analysis suggests that quality engineers may not always possess the advanced digital competencies or pedagogical skills required to design and deliver training adapted to software driven environments. In addition, the absence of a clearly defined digitalisation oriented training strategy limits the alignment between training priorities and broader organisational transformation goals. Training

initiatives tend to be driven primarily by compliance requirements and short term operational needs rather than by long term capability development and digital skill building.

Environmental and technological aspects further reinforce these challenges. High operational workloads reduce the time available for continuous learning and regular training updates, while the limited use of simulations, collaborative platforms, and modern learning technologies constrains the development of practical digital competencies. Consequently, although training processes are structured and supported by adequate infrastructure, their contribution to digitalisation outcomes remains limited.

Overall, the Ishikawa analysis confirms and complements the EFQM assessment results, which indicated lower maturity levels in performance management and data driven insights compared to innovation and technological capabilities. Together, these findings emphasise that improving training effectiveness requires a systemic approach that simultaneously addresses measurement practices, instructional design, organisational priorities, and capability development.

V. Discussion

This study analysed the effectiveness of training provided by quality engineers in supporting digitalisation needs within an automotive software context, based on the assumption that such training should play a key role in enabling digital transformation. The results reveal a clear discrepancy between the organisation's relatively high level of technological maturity and the limited effectiveness of training in supporting the development and practical application of digital competencies. While higher maturity levels were observed in innovation, technology adoption, organisational transformation, and resource management, lower scores in performance management and data driven insights indicate difficulties in translating training activities into measurable digitalisation outcomes.

These findings are consistent with previous research on digital transformation, which shows that technological progress often advances faster than the development of human capabilities and organisational learning mechanisms. Existing studies emphasise that training effectiveness is frequently assessed using administrative indicators such as attendance, completion rates, or participant satisfaction, rather than indicators linked to skill application, behavioural change, or performance improvement. The results of this study reinforce these observations, showing that training outcomes are insufficiently connected to project performance, software development practices, and the real use of digital tools.

The Ishikawa analysis provides additional insight into the underlying causes of this gap. The results indicate that the limitations identified are not primarily related to the availability of technology or infrastructure. Instead, they are strongly associated with organisational methods, measurement practices, people related factors, and management decisions. In particular, the findings suggest that quality engineers, despite their technical expertise, may lack advanced digital competencies and pedagogical preparation required to design and deliver training programmes aligned with digitalisation objectives. Similar challenges have been highlighted in prior research, which stresses the importance of instructional design, trainer capabilities, and learner engagement in digitally oriented training environments.

Another important observation emerging from the analysis is the limited integration between training activities and broader organisational performance systems. Training is often treated as a standalone activity rather than as an integral component of performance management and continuous improvement. As a result, organisations may invest significant resources in training without fully understanding its contribution to digital transformation goals.

From a broader perspective, these results underline that successful digital transformation requires more than technological investment alone. Effective alignment between training design, delivery, and evaluation is essential to ensure that digital competencies are developed and applied in practice. Without systematic measurement of training impact and stronger integration of digital skills into training content, organisations risk underutilising their technological potential and slowing the pace of transformation.

These findings also open directions for future research, particularly in developing quantitative measures to better capture the impact of training on digital skill development and digitalisation outcomes across different organisational contexts. Further studies could explore how different training design approaches influence learning transfer and long term performance improvement.

VI. Conclusions

This study demonstrates that, despite relatively high maturity levels in innovation, technology, and organisational transformation, training provided by quality engineers does not fully support the development and application of digital competencies in a software driven automotive environment. Lower maturity levels in performance management and data driven insights reveal a clear gap between technological capabilities and the organisation's ability to translate training efforts into measurable digitalisation outcomes.

The findings emphasise the need for organisations to move beyond compliance oriented training approaches and towards more practice based, data driven, and strategically aligned learning initiatives. Strengthening the connection between training outcomes, digital skill development, and organisational performance represents a key step in improving the effectiveness of training programmes.

Several limitations of this study should be acknowledged. The analysis is based on an EFQM self assessment approach supported by aggregated stakeholder input, which may involve subjective perceptions despite the use of objective evidence. In addition, the research focuses on a single organisational context within the automotive software industry, which limits the generalisability of the findings. Training effectiveness was assessed mainly through organisational processes and perceived outcomes rather than direct quantitative measures of individual skill acquisition or long term performance impact.

Future research could address these limitations by incorporating quantitative indicators of digital skill development and linking training participation to project and organisational performance outcomes. Longitudinal studies may provide further insight into how improvements in training design and evaluation influence digitalisation results over time. Comparative studies across multiple organisations or industries could also enhance understanding of contextual factors affecting training effectiveness. Finally, future research could explore the use of advanced analytics and learning data to support more systematic and evidence based evaluation of training practices within EFQM based frameworks.

References

Bhargavi. S (2024), "Impact of Training & Development on the Employee Performance with Respect to Automotive Industry", Shanlax International Journal of Management, Vol. 11 No. 1, pp. 179–185.

Dias, A. R., Dias, A. M., Carvalho, A. M., Domingues, P. and Sampaio, P, (2023), "From Concept to Industry: Development of a Capability Roadmap Towards Quality 4.0", Proceedings of the 6th European Conference on Industrial Engineering and Operations Management, pp. 319-327.

EFQM (2024), "The EFQM Model 2025: Driving Sustainable Performance Together", EFQM, Brussels.

Garad, A. and Rozana, H. (2024), "Integrating Quality 4.0 and 5.0 with Organisational Learning: Shaping the Future of Quality and Organisational Excellence", in Proceedings of the International Conference on Quality Engineering and Management, pp. 1202–1226.

Glevitzky, M., Glevitzky, I., Mucea-Ștef, P., Popa, M., Dumitrel, G.-A. and Vică, M.L. (2025), "Integrated Risk Framework (IRF)—Interconnection of the Ishikawa Diagram with the Enhanced HACCP System in Risk Assessment for the Sustainable Food Industry", Sustainability, Vol. 17.

Hariri, A., Domingues, J. P. T. and Sampaio, P. (2024) "Integrating Sustainability into Excellence Models", Proceedings book of the 6th International Conference on Quality Engineering and Management, pp. 1069-1082.

Hernández-Chávez, M., Cortés-Caballero, J.M., Pérez-Martínez, Á.A., Hernández-Quintanar, L.F., Roa-Tort, K., Rivera-Fernández, J.D. and Fabila-Bustos, D.A. (2021), "Development of virtual reality automotive lab for training in engineering students", Sustainability, Vol. 13.

Krivokuca, M. (2019), "Extending quality culture in the digital age", Proceedings of the International Conference on Industrial Engineering and Operations Management, pp. 226–231.

Maheswari, P. and Nalini V. (2025), "Bridging the Gap Between Training and Practice: An Empirical Study on Training Transfer and Employee Performance", Journal of Marketing & Social Research, Vol. 2 No. 6, pp. 212-216.

Mendi, A.F. (2022), "A digital twin case study on automotive production line", Sensors, Vol. 22 No. 18.

Naing, T. S. and Fang, W. (2026), "The Impact of On-the-Job Training on Employee Performance: Mediating and Moderating Mechanisms in Myanmar's Manufacturing Sector", Administrative Sciences, Vol. 16 No. 1.

Ponsignon, F., Kleinhans, S. and Bressolles, G. (2019), "The contribution of quality management to an organisation's digital transformation: a qualitative study", Total Quality Management & Business Excellence, Vol. 30 No. sup1, pp. S17–S34.

Ralea, C., Dobrin, O.C., Barbu, C. and Tănase, C. (2019), "Looking to the future: digital transformation of quality management", Proceedings of the 13th International Management Conference "Management Strategies for High Performance", Bucharest, Romania, pp. 121–132.

Robertsone, G. and Lapiņa, I. (2021), "Digital transformation and the changing shape of quality management practices", RTU 62nd International Scientific Conference on Economics and Entrepreneurship SCEE'2021 Proceedings, pp. 82–99.

Supiani, D.A. (2024), "Evaluating learning management system (LMS) effectiveness: an LPOMR model approach", Pedagogia: Jurnal Ilmiah Pendidikan, Vol. 16 No. 2, pp. 71–77.

Tafakur*, A.S., Dewi, A.S. and Solikin, M. (2023), "Automotive Industry Employee Online Training Platform: How Should It Develop Based on the Needs of Trainees?", International Journal of Information and Education Technology, Vol. 13 No. 11, pp. 1713–1719.

Zalukhu, R. R., Lase, D. and Telaumbanua, A. (2025), "The Effectiveness of Employee Training and Development: A Systematic Literature Review from a Human Resource Management Perspective", Jurnal Manajemen, Kepemimpinan, dan Supervisi Pendidikan, Vol. 10 No. 2, pp. 1027–1043.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Professional Training in Lean Six Sigma: An analysis of Yellow Belt and Green Belt Programs

Mónica Puertas[1], Edgar Yalta [2], José Carlos Flores-Molina[3]

[1] Pontificia Universidad Católica del Perú; mpuerta@pucp.edu.pe;
[2] Pontificia Universidad Católica del Perú; edyaltag@pucp.edu.pe;
[3] Pontificia Universidad Católica del Perú; jflores@pucp.edu.pe;

***Abstract*— This study presents an analysis of two Lean Six Sigma training programmes offered by a continuing education institution in Peru: Lean Six Sigma Yellow Belt (LSSYB) programme, aimed at university students and recent graduates, and Lean Six Sigma Green Belt (LSSGB) programme, aimed at practising professionals. Using data collected between 2016 and 2025, statistical analyses (Kruskal-Wallis and Mann-Whitney) and segmentation techniques (k-means clustering) were applied to identify the factors influencing academic performance and to characterise the participants' profiles. In addition, the final projects from the LSSGB programme were analysed to identify the sectors and processes most frequently addressed in improvement initiatives. The results reveal differences between the two programmes in terms of the factors affecting performance, as well as distinct profiles based on gender, age, work experience, roles and specialization. This study aims to provide empirical evidence on the effectiveness and adaptability of Lean Six Sigma training in educational and professional contexts; it also offers recommendations for the design of inclusive training programmes aligned with the principles of continuous improvement and organisational excellence.**

Keywords – Lean Six Sigma; Professional Training; Higher Education; Continuous Improvement; Quality Education

I. Introduction

The continued pursuit of operational excellence in organisations has driven the adoption of continuous improvement methodologies. Lean Six Sigma (LSS) has become established as a prominent methodology for advancing quality management and process optimisation by combining Lean principles (focused on waste elimination) with Six Sigma techniques (statistical rigour to reduce variability), thereby achieving significant improvements in cost, quality and customer satisfaction (Antony et al., 2018; Kwak & Anbari, 2006; George, 2002; Munro et al., 2022; Snee, 2010). Since its emergence, the application of LSS has not been confined to the manufacturing industry; rather, it has expanded into other sectors (Makwana & Patange, 2021). This expansion reflects the adaptability of LSS as a quality improvement strategy across different organisational contexts.

With regard to professional training in LSS, it is structured into belt levels (typically Yellow Belt, Green Belt and Black Belt) that represent progressively more advanced competencies and roles in improvement projects (Pyzdek & Keller, 2018). Each level entails expectations regarding a professional's ability to apply LSS within an improvement project. For instance, a Green Belt professional should be prepared to lead medium-scope improvement projects using the DMAIC cycle (Define, Measure, Analyse, Improve and Control) (Munro et al., 2022), whereas a Black Belt professional undertakes projects of high complexity (Antony et al., 2017; Mcshane-Vaughn, 2022; Pyzdek & Keller, 2018). Overall, LSS training programmes aim to ensure that participants develop robust competencies in process analysis, data-driven decision-making and leadership (Antony et al., 2018; DeRuntz & Meier, 2009).

Despite growing interest in LSS training, both in industrial and educational settings, gaps remain in the literature. From a general perspective, prior LSS research has focused on cases of improvement project implementation in companies or on organisational impact—i.e., industrial contexts (Buestan et al., 2025). In the context of professional training, most studies have described the perceptions of certified professionals (Green, 2006), explored the critical success factors for implementing LSS in higher education or other educational and service institutions (Antony et al., 2017; Kuwaiti, 2019; Nadeau, 2017), examined the standardisation of content in advanced training courses (Louzada et al., 2023), or characterised participant profiles and analysed performance within a Black Belt certification programme (Flores-Molina et al., 2025). Against this background, there is a clear need to document the characteristics and outcomes associated with LSS programmes aimed at Yellow Belt and Green Belt audiences, who are typically university students or recent graduates and practising professionals, respectively.

This study seeks to provide empirical evidence on the effectiveness and adaptability of LSS training in educational and professional contexts through the analysis of two training programmes in Peru. This setting is of particular interest due to the coexistence of organisations at different levels of organisational maturity (e.g., small and medium-sized enterprises (SMEs), informal firms, digital transformation initiatives), wide sectoral diversity, and the growing demand for professional certifications (INEI, 2022; OECD, 2025). Specifically, the study examines a Lean Six Sigma Yellow Belt (LSSYB) programme, aimed at final-year university students and recent graduates, and a Lean Six Sigma Green Belt (LSSGB) programme, aimed at practising professionals. The LSSYB programme provides foundational knowledge, whereas the LSSGB programme offers more in-depth training and requires the completion of an improvement project. Delivered by

a Peruvian continuing education institution, these training programmes represent two articulated levels within LSS education.

Analysing participants between 2017 and 2025 makes it possible to identify factors that affect participants' academic performance and to characterise typical profiles using segmentation techniques. In addition, the practical application of LSS is assessed by collecting information on the improvement projects developed as part of the LSSGB programme. The findings are intended to inform recommendations for the design of LSS training programmes aligned with the principles of continuous improvement and organisational excellence, and tailored to participants' characteristics.

Accordingly, three objectives are established:

(1) To determine which participant profile factors significantly influence academic performance in each training programme (Lean Six Sigma Yellow Belt programme and Lean Six Sigma Green Belt programme). Specifically, the study evaluates the effect of variables such as modality, gender, age, speciality and—depending on the programme— employment status or role on final marks (academic performance).

(2) To characterise typical participant profiles in both training programmes through cluster analysis, in order to identify distinctive demographic segments.

(3) To explore trends in the practical application of Lean Six Sigma through the improvement projects undertaken within the Lean Six Sigma Green Belt programme, identifying the sectors in which it is most frequently applied and the types of processes most commonly addressed.

II. METHODOLOGY

A. Design and variables

The study was conducted using a case study design (Case & Light, 2011). The approach is quantitative with a descriptive scope. A Peruvian continuing education institution, recognised for its training programmes in quality management, was selected in order to analyse in detail two of its LSS offerings: the LSSYB programme and the LSSGB programme. Although both programmes are delivered by the same institution, participants come from a wide variety of educational institutions nationwide. For example, LSSYB participants come from 49 higher education institutions and 13 regions of Peru (Lima, Callao, Arequipa, Ica, Piura, Ancash, La Libertad, Ayacucho, Cusco, Junín, Apurímac, Lambayeque and Puno), while LSSGB participants come from 53 higher education institutions and 14 regions (Lima, Arequipa, Callao, La Libertad, Piura, Lambayeque, Cajamarca, Junín, Ancash, Ica, Cerro de Pasco, Apurímac, Tacna and Moquegua). This indicates that the study is not limited to participants from a single organisation or institution; rather, it reflects a multi-institutional and multi-regional context for the adoption of LSS in Peru.

The dependent variable was defined as each participant's final grade (academic performance), measured on a 20-point scale (0–20), with 11 as the minimum passing score. The independent variables including in both training programmes are as follows:

a) *Modality*: a dichotomous variable (face-to-face and online), where each mode differs in terms of content delivery, interaction with the instructor and the assessment system.

b) *Gender*: a variable recording the participant's self-declared gender (male and female), based on the information provided in the institution's enrolment form.

c) *Age Group*: a categorical variable divided into three groups that vary by training programme. For LSSYB, the categories were 19–23 years, 24–26 years and 27 years and above; for LSSGB, the categories were 21–29 years, 30–39 years and 40 years and above. This categorisation allows for the analysis of performance differences across age groups, taking into account each programme's target audience.

d) *Speciality*: a categorical variable classifying participant as either engineering (industrial engineering, civil engineering, software engineering, among others) or non-engineering (disciplines such as economics, social sciences, business administration, among others). This variable is included because the institution predominantly serves an engineering-oriented student population. However, increasing demand has been observed among participants from non-engineering backgrounds.

e) *Type of institution*: a categorical variable classifying the participant's higher education institution of origin as either public or private. This distinction is important given the substantial structural and pedagogical differences between public and private higher education in Peru, which influence students' prior academic preparation. Such disparities may affect the development of analytical and quantitative competencies required for successful learning in LSS.

Because the training programmes target different audiences, programme-specific variables were also included, as follows:

a) *Employment status*: a binary variable indicating whether an LSSYB programme participant was employed at the time of the programme (yes/no). This variable allows the examination of the potential impact of studying and working simultaneously during LSS training.

b) *Role*: a three-category variable capturing the organisational hierarchy level of LSSGB programme participants (senior management and heads of department, supervisors, and technical or operational staff). This classification reflects the typical organisational structure of Peruvian organisations and enables analysis of whether professionals' level of responsibility, often linked to experience and acquired competencies, may influence academic performance.

c) *Project implementation sector*: variable that takes into account the sector in which the LSSGB programme student developed their final project (e. g. mining, food, health, etc.).

d) *Type of project improvement*: variable that takes into account the type of improvement addressed by the student on the LSSGB programme (*time reduction*, *reduction of nonconformities*, or *waste elimination*). Projects focused on *waste elimination* are associated with 'wastage', an inevitable loss that occurs during the process, rather than a non-conformity. In contrast, projects focused on *nonconformity reduction* are associated with the failure to meet an established requirement, without necessarily constituting waste. Lastly, projects focused on *time reduction* are associated with the optimisation of time in processes.

B. Training programmes

The Lean Six Sigma Yellow Belt (LSSYB) programme is a 36-hour introductory programme designed to familiarize participants with the fundamental principles of Lean and Six Sigma. It is primarily targeted at final-year university students and recent graduates, most commonly from engineering or management programmes, although it is open to any field of study. The curriculum covers the Lean philosophy, the DMAIC methodology and basic quality tools. Because it is an entry-level programme, a final improvement project is not required; therefore, assessment is based on coursework and intermediate deliverables.

In contrast, the Lean Six Sigma Green Belt (LSSGB) programme is a more extensive 120-hour programme that provides in-depth training in statistical and Lean techniques. It is aimed at practising professionals seeking to certify their competencies at the Green Belt level. As a final requirement, participants must complete a team-based improvement project applied to a real organisational process, making use of the tools addressed throughout the programme. These Projects are carried out in teams of four to six participants with guidance from a supervisor.

Each training programme comprises a set of modules (LSSYB: 3 modules, LSSGB: 8 modules). Each module is assessed in a different way depending on the assessment method (e.g., exam, group assignment, final project, etc.). The final score (academic performance) is calculated as the average of the scores for each module. Both training programmes were offered in face-to-face modality (classroom-based sessions) and online modality (live synchronous sessions via a digital platform). In recent years, however, there has been a shift towards a greater proportion of online cohorts, driven by increased demand (particularly in the case of LSSGB, where geographical flexibility is valuable for working professionals) and the impact of the COVID-19 pandemic.

C. Sample and variables

A historical dataset of participants and their academic results was compiled for cohorts delivered between 2016 and 2025. For the LSSGB programme, data were gathered for the 2017–2025 editions, whereas for the LSSYB programme, data were gathered for the 2018–2025 editions—periods in which the available information was complete and comparable across cohorts. The final sample comprises 450 LSSGB programme participants and 388 LSSYB programme participants. The frequency distribution of participants' scores (academic performance) is shown in **Figure I**.

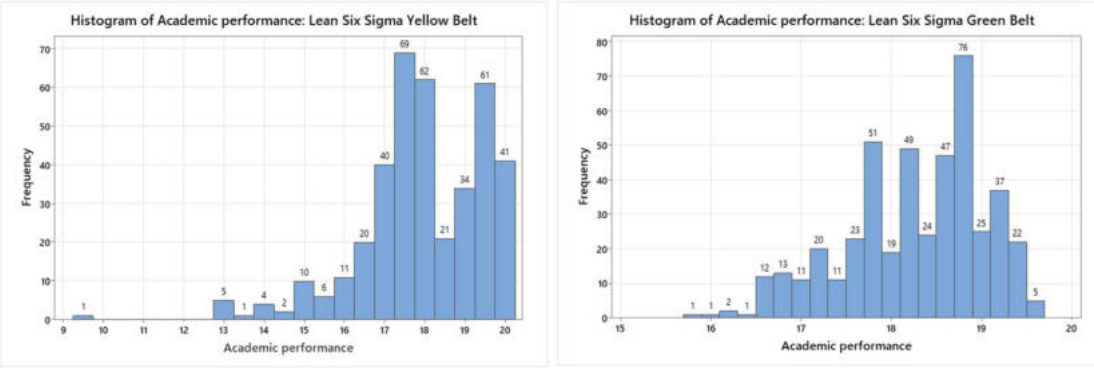


Fig. 1. Histogram: frequency distribution of the dependent variable (academic performance).

Table I presents the demographic data for the Lean Six Sigma Yellow Belt programme. This table provides a clear characterisation of the university-based profile of participants in the LSSYB programme. The data show a relatively balanced distribution across age groups, alongside a marked predominance of students or graduates from engineering programmes (84.54%) and higher participation of men than women. A further notable feature is the

prevalence of the online modality, together with a larger proportion of participants who were not employed at the time of the programme (57.73%). It is also important to highlight that most participants were either studying at, or had graduated from, private higher education institutions (62.10%).

Table II presents the demographic data for the Lean Six Sigma Green Belt (LSSGB)programme. In line with the programme´s professional orientation, ages are concentrated in early and mid-career stages, with a higher proportion of male participants, and a greater representation of individuals occupying intermediate organisational roles. The predominance of the online modality suggest that training is integrated with participant's work routines, while the high proportion of engineers (83.75%) reflects a specialisation profile similar to that observed in the LSSYB programme, albeit with a more consolidated focus on professional practice. Unlike the LSSYB programme, the distribution by type of higher education institution reported by participants is balanced.

Tables I and **II** also present an exploratory analysis of the dependent variable (academic performance), specifically the following descriptive statistics for each variable: mean, median and standard deviation. This made it possible to identify patterns of potential differences between groups and to guide subsequent inferential analyses.

Table I. Demographic data and exploratory analysis for the Lean Six Sigma Yellow Belt (2018–2025)[a]

Variable	Category	Frequency	Percentage	Academic performance		
				Mean	*Median*	*Standard deviation*
Modality	Face-to-face	155	39.95%	17.60	18.0	1.51
	Online	233	60.05%	18.21	18.0	1.52
Gender	Male	221	57.00%	18.08	18.0	1.67
	Female	167	43.00%	17.88	18.0	1.35
Age group	19-23 years	143	36.86%	18.02	18.0	1.35
	24-26 years	156	40.20%	17.84	18.0	1.55
	27 years and older	89	22.94%	18.10	18.0	1.79
Speciality	Engineering	328	84.54%	17.97	18.0	1.51
	Non-engineering	60	15.46%	17.93	18.0	1.73
Employment status	Employed	164	42.27%	18.20	18.3	1.58
	Unemployed	224	57.73%	17.80	17.7	1.49
Type of institution	Public	147	37.90%	18.10	18.0	1.40
	Private	241	62.10%	17.88	18.0	1.62

[a]The sample consisted of 388 participants

Table II. Demographic data and exploratory analysis for the Lean Six Sigma Green Belt (2017–2025)[a]

Variable	Category	Frequency	Percentage	Academic performance		
				Mean	*Median*	*Standard deviation*
Modality	Face-to-face	121	26.89%	18.13	18.2	0.80
	Online	329	73.11%	18.33	18.4	0.75
Gender	Male	293	65.11%	18.29	18.4	0.76
	Female	157	34.89%	18.24	18.2	0.78
Age group	21-29 years	206	45.78%	18.39	18.6	0.77
	30-39 years	169	37.55%	18.22	18.2	0.73
	40 years and older	75	16.67%	18.11	18.3	0.80
Speciality	Engineering	376	83.60%	18.28	18.4	0.77
	Non-engineering	74	16.40%	18.26	18.3	0.75
Role	Managers and leaders	78	17.33%	18.37	18.4	0.64
	Supervisors	272	60.45%	18.25	18.3	0.77
	Technicians and operators	100	22.22%	18.25	18.4	0.84
Type of institution	Public	223	49.56%	18.57	18.6	0.64
	Private	237	50.44%	18.98	18.0	0.77

[a]The sample consisted of 450 participants

In summary, Tables I and II confirm expected differences between the two programmes. For example, the LSSYB audience consists largely of young participants, many of whom are not employed at the time of enrolment, whereas all those enrolled in LSSGB are practising professionals holding roles with varying levels of responsibility. The tables also

show a notable level of female participation in both programmes, suggesting a growing presence of women professionals pursuing Lean Six Sigma certification.

D. Data analysis

To achieve Objective 1, inferential statistical techniques were employed. For each programme, the dependent variable was the participant's final mark (quantitative scale from 0 to 20). The independent variables for the LSSYB programme included modality (face-to-face vs online), gender (male vs female), age group (varies by training programme), speciality (engineering vs non-engineering), and employment status (employed vs not employed); For the LSSGB programme, the same set of variables were considered with employment status replaced by role within the organisation (technical/operational staff, supervisors, and managers/leaders). Since the data did not follow a normal distribution (Lean Six Sigma Yellow Belt: Kolmogorov-Smirnov, $p < 0.010$; Lean Six Sigma Green Belt: Kolmogorov-Smirnov, $p < 0.010$), non-parametric tests were applied in each case to determine whether there were statistically significant differences in final marks between the groups. A significance level of $p = 0.05$ was adopted. For variables with two categories, the Mann-Whitney U test was used, and for variables with three or more categories, the Kruskal-Wallis H test was used.

To achieve Objective 2, a k-means cluster analysis was performed to identify groups of participants with similar characteristics (Hair et al., 2019). Separate analyses were conducted for the population of each programme. The segmentation variables consisted of key demographic characteristics: gender, age, speciality and employment status for LSSYB; and gender, age, speciality and role for LSSGB. As required by the k-means algorithm, these categorical variables were numerically encoded. In both cases, a three-cluster ($k = 3$) solution was selected, as it produced clearly interpretable clusters of adequate size, with no cluster representing less than 15% of the sample. Mean profiles were calculated for each cluster, and χ^2 tests confirmed that all variables significantly differentiated the clusters ($p < 0.001$ for all variables), demonstrating the statistical validity of the segmentation. The resulting clusters therefore represent typical participant profiles within the programmes.

To achieve Objective 3, a descriptive analysis of the final projects from the LSSGB programme was conducted. Data were compiled for 44 group projects developed between 2018 and 2025, typically corresponding to one project per group of four to six participants. The variables 'Project implementation sector' and 'Type of project improvement' were used. Absolute and relative frequencies were calculated for each category to identify the most frequently addressed sectors and improvement types. As this analysis was exploratory and illustrative in nature, no hypothesis testing was performed. Its purpose was to illustrate practical trends in the application of Lean Six Sigma tools, complementing the assessment of academic performance with an applied dimension.

All statistical calculations were performed using Minitab 22. Results are presented separately for each programme (LSSYB followed by LSSGB), combining quantitative reporting with interpretative discussion, in line with the study objectives.

III. RESULTS

A. Six Sigma Yellow Belt

The results for the Lean Six Sigma Yellow Belt programme are presented below. First, to address Objective 1, non-parametric tests were conducted. The results are shown in the following table:

Table III. Results of Non-parametric tests: Lean Six Sigma Yellow Belt

Variable	Non-parametric tests		p-value
	Mann-Whitney U test	Kruskal-Wallis H test	
Modality	13855.00		< 0.001
Gender	19246.50		0.466
Age group		4.196	0.123
Speciality	9720.50		0.880
Employment status	15099.00		0.003
Type of institution	16563.00		0.281

For the LSSYB programme, significant differences were observed in mean final marks between participants enrolled in face-to-face and online modalities ($p < 0.001$), as well as between those who were employed and those who were not employed during the programme ($p = 0.003$). In contrast, no statistically significant differences in academic performance were identified with respect to gender, age groups, speciality or type of institution. In summary, while academic performance was broadly comparable across participants in this programme, both the instructional modality (online or face-to-face) and the presence of concurrent employment were both associated with differences in their academic outcomes.

To better understand these significant differences in the LSSYB programme, mean comparisons were conducted. **Figure 2** presents the results of comparisons across the categories of the modality and employment status variables:

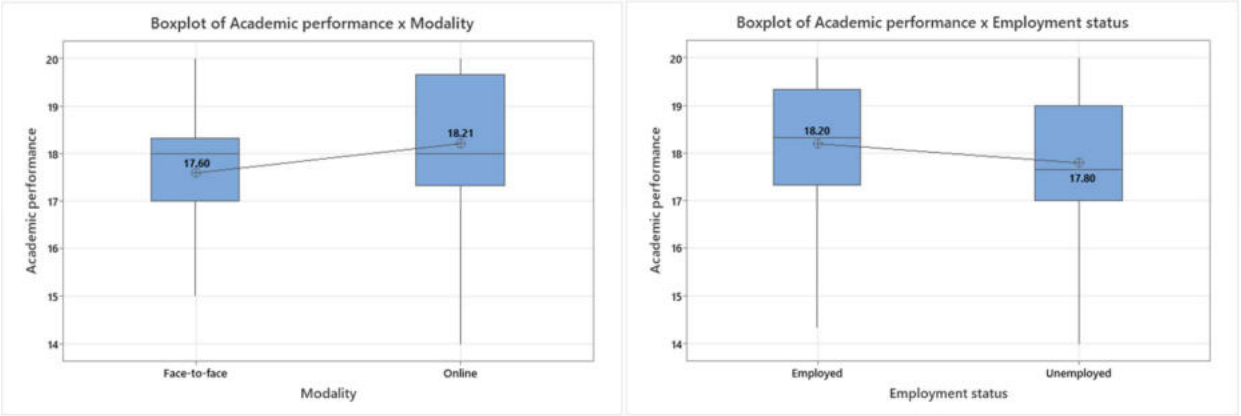


Fig. 2. Boxplot of Academic performance x modality ($p < 0.001$) and academic performance x employment status ($p = 0.003$).

With respect to modality, participants enrolled in the online modality achieved higher marks on average (M = 18.21, SD = 0.099) than participants in the face-to-face modality (M = 17.60, SD = 0.120), with the difference being statistically significant ($p < 0.001$). This finding suggests a performance advantage associated with synchronous online delivery at the Yellow Belt level, possibly due to the greater flexibility afforded by the online modality. In addition, participants who were employed during the programme obtained slightly higher marks on average (M = 18.20, SD = 0.120) compared with those who were not employed (M = 17.80, SD = 0.140), and this difference was also statistically significant ($p < 0.001$). This modest gap may indicate that concurrent workplace experience (i.e., exposure to practical improvement contexts) positively influences learning during the programme.

To address Objective 2, a k-means cluster analysis was conducted. Grouping the 388 LSSYB programme participants into three clusters resulted in clearly differentiated demographic profiles. The characteristics of the identified clusters (Clusters 1, 2 and 3) across the respective variables are shown in the following figure:

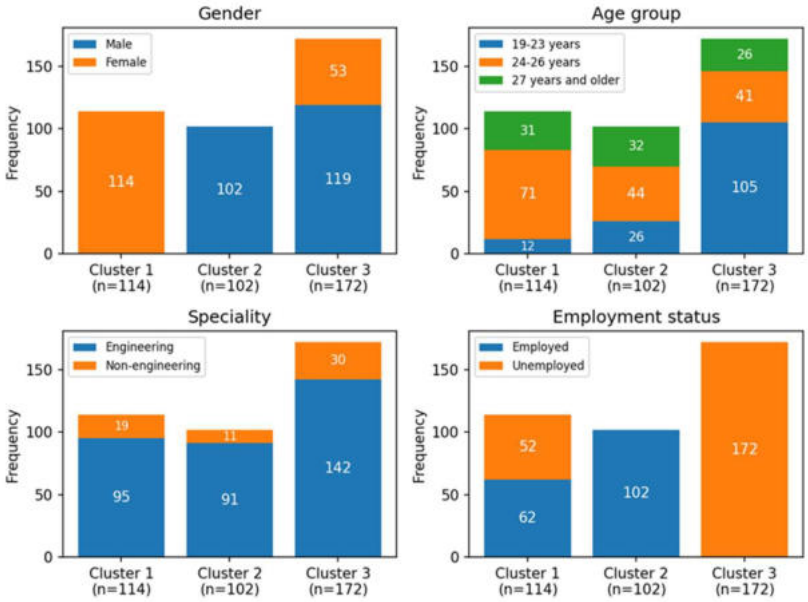


Fig. 3. Analysis of demographic profiles (k-means clusters) of the LSSYB. The distribution in tables was statistically significant (chi-squared test: $p < 0.001$).

Cluster 1 is characterised by young female engineers aged 24–26 who are moderately integrated into the labour market. This cluster (n = 114) is composed exclusively of women, almost all of whom have an engineering background. The predominant age falls within the second age group (24–26 years). Approximately half were employed at the time of the programme, while the remainder were recent graduates or final-year students without employment. This profile represents young women who are either entering the labour market or seeking certification before joining an organisation. From a gender perspective, this cluster highlights the growing interest and participation of women in Lean Six Sigma training.

Cluster 2 represents young, employed male engineers. This group (n = 102) is composed entirely of men, the majority of whom are engineering graduates (n = 91). All members of this cluster were employed while taking the programme. The age distribution is similar to that of Cluster 1, as it is concentrated in the 24–26 age group. This segment corresponds to early-carrier engineers who enrol in LSS training either to apply Six Sigma within their organisations or to strengthen skills relevant to professional advancement.

Cluster 3 represents, predominantly, very young male participants with no work experience. This is the largest cluster (n = 172) and, unlike Clusters 1 and 2, it presents a mixed gender composition, although men predominate and it is distinguished by the absence of employment among its members. They were therefore likely students with no work experience or recent graduates who had not yet entered the labour market. Moreover, this cluster concentrates the youngest participants (19–23 years). With respect to educational background, although most participants come from engineering programmes, this cluster also includes the highest proportion of participants from non-engineering disciplines (n = 30). Overall, Cluster 3 reflects the traditional Yellow Belt participant profile: university students or recent graduates without experience who seek Lean Six Sigma certification to complement their academic training and enhance future employability.

These findings confirm the presence of distinct participant subgroups within the population, ranging from young female engineering graduates (Cluster 1) to inexperience students and recent graduates (Cluster 3). Such heterogeneity suggests the need of differentiated pedagogical approaches. For example, participants in Cluster 3 may benefit from greater practical contextualisation due to their limited work experience, whereas Cluster 2 (employed participants) may contribute real-world examples but also face constraints related to time availability. This finding underscores the importance of segmenting instructional strategies within an apparently homogeneous programme to maximise learning outcomes for each group.

B. Lean Six Sigma Green Belt

The results for the Lean Six Sigma Green Belt (LSSGB) programme are presented below. To address Objective 1, non-parametric tests were conducted. The results are shown in **table IV**.

Non-parametric tests indicated a statistically significant difference between participants undertaking the programme online versus face-to-face (p = 0.021), as well as between age groups (p = 0.006). In contrast, no statistically significant differences in final marks were observed with respect to gender, speciality, organisational role, or participants' institution type of origin.

Table IV. Results of Non-parametric tests: Lean Six Sigma Green Belt

Variable	Non-parametric tests		p-value
	Mann-Whitney U test	Kruskal-Wallis H test	
Modality	17076.50		0.021
Gender	22036.50		0.443
Age group		10.253	0.006
Speciality	13516.00		0.595
Role		0.793	0.673
Type of institution	24804.50		0.713

To interpret these differences, mean comparisons were conducted. **Figure 4** presents the results of comparisons across categories of the modality and age group variables.

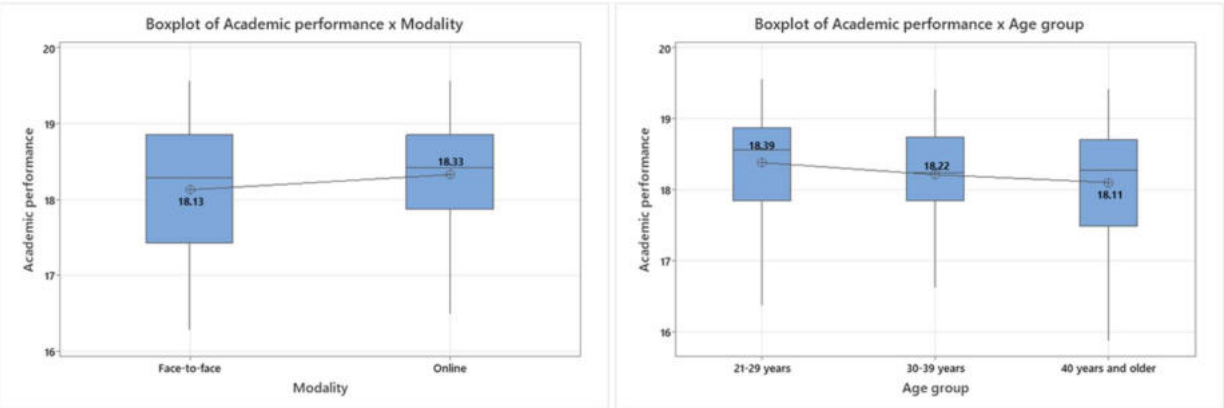


Fig. 4. Boxplot of Academic performance x modality (p = 0.021) and academic performance x age group (p = 0.006).

As observed for the LSSYB programme, participants in the online modality achieved higher performance on average (M = 18.33, SD = 0.751) than those attending the face-to-face sessions (M = 18.13, SD = 0.807), with a statistically significant difference (p = 0.021). This finding may be attributed to the affordances of online training for working professionals, including geographical and scheduling flexibility, which can facilitate programme engagement alongside professional responsibilities. Regarding age group, participants in the youngest group (21–29 years) achieved the highest mean final mark (M = 18.39, SD = 0.799), followed by those in the second age group (30–39 years) (M = 18.22, SD = 0.732) and the third age group (40 years and above) (M = 18.11, SD = 0.793). The difference between age groups were statistically significant difference (p = 0.006), suggesting a modest decline in performance with an increasing age, potentially reflecting differing time constraints or learning contexts among more senior professionals.

To address Objective 2, a k-means cluster analysis was conducted. Since k-means requires 'k' to be set a priori, alternative solutions (e.g., k = 2, k = 3 and k = 4) were evaluated using fit criteria and substantive criteria (interpretability and minimum cluster size). K = 3 was selected as it offered an appropriate balance between group separation, interpretative stability and sufficient cluster sizes. These clusters revealed demographic segments primarily differentiated by combinations of age, gender, speciality and professional role. The resulting profiles (Clusters 1, 2 and 3) and their distribution across the relevant variables, are shown in **Figure 5**.

Cluster 1 represents young technical professionals with a relatively balanced gender composition. This group (n = 106) is characterised by a strong concentration of participants in the youngest age group (21–29 years) and a more gender diversity than the other clusters. Members primarily hold technical or operational roles in their organisations. Another distinctive feature is that participants from non-engineering fields (n = 58) outnumber those from engineering backgrounds. In summary, this profile corresponds to early-career professionals with limited experience who perceive the LSSGB programme as a means to enhance their effectiveness in technical roles.

Cluster 2 represents mid-career supervisors and constitutes the largest group (n = 251), representing the typical participant profile in this program. It comprises professionals predominantly from engineering backgrounds (n = 251), with representation of both young adults (21–29 years) and mid-career adults (30–39 years). The gender distribution is predominantly men (n = 160). Most members (n = 201) hold intermediate or supervisory positions within their organisations, involving responsibility for processes and teams, though without occupying senior executive roles.

Cluster 3 represents experienced senior managers. This group (n = 93) includes the oldest participants (40 years and above), a substantial majority of men (n = 70), and a predominance of engineering backgrounds (n = 77). Roles tend to be managerial and head-of-department positions, including area directors, plant managers, and similar roles. This profile corresponds to senior professionals who pursue Green Belt training to formalise experiential knowledge and to lead continuous improvement initiatives within their organisations. The presence of this cluster highlighted that LSS attracts not only technical and middle-level professionals, but also organisational leaders interested in advancing operational excellence.

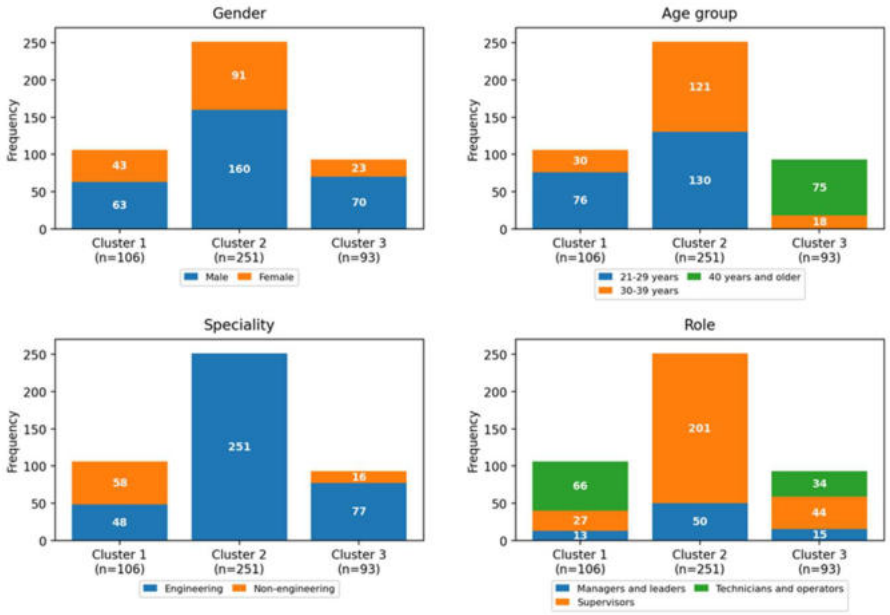


Fig. 5. Analysis of demographic profiles (k-means clusters) of the LSSYB. The distribution in tables was statistically significant (chi-squared test: p < 0.001).

To address Objective 3, a descriptive frequency analysis of the final projects developed in the LSSGB programme was conducted. This analysis enabled the identification of patterns related to the organisational sector in which projects were implemented and the types of improvement objectives pursued. **Figure 6** summarises the most frequently

represented organisational sectors and the most frequent improvement foci. As shown in **Figure 6**, 40% of projects were developed in organisations operating within the industrial and manufacturing sector, followed by the food sector, which accounted for 31% of the projects. It is also notable that 11% of projects were carried out in mining organisations. This sectorial distribution aligns with Peru's main economic activities, which are strongly oriented towards extractive and processing industries. Smaller proportion of projects were observed in the services, finance and trade sectors (8%), healthcare (6%) and construction (4%). Regarding improvement focus, **Figure 6** shows that the most frequent objective—accounting for more than 50% of projects—was the reduction of nonconformities (53%). The second most frequent objective was waste elimination (26%). Finally, 21% of the projects focused on time reduction. These results suggest a strong emphasis on the systematic improvement of quality and the enhancement of operational efficiency, which are fundamental principles of the Lean Six Sigma methodology.

Overall, project teams sought to implement their projects successfully through measurable and quantifiable performance improvements, which were formally presented and justified during the programme's final assessment. This high level of practical application indicates were able to translate theoretical knowledge into concrete organisational outcomes, providing a key indicator of the effectiveness and success of a training programme.

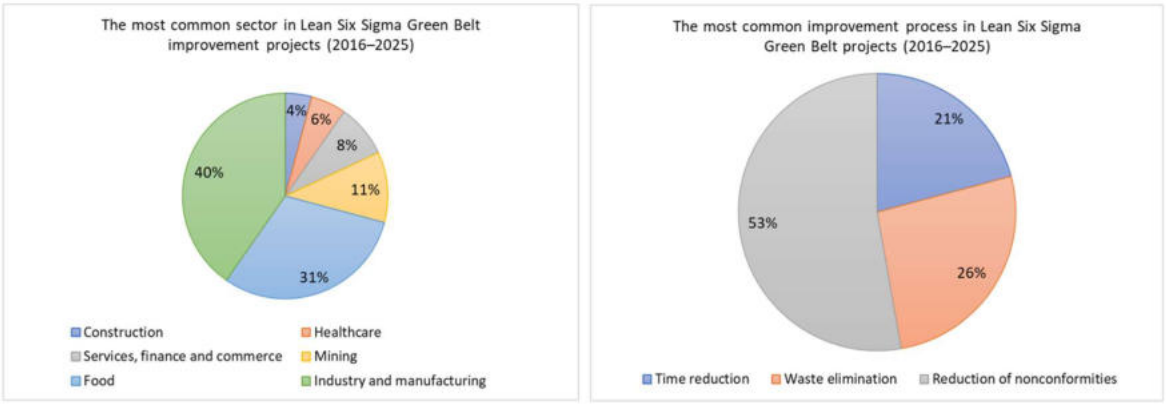


Fig. 6. The most common sector in Lean Six Sigma Green Belt improvement projects (2016–2025) and the most common improvement process in Lean Six Sigma Green Belt projects (2016–2025).

IV. Discussions and Implications

The findings confirm that the effectiveness of LSS training depends not only on the technical content delivered, but also on the training context and the participant's demographic profile. These effects vary by certification level. In both the LSSYB and LSSGB programmes, training modality showed a significant association with academic performance, consistent with previous studies highlighting the importance of flexibility and learner autonomy in continuous improvement programmes designed for adults and early-career professionals (Antony et al., 2017; Nadeau, 2017; Flores-Molina et al., 2025). The study also indicates that other relevant factors differ across training levels. Employment status is a significant variable at the Yellow Belt level, whereas age emerges as a key factor at Green Belt level. This suggests that the influence of individual characteristics is mediated by the level of practical complexity associated with each training programme (McShane-Vaughn, 2022; Munro et al., 2022; Paneerselvam et al., 2025). Moreover, the cluster analysis shows that the participant population is not structurally uniform, reinforcing the need for more adaptive and differentiated approaches (Hollingshed, 2022; Hutwelker & Ott, 2022). Finally, with respect to LSSGB projects, the findings support the argument that evaluating the effective transfer of learning into professional practice (via a final improvement project) is essential for assessing the maturity and practical value of LSS programmes (Antony et al., 2018; Pyzdek & Keller, 2018).

Several practical implications may be considered:

- *Pedagogical differentiation based on profiles within the same certification level*: identifying differences in participant´s educational background, professional experience and academic performance enables the design of teaching strategies and learning supports tailored to each learner profile. This approach enhances learning effectiveness by reducing performance gaps and ensuring instructional methods aligned with empirical evidence. For example, differentiated interactive online materials could be developed for participants who need to reinforce foundational concepts, alongside alternative materials for those with higher prior knowledge or experience.
- *Aligning professionals' work experience with the practical demands of each training programme*: it is advisable to adjust the complexity and applicability of practical exercises and projects according to participants' level of experience, which may enhance the transfer of learning to the workplace and increase the practical value of the certification in that environment. For less experienced professionals,

simulated or case-based projects may be more appropriate; by contrast, more experienced professionals can be encouraged to develop projects directly within their organisational contexts.

- *Strategic segmentation of LSS training into differentiated learning pathways*: characterising participant profiles by sector, professional role, speciality and employment status allows the design of specific learning trajectories within LSS training programmes. Such segmentation can improve programme and, consequently, enhance participants' academic performance. For instance, common or core sessions could be interleaved with parallel workshops tailored to specific roles (managers, supervisors, analysts, interns/trainees, among others).
- *Comprehensive evaluation of training impact*: expanding evaluation criteria to include real-world application, implemented improvements and competency development provides a more complete assessment of training effectiveness, facilitating continuous program improvement. For example, in addition to progress-based assessments and participation metrics, rubrics could be introduced to evaluate not only technical proficiency in applying the DMAIC methodology, but also transversal competencies such as critical thinking, leadership, effective communication and teamwork.
- *Prioritising sectors and processes for improvement projects*: Systematic analysis of project characteristics can help to identify sectors and processes with the greatest demand for improvement. This can support training provision aligned with market needs and the creation of strategic partnerships with organisations. For example, representatives from leading companies could be invited to present real organisational challenges, guiding participants toward relevant improvement problems, or to form sector-specific project teams to analyse and addressed targeted issues.

The results derived from the Peruvian case provide evidence from a context characterised by high sectoral diversity and varying levels of organisational maturity—conditions that are typical of emerging economies but remain underrepresented in the Lean Six Sigma literature. Unlike studies focusing on large, highly standardised organisations, this research shows how LSS training can be effectively implemented and can yield consistent outcomes in an environment where extractive industries, manufacturing, service sector, and organisations with heterogeneous resources coexist (OECD, 2025). While the findings are not intended to be generalised, the study demonstrates the adaptability of LSS principles and certification schemes to contexts with diverse organisational structures. This reinforces the relevance of LSS beyond highly industrialised economies (Makwana & Patange, 2021). In this sense, the Peruvian case provides a useful reference point for educational institutions and organisations in similar settings aiming to implement continuous improvement programmes that are closely aligned with their operational realities.

In conclusion, the findings indicate that academic performance is influenced by different factors depending on programme level: modality emerges as a common factor, while employment status and age group are programme-specific factors (LSSYB and LSSGB, respectively). The cluster analysis further reveals the presence of differentiated participant profiles within each programme, while project analysis evidenced effective transfer of learning towards improvements in real organisational processes. Collectively, these findings reinforce the importance of designing Lean Six Sigma training programmes that explicitly consider participant diversity and the that deliberately integrate learning activities with organisational practice.

One of the limitations identified in the research is the potential bias in the method of evaluation, given that the sample comes from a single institution. Another limitation is the lack of comprehensive information on the projects, as only their titles are recorded, which may lead to a bias in the assignment of the 'type of project improvement' categories.

This study highlights the need to move beyond standardised Lean Six Sigma training towards more adaptive and practice-oriented models. Specifically, training programmes should be deliberately designed to account for the diversity of participant profiles, including differences in experience, professional roles, and learning contexts, while ensuring a strong integration between instructional content and real organisational application, thereby enhancing both academic outcomes and practical impact on operational excellence.

REFERENCES

Antony, J., Snee, R. and Hoerl, R. (2017), "Lean Six Sigma: yesterday, today and tomorrow", International Journal of Quality & Reliability Management, Vol. 34 No. 7, pp. 1073–1093, doi: 10.1108/IJQRM-03-2016-0035.

Antony, J., Gupta, S., Sunder M., V. and Gijo, E.V. (2018), "Ten commandments of Lean Six Sigma: a practitioners' perspective", International Journal of Productivity and Performance Management, Vol. 67 No. 6, pp. 1033–1044, doi: 10.1108/IJPPM-07-2017-0170.

Buestan, M., Perez, C.C. and Rodríguez-Zurita, D. (2025), "Lean six sigma for health care: multiple case studies in Latin America", International Journal of Lean Six Sigma, Vol. 16 No. 1, pp. 172–196, doi: 10.1108/IJLSS-10-2023-0169.

Case, J.M. and Light, G. (2011), "Emerging Research Methodologies in Engineering Education Research", Journal of Engineering Education, Vol. 100 No. 1, pp. 186–210, doi: 10.1002/j.2168-9830.2011.tb00008.x.

DeRunt, B. and Meier, R. (2009), "Trainers' perceptions of the relative importance of the ten topics included in the American society for quality's Six Sigma Black Belt certification", Journal of Industrial Technology, Vol. 25 No. 3, pp. 1–13.

Flores-Molina, J.C., A. Puertas, M. and Yalta, E. (2025), "Lean Six Sigma Black Belt training for professionals: analysis of profiles, projects and academic performance", Proceedings of the International Conference on Industrial Engineering and Operations

Management, in 10th North American International Conference on Industrial Engineering and Operations Management, IEOM Society International, Orlando, USA, pp. 886-895, doi: 10.46254/NA10.20250256.

George, M.L. (2002), Lean Six Sigma: Combining Six Sigma Quality with Lean Speed, McGraw-Hill, New York.

Green, F.B. (2006), "Six Sigma and the green belt perspective: a study in five companies", International Journal of Six Sigma and Competitive Advantage, Vol. 2 No. 3, p. 291, doi: 10.1504/IJSSCA.2006.011119.

Hair, J.F. (2019), Multivariate Data Analysis, Eighth edition., Cengage, Andover, Hampshire.

Hollingshed, M. (2022), "Standardizing Six Sigma Green Belt training: identification of the most frequently used measure phase DMAIC tools", *International Journal of Lean Six Sigma*, Vol. 13 No. 2, pp. 276–294, doi: 10.1108/IJLSS-12-2020-0220.

Hutwelker, R. and Ott, H. (2022), "A Lean Six Sigma Certification MOOC: Balancing Standard Material and Individual Guidance to Help Participants Implement Real Projects", in Guralnick, D., Auer, M.E. and Poce, A. (Eds.), *Innovations in Learning and Technology for the Workplace and Higher Education*, Vol. 349, Springer International Publishing, Cham, pp. 122–132, doi: 10.1007/978-3-030-90677-1_12.

INEI. (2022), "Perú: estructura empresarial", available at: https://www.gob.pe/institucion/inei/informes-publicaciones/6330149-peru-estructura-empresarial-2022 (accessed 6 March 2026).

Kwak, Y.H. and Anbari, F.T. (2006), "Benefits, obstacles, and future of six sigma approach", Technovation, Vol. 26 No. 5–6, pp. 708–715, doi: 10.1016/j.technovation.2004.10.003.

Kuwaiti, A.A. (2020), "Use of Six Sigma and Lean Six Sigma Methodologies in Higher Education Institutions: A Systematic Review", Lean Six Sigma for Higher Education, WORLD SCIENTIFIC (EUROPE), pp. 1–20, doi: 10.1142/9781786348500_0001.

Louzada, P.D.S., Sigahi, T.F.A.C., Moraes, G.H.S.M.D., Rampasso, I.S., Anholon, R., Antony, J. and Cudney, E.A. (2023), "Critical analysis of Lean Six Sigma black belt certification courses offered in Brazil", The TQM Journal, Vol. 35 No. 7, pp. 1980–2002, doi: 10.1108/TQM-08-2022-0254.

Makwana, A.D. and Patange, G.S. (2021), "A methodical literature review on application of Lean & Six Sigma in various industries", Australian Journal of Mechanical Engineering, Vol. 19 No. 1, pp. 107–121, doi: 10.1080/14484846.2019.1585225.

McShane-Vaughn, M. (2023), The ASQ Certified Six Sigma Black Belt Handbook, ASQ Quality Press, La Vergne.

Munro, R.A., Ramu, G. and Zrymiak, D.J. (Eds.). (2022), The ASQ Certified Six Sigma Green Belt Handbook, Third Edition, Third., ASQExcellence, Milwaukee.

Nadeau, S. (2017), "Lean, Six Sigma and Lean Six Sigma in Higher Education: A Review of Experiences around the World", American Journal of Industrial and Business Management, Vol. 07 No. 05, pp. 591–603, doi: 10.4236/ajibm.2017.75044.

OECD. (2025), Estudios Económicos de la OCDE: Perú 2025, Vol. 2025, OECD Publishing, Paris, doi: 10.1787/626594d0-es.

Paneerselvam, N., Muhammad, N.A., Azhan, A.M. and Muhammad, N. (2025), "Analyzing critical success factors in Lean Six Sigma training", International Journal of Productivity and Performance Management, Vol. 74 No. 4, pp. 1400–1424, doi: 10.1108/IJPPM-11-2023-0627.

Pyzdek, T. and Keller, P.A. (2018), The Six Sigma Handbook, Fifth Edition., McGraw-Hill Education, New York.

Snee, R.D. (2010), "Lean Six Sigma – getting better all the time", International Journal of Lean Six Sigma, Vol. 1 No. 1, pp. 9–29, doi: 10.1108/20401461011033130.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Transformational Leadership and ISO/DIS 9001:2026: Are the Quality Managers transformational leaders?

Noémia Almeida [1], Cláudia Sousa Silva [1] and Magda Monteiro [1]

[1] Águeda School of Technology and Management, University of Aveiro, Aveiro, noemiaalmeida@ua.pt
GOVCOPP, Águeda School of Technology and Management, University of Aveiro, Aveiro, claudia.margarida@ua.pt; orcid:0000 – 0002 – 3837-2809

[2] Águeda School of Technology and Management & Center for Research and Development in Mathematics and Applications, University of Aveiro, Aveiro, Portugal; msvm@ua.pt orcid: 0000-0001-8585-4440~

***Abstract*— The research explores the connections between Transformational Leadership (TL) and ISO/DIS 9001:2026, within the broader context of evolving Quality Management Systems (QMS). A mixed-methods approach was adopted, combining qualitative content analysis of the draft ISO 9001:2026 requirements with a quantitative survey of 279 quality management professionals in Portugal, using the Global Transformational Leadership (GTL) scale. The qualitative findings discuss the alignment between emerging ISO 9001 requirements—such as ethical quality culture, leadership involvement, continuous improvement, and opportunity-based thinking—and core TL dimensions. Quantitative results indicate a high self-perception of transformational leadership (M = 4.323), which enhances empowerment, leading by example, and staff development behaviours. Significant differences were found across organizational variables, particularly position and company size. By combining qualitative and quantitative findings, this research highlights how leadership behaviours align with the new normative requirements and identifies key areas of convergence and potential gaps. This study contributes to theory by reinforcing the TL–QMS relationship and offers insights for leadership development in quality management.**

Keywords – Transformational Leadership, Quality Management Systems, ISO 9001.

I. Introduction

ISO 9001 is currently the most widely adopted quality management standard worldwide and the only certifiable standard within the ISO 9000 family (Sanchez-Lizarraga et al., 2020). According to the ISO Survey (2022), more than 1.2 million valid certificates are registered globally, reflecting the sustained growth in its adoption. Research on ISO 9001 certification has identified two main types of motivation for adoption: internal motivations, related to organizational improvement objectives, and external motivations, primarily associated with market pressures and customer requirements (Sampaio et al., 2009).

ISO 9001 is subject to periodic revisions to ensure its continued relevance in response to changes in the organizational environment. The need for a new revision was formally recognized in 2023 following extensive international consultation, leading to the process that will result in the publication of ISO 9001:2026 (SGS, 2025b). Limón-Romero et al. (2024) highlight that the upcoming revision is expected to address digital transformation and sustainability while adopting simplified language to improve consistency in requirement interpretation. The new version is expected to be published in the third or fourth quarter of 2026, with a three-year transition period for currently certified organizations. The revised standard will maintain the Harmonized Structure of the 2015 version, ensuring continuity and integration with other management systems (ISO, 2026; SGS, 2025a, 2025b).

The 2026 revision will introduce two interrelated normative requirements that reinforce leadership's role within organizations. Clause 5 will require top management to promote and demonstrate a culture of quality and ethical behavior, including guidance on how to evidence this culture in practice. Clause 7.3 expands the awareness requirements by requiring all employees to understand these values. This articulation highlights that it is not sufficient to declare such values; they must be genuinely internalized across the organization (APCER, 2025; ISO, 2026; SGS, 2025a).

The incorporation of dimensions such as quality culture, and ethics signals a clear trend toward increased demands on organizational leaders, making research on the role of transformational leadership in the effective implementation of ISO 9001 principles even more relevant. A key question is how these cultures become genuinely embedded among employees. Transformational leadership theory offers relevant insights, as inspirational motivation helps employees understand the purpose and values underlying their work, while idealized influence positions leaders as ethical role models, encouraging alignment with those values (Bass, 1990). Together, these dimensions create the conditions for the genuine awareness required by Clauses 5 and 7.3.

Furthermore, the transition to a new version of the standard represents, in itself, an organizational change, and the literature identifies transformational leadership as the most appropriate style for leading such transformations (Boonstra, 2026; Gang Wang et al., 2011; Islam et al., 2025; Karimi et al., 2023).

Accordingly, this study aims to contribute to the discussion on the importance of transformational leadership in the transition to ISO/DIS 9001:2026, as this process requires not only the adoption of new requirements but also new approaches that should lead to genuine organizational change, rather than mere compliance supported by evidence. Therefore, the way quality management professionals lead this transition emerges as a central focus of this research.

In this context, the research follows a mixed-methods approach — combining qualitative analysis with quantitative survey data to achieve the following main goals:

- To review existing literature on the intersection of Quality Management Systems (QMS) ISO 9001 and Transformational Leadership (TL).
- To analyze the new requirements and changes of the ISO/DIS 9001:2026 version, establishing connections with TL
- To diagnose how Quality Professionals self-perceive as Transformational Leaders, analyzing differences according to socio-demographic and organizational variables.

II. Literature Review

This section systematizes the main concepts and models related to transformational leadership within the context of quality management, analyzing the interconnections between Transformational Leadership and Quality Management Systems (QMS), with particular attention to the emerging requirements of ISO 9001 to be published in 2026.

A. *Transformational Leadership*

The conceptual origins of Transformational Leadership can be traced back to Burns (1978, as cited in Judge & Piccolo, 2004; van Beveren et al., 2017), who introduced the distinction between transactional and transformational leadership in the context of political leadership. According to Burns, transactional leadership is based on an exchange process between leaders and followers, in which rewards are offered in return for the achievement of agreed objectives, appealing to individual interests. In contrast, transformational leadership represents a qualitatively different relationship, in which leaders and followers elevate each other to higher levels of motivation and morality. The transformational leader provides a sense of purpose that goes beyond immediate goals, appealing to followers' values and sense of mission (Judge & Piccolo, 2004).

At its core, transformational leadership can be defined as a process of influence capable of mobilizing social systems, aiming to turn employees into agents of change for organizational development and progress (Burns, 1978, as cited in van Beveren et al., 2017).

The multidimensional model of Bass and Avolio constitutes the main conceptual framework for characterizing transformational leadership behaviors. In its consolidated form, the model proposes four interrelated dimensions, commonly referred to as the "4 I's": Idealized Influence, Inspirational Motivation, Intellectual Stimulation, and Individualized Consideration (Judge & Piccolo, 2004):

- Idealized Influence, also known as charisma, refers to the degree to which the leader behaves in admirable ways, leading followers to identify with them. Bass, (1990) described such leaders as possessing a sense of mission and vision, capable of instilling pride and earning respect and trust.
- Inspirational Motivation refers to the extent to which the leader articulates an inspiring vision. Leaders high in this dimension challenge followers with high expectations, communicate optimism about future goals, and give meaning to tasks (Judge & Piccolo, 2004).
- Intellectual Stimulation refers to the degree to which the leader challenges assumptions, takes risks, and encourages followers to contribute ideas. Such leaders promote rationality and careful problem-solving, encouraging questioning of the status quo and the development of innovative solutions (Bass, 1990).
- Individualized Consideration refers to the degree to which the leader attends to the needs of each follower, acting as a mentor and listening to individual concerns (Judge & Piccolo, 2004). These leaders recognize individual differences and support personal development (Bass, 1990).

Measurement of Transformational Leadership

Various instruments have been developed to assess transformational leadership. The most widely used include the Multifactor Leadership Questionnaire (MLQ; Avolio et al., 1995), the Conger-Kanungo Scale, the Leadership Practices Inventory (LPI; Kouzes & Posner, 1990), and the Transformational Leadership Questionnaire (TLQ; Alban-Metcalfe & Alimo-Metcalfe, 2000) (van Beveren et al., 2017). A common limitation of these instruments is their length, which makes them time-consuming and difficult to apply. Considering this limitation, Carless et al. (2000) proposed a brief, practical, and equally valid instrument: the Global Transformational Leadership Scale (GTL).

The GTL was developed based on seven transformational leadership behaviors identified in the literature: (1) communicating a clear and positive vision of the future; (2) treating employees as individuals, supporting and encouraging their development; (3) providing encouragement and recognition; (4) fostering trust, engagement, and cooperation among team members; (5) promoting creative thinking and challenging assumptions; (6) being clear about the values and practices one upholds; and (7) instilling pride and respect in others and inspiring through demonstrated competence. Each behavior was operationalized as a single item on a 5-point Likert-type response scale (1 = rarely or never; 5 = very often or always). The term "Global" underscores the intention to capture an integrative, unidimensional measure of transformational leadership, rather than a breakdown into subscales.(Carless et al., 2000).

The original validation was conducted with a sample of 1,432 subordinates who evaluated 695 branch managers at an Australian financial organization (Carless et al., 2000). Exploratory factor analysis revealed a single factor with an eigenvalue of 5.0, explaining 71% of the total variance, with factor loadings ranging from .78 to .88. Internal consistency was $\alpha = .93$. Convergent validity was established through significant correlations with the MLQ and the LPI (Carless et al., 2000). The scale's

unidimensional structure justifies using a single composite score as a measure of transformational leadership, making it simpler and more efficient in empirical research.

van Beveren et al., (2017) adapted and validated the GTL for the Portuguese context and language in a study conducted with 456 members of 70 work groups from 26 organizations across various sectors, located in Central and North-Central Portugal. The adaptation process followed rigorous procedures: translation by bilingual experts, review by a panel of researchers in organizational psychology, and back-translation into English to verify semantic equivalence.

The psychometric results confirmed the scale's unidimensional structure, with AFE explaining 70.23% of the variance (eigenvalue = 5.76; loadings .783–.872) and AFC showing good fit (CFI = .98; RMSEA = .097), high internal consistency (α= .93–.96), and stability (Tucker = .999), demonstrating theoretical validity through positive correlations with autonomy (r = .55) and group experience (r = .53), which supports its reliability and utility in research and organizational contexts (van Beveren et al., 2017).

The one-dimensionality of GTL has direct methodological implications for this research. The accumulated evidence, from the original study by Carless et al., (2000) to the Portuguese validation study by van Beveren et al., (2017), supports the view that transformational leadership can be treated as a single global factor when using this scale, and there is no basis for decomposing the score into subscales. This is the approach adopted in this study, consistent with the psychometric properties demonstrated for the Portuguese version of the instrument.

B. ISO 9001 Quality Management Systems

Quality management systems (QMS) have emerged as the operational framework that translates quality management principles into concrete processes, procedures, and organizational structures. A QMS can be understood as the set of activities and decisions within an organization aimed at producing and maintaining products or services at a desired level of quality, aligning operations with customer needs and, continuous improvement (Limón-Romero et al., 2024).

ISO 9001 is the international reference standard for quality management, published by the International Organization for Standardization, commonly referred to as ISO. The current version, ISO 9001:2015, specifies the requirements for a QMS and enables organizations to demonstrate their ability to provide products/services that meet customer requirements and to continuously improve (Sanchez-Lizarraga et al., 2020). The standard is based on the process approach, risk-based thinking, and the seven Principles of Quality Management: Customer focus; Leadership, Engagement of people, Process approach, Improvement, Evidence-based decision making, Relationship management (ISO, 2025) The process-based approach relies on the PDCA (Plan-Do-Check-Act) cycle to help organizations better plan their internal processes. Risk-based thinking enables the proactive identification of circumstances that could cause problems in the QMS, allowing for the implementation of preventive measures and capitalizing on opportunities (Limón-Romero et al., 2024; Sanchez-Lizarraga et al., 2020).

Since ISO 9001:2026 had not yet been published at the time this research was conducted, the anticipated changes are described based on prospective technical documentation (APCER, 2025; ISO, 2026; SGS, 2025a). According to this documentation, ISO 9001:2026 is expected to be published in the third or fourth quarter of 2026.

Among the most relevant changes, Clause 5 (Leadership and Commitment) stands out, with a new emphasis on promoting a culture of quality and ethical behavior, and guidance on how these can be demonstrated in practice (APCER, 2025; ISO, 2026; SGS, 2025a). Clause 6 introduces a more explicit distinction between risks and opportunities, moving the text on risk-based thinking to Annex A and adding the concept of opportunity-based thinking (APCER, 2025). Clause 10 (Improvement) should provide additional guidance aligned with the leadership expectations of Clause 5 (SGS, 2025a). The revision also incorporates, in Clause 4.1, the 2024 amendment requiring organizations to determine whether climate change is a relevant issue for their context (APCER, 2025).

The inclusion of dimensions such as a culture of quality and ethics in the ongoing review signals a clear trend toward placing greater demands on organizational leaders, making research on the role of transformational leadership in the effective implementation of the quality management principles inherent in the structure of an ISO 9001 Quality Management System even more relevant.

C. Transformational Leadership (TL) and Quality Management (QM)

The relationship between TL and QM has been the subject of growing empirical attention over the past few decades, as researchers have sought to identify which leadership styles are best suited to support the effective implementation of quality philosophies and systems. Based on a narrative review of the literature, Table I identifies some academic studies focused on the relationship between TL and QM, adding the main findings.

TABLE I. STUDIES ON TRANSFORMATIONAL LEADERSHIP AND QUALITY MANAGEMENT

Study scope	Authors / Year	Methods / Statistical Validation	Main Results
Greek companies (manufacturing and services)	Bouranta, (2021)	Structural Equation Modeling (SEM); β coefficients; correlations ($p < .001$; $p < .01$)	Positive and significant impact of TL on all TQM practices; high coefficients; differences across sectors
Pakistan's industrial sector	Ma et al., (2025)	Mediation and moderation analysis; significance tests	TL influences TQM practices; trust as a mediator; collectivism as a moderator; some relationships are not significant
Southern Taiwan Science Park	Chen et al., (2020)	Structural equation modeling; serial mediation (β; $p < .001$)	TQM influences TL, which influences sustainable

Study scope	Authors / Year	Methods / Statistical Validation	Main Results
			competitive advantage; mediating role of TL.
Bahrain Authority for Culture and Antiquities	Al Kohji et al., (2024)	Regression (R^2 = .732); β coefficients (p = .000)	Strong impact of TL on the implementation of TQM
Portugal (quality managers)	Silva et al., (2021)	Systematic review + descriptive analysis (means)	TL is the predominant style; emphasis on inspirational motivation; effective for change and sustainability
Industrial companies in Mexico	Limón-Romero et al., (2024)	Structural Equation Modeling (SEM)	Top management commitment as the main critical factor in the implementation of the ISO 9001 QMS
United Arab Emirates (448 participants)	Lasrado & Kassem, (2020)	SEM with mediation	TL influences organizational excellence via a culture of engagement

Bouranta, (2021) argues that the attributes of TL are compatible with a TQM environment; consequently, TL has been consistently identified as the leadership style best suited for the effective implementation of TQM practices (Dean & Bowen, 1994, cited in Bouranta, 2021), often through mediating variables such as trust and organizational culture, and with moderating effects such as collectivism (Ma et al., 2025). TL increases employee engagement, a vital condition for implementing the TQM philosophy (Bouranta, 2021). In some cases, TL not only supports QM practices but is also influenced by them, contributing to outcomes such as sustainable competitive advantage and organizational excellence (Chen et al., 2020; Silva et al., 2021).

III. RESEARCH METHODOLOGY

This study aims to highlight the importance of transformational leadership in the current context, which is characterized by new regulatory requirements for Quality Management Systems, to be published in 2026, following a research design outlined in Figure 1.

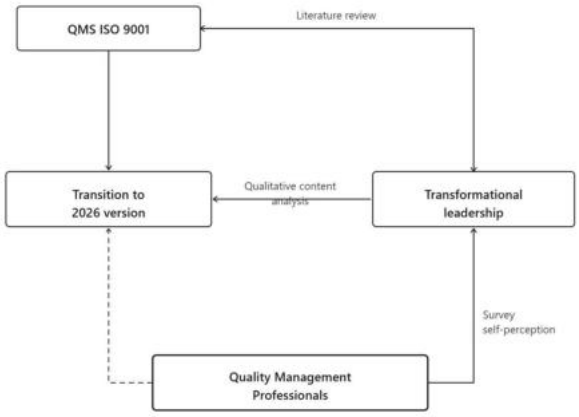


Fig. 1. Research Design

The study is grounded in a Literature Review that connects two central constructs: QMS ISO 9001 and Transformational Leadership. The bidirectional relationship between these two concepts suggests that literature treats them as mutually influential — leadership styles shape quality management practices, and quality systems, in turn, inform leadership demands. Then, two distinct data collection and analysis methods are employed:

a) Qualitative approach: The content analysis technique was applied to the two documents under review: ISO/DIS 9001:2026 and the GLT Scale developed by Carless et al. (2000). The main categories used in the analysis were based on the dimensions of the GLT Scale, which were subsequently linked to the requirements of the ISO/DIS 9001:2026 standard. The content analysis was conducted in two iterations. In the first iteration, one researcher analyzed the standard and proposed associations between the priorities of ISO 9001:2026 and transformational leadership (TL) behaviors. In the second iteration, these associations were reviewed and validated by another researcher.

b) Quantitative approach: An exploratory study was conducted, based on a survey of quality management professionals in Portugal. As the data collection instrument, the questionnaire developed by Carless et al., (2000) was adopted (a more detailed description is provided in Section II. A), incorporating sociodemographic variables. Participants were invited via the professional media platform to complete the online questionnaire on an electronic platform, to which they were given access via a link. Ethical considerations were fully ensured through the voluntary participation of respondents, who were previously

informed of the questionnaire's strictly anonymous nature, in accordance with the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679, 2016). Data collection followed a cross-sectional design, with a convenience sample of 279 valid responses used for analysis after data screening procedures. A sample size of over 200 was targeted to ensure the application of Confirmatory Factor Analysis (CFA) to models of moderate complexity since this value aligns with a median sample of N=200 with the median observed in comparable methodological studies within this field (Kline, 2016). This choice is further supported by simulation studies conducted by Wolf et al. (2013), which demonstrate that required sample sizes decrease as indicator reliability increases; specifically, when factor loadings are anticipated to be strong, an N in the range of 150 to 200 is sufficient to maintain high power and minimize bias.

Data analysis was performed using Microsoft Excel, IBM SPSS Statistics (version 30) and the package Lavaan of the R software for descriptive statistical analysis, reliability assessment and Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA)

Transformational leadership will be treated as a single variable in this study, with an overall transformational leadership score calculated from the CFA scores. This approach is consistent with the validation study of the Portuguese version of the GTL, which demonstrates that the scale has a unidimensional structure, with all items measuring the same factor. Since the scale was used to measure self-perception, exploratory factor analysis and confirmatory factor analysis were conducted. In the latter case, given the ordinal nature of the items (5-point Likert scale) and the asymmetry in the distribution of responses (with a ceiling effect), the Diagonally Weighted Least Squares (DWLS) estimator with corrections, based on polychoric correlation matrices, was used.

Thus, by using the overall GTL score as an independent variable, the present study maintains methodological consistency with the original validation and application process of the scale. To explore heterogeneity in TL self-perceptions across subpopulations, a comparative analysis of this factor will be conducted using socio-demographic and organizational variables (gender, professional category, and age group). Given the ordinal nature of the original scale and the potential lack of normality in the distribution of the transformed scores, non-parametric statistical methods were selected, as they do not require the stringent normality assumptions of parametric tests. Specifically, the Mann-Whitney U test will be employed for comparisons between two independent groups, while the Kruskal-Wallis test will be used for variables with three or more categories.

IV. Results and discussion

A. *Qualitative content analysis : Transformational Leadership and the draft of ISO 9001:2026*

The literature review presented in section II -C argues that transformational leadership is a robust and theoretically grounded predictor in quality management. This section proposes a forward-looking analysis of this relationship, examining the extent to which the behaviors that define transformational leadership, as operationalized by the GTL scale, can be specifically linked to the priorities that the ISO/DIS 9001:2026 seeks to formalize and reinforce.

Thus, supported by a qualitative approach through a content analysis of the potential changes highlighted in the ISO 9001:2026 standard, Table II structures the alignment of the analyzed normative priorities considering the seven behaviors that the GTL scale operationalizes. It should be noted that the GTL, developed by Carless et al., (2000) and validated for the Portuguese context by van Beveren et al., (2017), measures transformational leadership through seven specific behaviors:

(1) Vision: communicating a clear and positive vision of the future;
(2) Staff development—developing employees;
(3) Supportive Leadership—providing support and recognition;
(4) Empowerment—promoting autonomy and cooperation;
(5) Innovative Thinking—stimulating innovative thinking;
(6) Lead by Example—leading by example;
(7) Charisma—inspiring through the example of competence.

TABLE II. Alignement between the priorities of ISO 9001:2026 and TL behaviors (GTL)

Clause of ISO 9001:2026	Normative Priority	Corresponding TL Behavior (GTL)
5 e 7.3	Top management must promote and demonstrate a culture of quality and ethical behavior; employees must be made aware of these values	Lead by Example (6): the leader is clear about their values and acts in accordance with them; Vision (1): communicates a vision that guides employees toward the purpose of quality (7) Charisma: inspire through the example of competence.
10	Leadership must demonstrate active influence in continuous improvement cycles, explicitly aligned with the expectations of Clause 5	Innovative Thinking (5): encourages thinking about problems in new ways; Empowerment (4): fosters autonomy and cooperation among team members; (2): Staff development—developing employees;
4	Maintenance and reinforcement of consideration for climate change and sustainability within the organizational context	Vision (1): communicates a clear and positive vision of the future that transcends immediate operational objectives

Clause of ISO 9001:2026	Normative Priority	Corresponding TL Behavior (GTL)
6.1	Structured separation of risks and opportunities; explicit introduction of opportunity-based thinking	Vision (1): communicates a clear and positive vision of the future; Innovative Thinking (5): proactively guides the organization toward new approaches and opportunities

These findings indicate that the seven dimensions of TL, as operationalized by the Global Transformational Leadership (GTL) scale, correspond to key normative requirements, including the promotion of an ethical quality culture, active leadership involvement in continuous improvement, integration of sustainability concerns, and the adoption of opportunity-based thinking. Behaviors such as articulating a compelling vision, leading by example, fostering innovation, empowering employees, and supporting staff development appear particularly relevant in addressing these updated requirements.

B. *Quantitative methodological approach: Quality management professionals' self-perception*

Table III presents the descriptive statistics for the socio-demographic and professional variables of the sample. The sample characterization (n = 279) reveals a predominantly female profile, accounting for 59.9% of respondents (n = 167). Regarding the age structure, the sample presents an estimated mean age of 39.82 years (SD = 9.17), with a significant concentration in the intermediate age groups: 36.2% of participants are between 35 and 44 years old, followed closely by the 25–34 (28.7%) and 45–54 (28.3%) age brackets.

In terms of the professional profile, there is a balanced distribution among decision-making roles within the quality sector, with Quality Managers (27.2%) and Quality Directors (24%) representing more than half of the sample. The estimated mean experience in the current position is 7.51 years (SD = 5.80), with the vast majority of respondents (69.9%) having held their roles for a period between 1 and 10 years.

At the organizational level, participants are mostly integrated into the Industry and Manufacturing sector (80.6%), operating primarily within SMEs (56.3%). Geographically, the sample shows a strong concentration in the Central (48.7%) and Northern (43%) regions of Portugal, with a minor representation from the South and Islands (8.2%). In summary, the data delineates a mature respondent profile with consolidated experience in quality management roles, primarily IN the Industry and Manufacturing sectors of the country's central and northern regions.

TABLE III. SOCIODEMOGRAPHIC AND PROFESSIONAL CHARACTERISTICS OF THE SAMPLE (N = 279)

Variable	Categorys	n (%)
Sex	Female	167 (59.9%)
	Male	112 (40.1%)
Position	Quality Directori	67 (24%)
	Quality Managmer	76 (27.2%)
	Quality Engineer	45 (16.1%)
	Quality Technician	28 (10%)
	Other	63 (22.6%)
Enterprise size	SME	157 (56.3%)
	Large Enterprise	122 (43.7%)
Sector	Industry/manufacturing	225 (80.6%)
	Services	26 (9.3%)
	Other	28 (10.1%
Region	North	120 (43.0%)
	Center	136 (48.7%)
	South and islands	23 (8.2%)
Age Groups	Less than 25 years	7 (2.5%)
	25-34 years	80 (28.7%)
	35-44 years	101 (36.2%)
	45-54 years	79 (28.3%)
	More than 54 years	12 (4.3%)
Estimated mean (SD) age derived from the age groups and frequencies: 39.82 (9.17)		
Experience on the position	Less than 1 year	28 (10%)
	1-10 years	195 (69.9%)
	11-20 years	42 (15.1%)
	More than 20 years	14 (5.0%)
Estimated mean (SD)experience derived from the experience intervals and frequencies: 7.51 (5.80)		

Regarding the organizational context, respondents were asked to identify the management system certifications held by their respective companies (see Table IV). The results indicate a clear prevalence of ISO 9001, reported by 76.3% (n = 213) of the participants. Environmental and occupational health and safety standards also show a significant presence in the respondents' professional environments, with ISO 14001 and ISO 45001 being reported by 59.9% and 41.9% of the participants, respectively.

Aligned with the industrial profile of the sample, the automotive-specific IATF 16949 standard was identified by 22.6% of the professionals. Sector-specific standards, such as ISO 22000 (Food Safety) and ISO 13485 (Medical Devices), were reported

by a smaller subset of the sample (6.1% and 5.7%, respectively). Furthermore, 24.7% of the quality managers indicated that their organizations hold additional certifications not specifically listed in the survey.

TABLE III. ORGANIZATIONAL CERTIFICATIONS REPORTED BY QUALITY PROFESSIONALS (N = 279)

ISO 9001	IATF16949	ISO14001	ISO45001	ISO22000	ISO13485	Other Certifications
213 (76.3%)	63 (22.6%)	167 (59.9%)	117 (41.9%)	17 (6.1%)	16 (5.7%)	24,7%

Before calculating the factor scores for the Transformational Leadership construct required for the intended analysis, the items comprising the construct were evaluated for internal consistency. The results showed excellent reliability, with a Cronbach's alpha of 0.922. Since this scale is being used to assess self-perception—a different context from that of Van Beveren et al. (2017)—an Exploratory Factor Analysis (EFA) was initially conducted. Like the previous study, this led to an unifactorial solution explaining 62.92% of the total variance. The viability of the EFA was assessed using three indicators: the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy; Bartlett's test of sphericity; and the anti-image matrix. The results of the KMO (0.916), Bartlett's test ($\chi^2(21) = 1284.99$, $p < .001$), and the Anti-image Matrix (with diagonal values ranging between 0.89 and 0.96 and small off-diagonal values) confirmed the matrix's adequacy for factor analysis.

Additionally, to verify the unifactorial structure of the Transformational Leadership construct, a Confirmatory Factor Analysis (CFA) was performed. Model fit was evaluated using Scaled indices. Good fit criteria included Comparative Fit Index (CFI) and Tucker–Lewis index (TLI) values above 0.95, SRMR below 0.08, and RMSEA preferably below 0.08 (Byrne, 2010; Kline, 2016). Table V presents the scaled global fit measures, which are very similar in magnitude to those found by the authors who validated the scale and considered the model acceptable. Although the RMSEA value is slightly higher than the literature suggests (Byrne, 2010; Kline, 2016), the low SRMR (Standardized Root Mean Square Residual) value, in conjunction with the other indicators, supports the adequacy of the model. Regarding the evaluation of parameter estimates, the results were highly favorable. Figure 2 displays the retained factor structure, in which all estimated parameters are statistically significant, item factor loadings are well above 0.50, and residual variances are low. Using these loadings, the Transformational Leadership factor was created for the sample based on factor scores, which were subsequently rescaled to the original 1 to 5 scale.

TABLE IV. GOODNESS OF FIT MEASURES IN CFA

$\chi 2(14,N=279)$	χ^2/df	CFI	TLI	RMSEA	SRMR
59.612 (p<0.001)	4.258	0.990	0.985	0.108	0.035

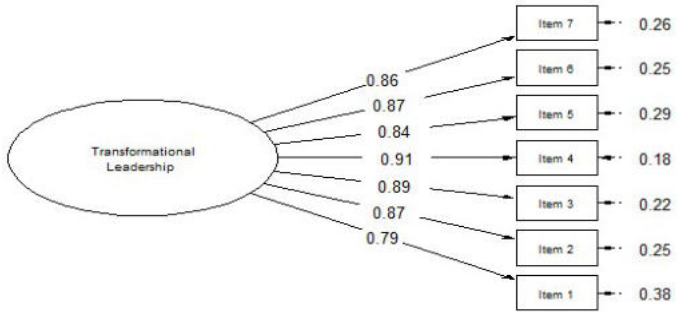


Fig. 2. GTL Factor Structure (standardized solution wherein values concerningerrors are proportions of variance not explained)

Figure 3 displays the sampling distribution of the Transformational Leadership factor scores, including the sample mean and standard deviation. Based on this sample, it can be concluded that professionals hold a highly positive self-perception of transformational leadership, with a mean score of 4.323 ± 0.059 at a 95% confidence level.

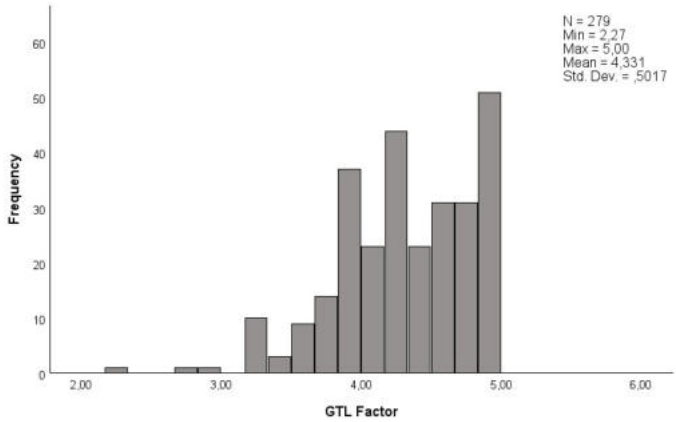


Fig. 3. Sampling distribnution of GTL factor

TABLE IV. GTL ITEM DESCRIPTIVE STATISTICS

Item n.º I...	Mean	SD	Sk	k
1.Vision: communicate a clear and positive vision of the future.	4.09	0.754	-0.651	0.651
2.Staff Development: treat staff as individuals, supporting and encouraging their development.	4.42	0.725	-1.402	2.933
3. Supportive Leadership: give encouragement and recognition to staff.	4.39	0.726	-1.092	1.267
4. Empowerment: foster trust, involvement and cooperation among team members.	4.48	0.704	-1.427	2.490
5.Innovative Thinking: encourage thinking about problems in new ways and questions assumptions.	4.40	0.722	-1.170	1.649
6. Lead by Example: am clear about my values and I practice what preach.	4.45	0.747	-1.529	2.969
7. Charisma: instill pride and respect in others and I inspire by being highly competent.	4.34	0.787	-1.164	1.505

The descriptive statistics reported in Table VI show that all items exhibit high mean scores, ranging from 4.09 to 4.48, confirming a consistently positive self-assessment across different aspects of transformational leadership. In particular, "Empowerment" (M = 4.48), "Lead by Example" (M = 4.45), and "Staff Development" (M = 4.42) stand out as the most highly rated dimensions, indicating strong emphasis on trust-building, role modeling, and individual support. Standard deviation values remain relatively low across items, suggesting limited dispersion and a high degree of agreement among respondents.

The item with the lowest mean scores in Table VI is "Vision" (M = 4.09), suggesting that, although still positively evaluated, this dimension is comparatively less emphasized within the overall transformational leadership profile. This pattern may indicate that leaders are somewhat less focused on articulating a compelling long-term vision. From a behavioral perspective, this could reflect a stronger orientation toward practical, relational leadership behaviors—such as supporting staff, fostering collaboration, and leading by example—rather than symbolic or visionary actions. In other words, these results could indicate that, while transformational leadership is clearly present, it may be expressed more through day-to-day interpersonal practices than through highly visible, inspirational, or future-oriented leadership behaviors.

C. Comparative Analysis of Transformational Leadership

The comparative analysis of Transformational Leadership (GTL) scores across socio-demographic and organizational context variables revealed statistically significant differences in two critical dimensions: position role and company size. Regarding company size, the Mann-Whitney U test revealed that managers in Large Companies have a significantly higher leadership self-perception (M = 4.460; SD = 0.439) than those in SMEs (M = 4.228; SD = 0.531; p < .001). The remaining variables, including gender (p = .218), sector (p = .564), region (p = .203), age group (p = .141), years in position (p = .657), and number of certifications (p = .086), did not show statistically significant differences. Nevertheless, a marginal trend was observed in the number of certifications, where organizations with 4 or more certifications tended to display higher leadership scores (M = 4.499).

TABLE V. MEAN (SD) OF THE GTL FACTOR BY GROUP, FOR EACH VARIABLE, AND THE P-VALUE OF THE NON-PARAMETRIC TEST PERFORMED

Variable	Categorys	GTL M (SD)	Non-parametric test
Sex	Female	167 (59.9%)	U M-W
	Male	112 (40.1%)	p=0.218
Position	Quality Director/ Manager	4.383 (0.496)	K-W
	Quality Engineer	4.197 (0.455)	P=0.046
	Quality Technician	4.247 (0.455)	
	Other	4.344 (0.488)	
Enterprise size	SME	4.228 (0.531)	U M-W
	Large Enterprise	4.460 (0.439)	P<0.001
Sector	Industry/manufacturing	4.315 (0.505)	K-W

	Services Other	4.360 (0.549) 4.427 (0.433)	P=0.564
Region	North Center South and islands	4.311 (0.533) 4.315 (0.483) 4.485(0.425)	K-W P=0.203
Age Groups	Less than 34 years 35-44 years More than 44 years	4.269 (0.502) 4.330 (0.508) 4.386 (0.508)	K-W P=0.141
Experience on the position	Less than 1 year 1-10 years More than 10 yeas	4.292 (0.422) 4.332 (0.528) 4.345 (0.449)	K-W P=0.657
N.º certifications	0-1 2-3 4 ou +	4.265 (0.561) 4.334 (0.461) 4.499(0.452)	K-W P=0.086

The findings suggest that transformational leadership is consistently perceived across most socio-demographic groups, with no statistically significant differences observed for gender, age, sector, region, or professional experience. This reinforces the idea that transformational leadership behaviors—such as empowerment, support, and leading by example—are broadly shared and not strongly conditioned by individual background characteristics. The absence of statistically significant differences in transformational leadership across most socio-demographic variables is consistent with prior research, which suggests that this leadership style is relatively universal and not strongly conditioned by individual characteristics (Attaallah, 2024; Chandaria et al., 2025; Kim & Shin, 2017).

However, significant differences emerge across organizational context, particularly position, role, and company size. Managers, especially Quality Directors, report higher levels of transformational leadership compared to engineers, and professionals in large organizations exhibit stronger leadership self-perceptions than those in SMEs. These results suggest that structural factors, such as hierarchical responsibility and resource availability, may play a more decisive role in shaping leadership behaviors than personal demographics, according to previous results by Khan et al. (2009) and Nguyen et al. (2023).

To strengthen the integration between the qualitative and quantitative phases, Table VIII was consequently reframed as a triangulation matrix that consolidates the conceptual relationships emerging priorities of ISO/DIS 9001:2026 with Transformational Leadership (TL) and the Quality professional's TL self-perception to identify key areas of convergence and potential gaps. So, the sequential mixed-methods logic adopted in the study examine whether the leadership behaviours theoretically associated with the revised ISO requirements were reflected in the perceptions of quality professionals.

TABLE VI. DISCUSSION FROM QUALITATIVE AND QUANTITATIVE RESULTS

ISO 9001:2026 Priorities (Qualitative)	TL Behaviors	Quantitative Evidence	Integrated Insight
Ethical quality culture (Clauses 5 & 7.3)	Lead by Example; Vision; Charisma	High mean: Lead by Example (M = 4.45)	Ethical leadership is strongly internalized, supporting authentic quality culture implementation
Continuous improvement (Clause 10)	Innovative Thinking; Empowerment; Staff Development	High scores: Empowerment (M = 4.48); Staff Development (M = 4.42)	Indicates a potential gap in strategic foresight and long-term orientation
Opportunity-based thinking	Vision; Innovative Thinking	Lower score in Vision (M = 4.09)	Vision (1): communicates a clear and positive vision of the future that goes beyond immediate operational goals
Sustainability and organizational context (Clause 4.1 – climate change relevance)	Vision; Charisma; Innovative Thinking	Strong overall TL (M = 4.323) but weaker Vision dimension	TL supports sustainability integration, but may require stronger strategic vision to address long-term environmental challenges
Leadership in organizational change (ISO transition)	All TL dimensions	High global TL + low dispersion	Professionals are well-positioned to lead ISO 9001:2026 transition processes
Culture internalization and awareness	Lead by Example and Staff development	High scores: Lead by Example (M = 4.45); Staff Development (M = 4.42)	Consistent leadership behaviors facilitate internalization of quality values
Organizational and structural factors	---	Significant differences: position ($p = .046$); company size ($p < .001$)	Leadership expression varies with hierarchy and organizational scale, affecting implementation capacity

V. CONCLUSION

This study addressed the role of transformational leadership in supporting organizations during the transition to ISO/DIS 9001:2026. The literature review examined the intersection between Quality Management Systems (QMS), particularly ISO 9001, and Transformational Leadership (TL), highlighting their bidirectional relationship. It showed that TL supports the implementation and continuous improvement of QMS, while quality practices also shape leadership behaviors. Moreover, this investigation identified a set of relevant studies that can guide future research and support quality management professionals in deepening their understanding of the relationship between transformational leadership and quality management.

The integration of qualitative and quantitative findings provides a coherent response to the study objectives. The qualitative analysis demonstrated that the emerging requirements of ISO/DIS 9001:2026—particularly the emphasis on ethical quality culture, leadership involvement, continuous improvement, and opportunity-based thinking—are aligned with the behavioral dimensions of Transformational Leadership (TL), as operationalized by the GTL scale. This alignment suggests that the transition to the new standard is not merely technical but fundamentally leadership driven.

The quantitative results reinforce this perspective by showing that quality professionals report a high self-perception of transformational leadership (M = 4.323), supported by strong psychometric properties (α = 0.922; explained variance = 62.92%). In particular, behaviors such as empowerment, leading by example, and staff development emerged as the most prominent dimensions, indicating a leadership profile focused on trust, support, and operational engagement. This empirical evidence supports the qualitative argument that transformational leadership behaviors are already embedded in quality management practice and are well-suited to meet the new ISO requirements.

Furthermore, the comparative analysis revealed that self-perceptions of transformational leadership vary significantly across organizational contexts, particularly by position and company size, with higher scores observed among quality directors/managers and professionals in large organizations. These findings suggest that while TL is broadly present, its intensity and expression are influenced by structural and organizational conditions. This is particularly relevant in the context of ISO/DIS 9001:2026, where leadership responsibility is expected to be more explicitly demonstrated across hierarchical level.

Despite its contributions, the study is limited by its reliance on self-reported data. Future research should examine how the self-reported TL behaviors of participating professionals contribute to the consolidation of the quality management principles advocated in quality management systems, thereby reinforcing the culture of quality management. It is also important to supplement the assessment of TL behaviors from the perspective of subordinates.

Acknowledgment

This work is partially supported by CIDMA (https://ror.org/05pm2mw36) under the Portuguese Foundation for Science and Technology (FCT, https://ror.org/ 00snfqn58), Grants UID/04106/2025 (https://doi.org/10.54499/UID/04106/2025) and UID/PRR/04106/2025.

This work was partially financially supported by UID/04058/2025 - Research Unit on Governance, Competitiveness and Public Policies, financed by national funds through FCT - Foundation for Science and Technology

References

Al Kohji, A. J., Al Deeb, H. M., Al Thawadi, M. A. A., and Shatat, A. (2024)," Impact of Transformational Leadership on the Application of Total Quality Management". *2024 International Conference on Decision Aid Sciences and Applications (DASA)*, 1–4. https://doi.org/10.1109/DASA63652.2024.10836495

APCER. (2025), "Atualizações da ISO 9001 e ISO 14001". *https://Apcergroup.Com/Pt/Noticias-e-Destaques/4912/Atualizacoes-Da-Iso-9001-e-Iso-14001-o-Que-Esperar*.

Attaallah, H. (2024), "Transformational and Transactional Leadership: From the Gender Lens", *Proceedings of the International Conference on Research in Business, Management and Finance*, Vol. *1*No.1,pp. 1–16. https://doi.org/10.33422/icrbmf.v1i1.514

Bass, B. M. (1990), "From transactional to transformational leadership: Learning to share the vision", *Organizational Dynamics*, Vol. *18*, No.3, pp. 19–31. https://doi.org/10.1016/0090-2616(90)90061-S

Boonstra, J. (2026), "Navigating Transformation: Review and Research Agenda", *Journal of Change Management*, Vol. 26, No.1, pp. 8–25. https://doi.org/10.1080/14697017.2026.2622508

Bouranta, N. (2021), "Does transformational leadership influence TQM practices? A comparison analysis between manufacturing and service firms", *The TQM Journal*, Vol. *33*, No.3, pp.706–728. https://doi.org/10.1108/TQM-12-2019-0296

Byrne, B. M. (2010), "*Structural equation modeling with Amos: Basic concepts, applications, and programming*", Taylor and Francis Group., Ed.; 2nd ed.

Carless, S. A., Wearing, A. J., & Mann, L. (2000), "A Short Measure of Transformational Leadership", *Journal of Business and Psychology*, Vol.*14*, No.3, pp. 389–405. https://doi.org/10.1023/A:1022991115523

Chandaria, M. B. N., Rajdev, A., and Bhuva, K. K. (2025), "Analysing Leadership Perception: The Role of Demographic and Professional Factors", *Advances in Consumer Research*, Vol. 2, pp. 4730–4738.

Chen, R., Lee, Y.-D., and Wang, C.-H. (2020, "Total quality management and sustainable competitive advantage: serial mediation of transformational leadership and executive ability", *Total Quality Management & Business Excellence*, Vol. *31, No.*5–6, pp. 451–468. https://doi.org/10.1080/14783363.2018.1476132

Gang Wang, Oh, I.-S., Courtright, S. H., and Colbert, A. E. (2011), "Transformational Leadership and Performance Across Criteria and Levels: A Meta-Analytic Review of 25 Years of Research", *Group & Organization Management*, Vol. 36, No. 2, pp. 223–270. https://doi.org/10.1177/1059601111401017

Islam, M. R., Prieto, L., and Talukder, M. F. (2025), "From Influence to Impact: How Transformational Leadership Shapes Employee Behavior Through Psychological Activation", *Administrative Sciences*, Vol.*15, No.*9, pp. 344. https://doi.org/10.3390/admsci15090344

International Organization for Standardization. (2026), " *ISO/DIS 9001 - Quality management systems: requirements*", Geneva: International Organization for Standardization

Judge, T. A., and Piccolo, R. F. (2004), "Transformational and Transactional Leadership: A Meta-Analytic Test of Their Relative Validity", *Journal of Applied Psychology*, Vol. *89*, No. *5*, pp. *755–768*. https://doi.org/10.1037/0021-9010.89.5.755

Karimi, S., Ahmadi Malek, F., Yaghoubi Farani, A., and Liobikienė, G. (2023), "The Role of Transformational Leadership in Developing Innovative Work Behaviors: The Mediating Role of Employees' Psychological Capital", *Sustainability*, Vol. *15, No.* 2, pp.1267. https://doi.org/10.3390/su15021267

Kim, S., and Shin, M. (2017), "The effectiveness of transformational leadership on empowerment", *Cross Cultural & Strategic Management*, Vol. *24*, No.2, pp. 271–287. https://doi.org/10.1108/CCSM-03-2016-0075

Kline, R. B. (2016), "*Principles and practice of structural equation modeling*",Guilford publications, Ed.

Lasrado, F., and Kassem, R. (2020), "Let's get everyone involved! The effects of transformational leadership and organizational culture on organizational excellence", *International Journal of Quality & Reliability Management*, Vol. *38, No.* 1, pp. 169–194. https://doi.org/10.1108/IJQRM-11-2019-0349

Limón-Romero, J., García-Alcaraz, J. L., Alberto Sanchez-Lizarraga, M., Gastelum-Acosta, C., Báez-López, Y., and Tlapa, D. (2024), "Implementing a Quality Management System Based on ISO 9001:2015 Standard: Modeling the Enablers' Relationships", *IEEE Access*, Vol. *12*, 95174–195187. https://doi.org/10.1109/ACCESS.2024.3520892

Ma, X., Waqas, A., and Malik, J. T. (2025), "Transformational leadership and total quality management practices: Moderating role of collectivism". *Acta Psychologica*, Vol. *260*, 105489. https://doi.org/10.1016/j.actpsy.2025.105489

Sampaio, P., Saraiva, P., and Guimarães Rodrigues, A. (2009), "ISO 9001 certification research: questions, answers and approaches", *International Journal of Quality & Reliability Management*, Vol. *26*, No.1, pp. 38–58. https://doi.org/10.1108/02656710910924161

Sanchez-Lizarraga, M. A., Limon-Romero, J., Tlapa, D., Baez-Lopez, Y., Puente, C., Puerta-Sierra, L., and Ontiveros, S. (2020),"ISO 9001 Standard: Developing and Validating a Survey Instrument", *IEEE Access*, Vol. *8*, 190677–190688. https://doi.org/10.1109/ACCESS.2020.3029744

SGS. (2025a), "*ISO 9001:2026—Anticipated changes*", Available at https://www.sgs.com/-/media/sgscorp/documents/corporate/brochures/sgs-ba-iso-9001-2026-anticipated-changes-en.cdn.en.pdf (Accessed: March 2026)

SGS. (2025b), "*The significance of ISO 9001 and why ISO 9001:2026?",* Available at *https://www.sgs.com/en/whitepapers/the-significance-of-iso-9001-and-why-iso-9001-2026-form* ((Accessed: March 2026)

Silva, C. S., Magano, J., Matos, A., and Nogueira, T. (2021), "Sustainable Quality Management Systems in the Current Paradigm: The Role of Leadership. *Sustainability*", Vol. *13, No.*4, pp. 2056. https://doi.org/10.3390/su13042056

van Beveren, P., Dórdio Dimas, I., Renato Lourenço, P., and Rebelo, T. (2017), "Psychometric properties of the Portuguese version of the Global Transformational Leadership (GTL) scale", *Revista de Psicología Del Trabajo y de Las Organizaciones*, Vol. *33*. No.2, 109–114. https://doi.org/10.1016/j.rpto.2017.02.004

Wolf, E. J., Mentzel, K. M., Miller, M. W., Levy, K. N., and Arieti, M. (2013, "Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*," Vol. 73, No.6, pp. 913–934. https://doi.org/10.1177/0013164413495237

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

The impact of lithium mining in Portugal: chemometric studies

Inês Timóteo [1], Susana Jesus [2, 3, 4], Marta Martins [3,4], Inês Coelho [2] and Ana Matos [5]

[1] Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal; i.timoteo@campus.fct.unl.pt
[2] Department of Food and Nutrition, National Institute of Health Doutor Ricardo Jorge (INSA, IP), Lisbon, Portugal
[3] Departamento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (NOVA FCT), 2829-516 Caparica, Portugal.
[4] Marine and Environmental Sciences Centre (MARE) & Aquatic Research Network Associated Laboratory (ARNET), NOVA FCT, 2829-516 Caparica, Portugal
[5] UNIDEMI – Research and Development Unit for Mechanical and Industrial Engineering, Department of Mechanical & Industrial Engineering, NOVA School of Science and Technology, Caparica, Portugal; Corresponding author: asvm@fct.unl.pt

***Abstract*— This study characterized the soil–water–plant system in agricultural contexts of lithium-bearing regions, with particular focus on the Guarda and Vila Real areas. An integrated statistical approach combining correlation analysis and principal component analysis was applied separately to soils, irrigation waters, and agricultural crops with contrasting physiological characteristics.**

The results indicate that crop responses vary according to the edible plant organ. In cabbage, significant associations between lithium concentrations in leaves and those in irrigation water were observed, suggesting that this compartment can play a significant role in lithium variability. In contrast, potato tubers exhibited stronger physiological regulation, limiting lithium transfer to the edible part and showing mainly indirect effects associated with nutrient redistribution rather than direct lithium accumulation.

Across all matrices, multivariate analyses did not reveal well-defined clusters or systematic separation of samples by proximity to current mining areas or regional origin. The continuous distribution of samples in multivariate space and the absence of dominant principal component analysis components suggest natural geochemical background variability and crop-specific physiological controls within the studied system.

Keywords – Lithium mining, ICP-MS, Soil-water-plant system, Multivariate analysis

I. Introduction

The global transition towards green energy and electric mobility has led to an exponential increase in demand for lithium (Li) in recent decades, driven by the need for lithium-ion batteries used in electric vehicles and energy storage systems (Chaves et al., 2021; Gbondo et al., 2025; Sadik-Zada et al., 2023). As a result, Li production has seen a significant growth, transforming this mineral into the 'white gold' of the fossil-fuel-free economy (Eniowo et al., 2026; Sadik-Zada et al., 2023). Recognising the economic importance of this resource and the high risk of long-term supply shortages, the European Union has classified Li as a critical raw material (Dunlap & Riquito, 2023; Jesus et al., 2024). In this global context, Portugal holds a prominent position, as it has the largest estimated hard-rock Li reserves in Europe. The country is already one of the world's ten largest Li producers. However, historically and until now, Portuguese Li mining has been exclusively for the ceramics and glass industries. Regions in the center and northern Portugal, have become the epicenter of this new focus, hosting projects for the construction of what could be the largest open-pit Li mines in Western Europe (Chaves et al., 2021; Dunlap & Riquito, 2023).

However, the expansion of open-pit Li mining raises significant environmental, social and public health concerns (Chaves et al., 2021; Dunlap & Riquito, 2023; Jesus et al., 2024). The extraction and processing of the mineral can have a severe impact on local ecosystems, leading to the release of Li and other elements into water, soil, and air (Bollaert et al., 2025). While Li is the primary element of concern due to its direct association with mining activities, other potentially toxic elements may also be mobilised and contribute to environmental contamination. Due to the high mobility of Li in aquatic and terrestrial systems, there is a risk of contamination of surface and groundwater, as well as its entry into the food chain via agricultural soils and irrigation water. In addition, the presence of other trace elements in these systems may further influence environmental and food safety. Since the chronic toxicity of Li and its safe limits for ecosystems and continuous human ingestion are not yet fully understood, monitoring its environmental impact is crucial (Bolan et al., 2021).

The aim of this article is to assess the potential effects associated with the presence of Li in Portugal through chemometric studies, covering different geographical areas and exploitation. Through multielemental analyses applied to environmental and agri-food samples, this study encompasses three Li mining areas in Portugal: the Barroso mine in Vila Real, and the Castanho and Alvarrões mines located in the Guarda district. The application of chemometric tools will enable the mapping of the current distribution patterns and relationships between matrices, as well as the assessment of potential differences associated with proximity to mining operations. This will contribute to a better understanding of the dynamics of these elements in the environment and provide scientific support for food safety monitoring and the ongoing risk assessment of exposed populations.

II. Materials and Methods

A. Study Area

The sampling strategy was designed to assess environmental and food matrices in areas influenced by Li mining activities located in central and northern Portugal.

The Barroso mining concession includes the parishes of Dornelas and Covas do Barroso and several nearby villages. Mining operations began in 2006, initially involving small open pits targeting lithium-bearing pegmatites. Current development plans foresee an expansion of the concession, including new open-pit operations planned in the vicinity of Covas do Barroso, Romaínho, and Muro. The other study sites are located in the Guarda region, in central-eastern Portugal, which is characterized by granitic formations and lithium-rich aplite-pegmatite veins. This area includes the Seixo Amarelo - Gonçalo vein field. Two open-pit mining concessions were selected: the Castanho mine, located in the Gonçalo district, and the Alvarrões mine, situated between Seixo Amarelo and Gonçalo. Both mines extract feldspar, lithium-bearing mica, and quartz, currently mainly for ceramic applications.

B. Sampling Plan and Collection

The sampling focused on subsistence farming areas surrounding Li mining sites in the Guarda and Barroso regions, considering proximity to current and projected mining activities. In the Guarda region, populations near the Castanho and Alvarrões mines were selected using a distance-based approach. For each mine, a near and a faraway site were defined. Farms located within 1.5 km of the mines were classified as near-mine sites (S1 for Castanho and S3 for Alvarrões), whereas farms located between 2 and 8.5 km were considered distant sites (S2 and S4). These locations correspond to the closest inhabited areas surrounding the mines. In the Barroso region, sampling targeted villages closest to the mining concession and areas expected to be influenced by its planned expansion. Samples were collected from subsistence farms grouped into two sites based on distance and relative position to current and future open-pit operations. Near-mine site (S5) includes eastern villages (Covas do Barroso, Romaínho, and Muro), located closer to the projected future pits, whereas faraway sites (S6) comprise northwest villages (Dornelas, Antigo, and Espertina), situated farther away, as illustrated in Table I. For each site, samples were collected from 12 subsistence farms, resulting in a total of 24 farms sampled per mining area.

TABLE I. Distribution of samples (plots) by location and proximity intervals to the mining areas.

	Castanho Mine		Alvarrões Mine		Barroso Mine	
Location	**Farms**	**Sites (km)**	**Farms**	**Sites (km)**	**Farms**	**Sites (km)**
Near	101	<0.5	301	<0.5	501	2.5-3.0
(S1/S3/S5)	102	<0.5	302	<1.5	502	2.5-3.0
	103	<1.5	303	<1.5	503	2.5-3.0
	104	<1.5	304	<1.5	504	2.5-3.0
	105	<1.5	305	<1.5	505	2.5-3.0
	106	<0.5	306	<0.5	506	2.5-3.0
	107	<0.5	307	<1.5	507	1.8-2.5
	108	<1.5	308	<1.5	508	1.8-2.5
	109	<1.5	309	<1.5	509	1.8-2.5
	110	<0.5	310	<1.5	510	1.8-2.5
	111	<1.5	311	<1.5	511	1.8-2.5
	112	<0.5	312	<1.5	512	1.8-2.5
Far	201	2.0-8.5	401	2.0-8.5	602	3.0-4.0
(S2/S4/S6)	202	2.0-8.5	402	2.0-8.5	603	3.0-4.0
	203	2.0-8.5	403	2.0-8.5	604	3.0-4.0
	204	2.0-8.5	404	2.0-8.5	605	3.0-4.0
	205	2.0-8.5	405	2.0-8.5	606	3.0-4.0
	206	2.0-8.5	406	2.0-8.5	607	3.0-4.0
	207	2.0-8.5	407	2.0-8.5	608	2.5-3.0
	208	2.0-8.5	408	2.0-8.5	609	3.0-4.0
	209	2.0-8.5	409	2.0-8.5	610	2.5-3.0
	210	2.0-8.5	410	2.0-8.5	611	2.5-3.0
	211	2.0-8.5	411	2.0-8.5	612	2.5-3.0
	212	2.0-8.5	412	2.0-8.5	602	3.0-4.0

The sampling methodology followed a total diet studies approach aiming to ensure representativeness of the local population's diet. Two horticultural products widely cultivated and consumed in portuguese rural regions were selected: cabbage and potato. At each farm, three samples of each crop were collected and analysed as pool sample. Corresponding soil samples were collected at a depth of 20 cm, as well as irrigation water samples. All samples were transported under appropriate conditions for analysis.

C. Reagents and Chemicals

All reagents used were of high analytical grade, and samples were handled in a cleanroom environment to prevent contamination. Ultrapure water was obtained through a Milli-Q Plus system (Millipore). Nitric acid (HNO_3, 65% v/v; Merck) was previously purified to ultrapure grade using an acid distillation system (Milestone, Unicam), and 30% hydrogen peroxide (H_2O_2) was employed.

Multielemental standard calibration solutions were prepared from high-purity single-element standards at a concentration of 1000 mg/L (SCP Science / Merck) in a 2% (v/v) HNO_3 solution. For Inductively Coupled Plasma Mass Spectrometry (ICP-MS) analysis, Germanium, Yttrium, and Indium at 1000 mg/L (Merck) were used as internal standards.

D. Sample Preparation and Multi-elemental Analysis

The edible parts of the crops (cabbage leaves and peeled potatoes) were washed with tap water, dried in an oven at 40°C until reaching a constant weight, and subsequently ground and homogenized in a mill.

Cabbage samples were digested using a closed-vessel microwave system (Ethos 1, Milestone) with the addition of 3 mL of ultrapure water, 4 mL of HNO_3, and 1 mL of H_2O_2, following a progressive heating program up to 210°C. For the potatoes, a 48-well heating block (DigiPREP, SCP Science) was used at 85°C for 180 min, using 7 mL of HNO_3 and 1 mL of H_2O_2.

Irrigation water samples were filtered (0.45 µm) and acidified with 2% (v/v) ultrapure HNO_3 prior to analysis. Soil samples were air-dried, sieved to 2 mm, and subjected to aqua regia extraction.

Elemental concentrations were determined using ICP-MS (Thermo X Series II, Thermo Fisher Scientific) for trace elements (As, B, Be, Cd, Co, Cu, Li, Mn, Ni, Pb, Se, Sn, Sr, Tl, Zn) and Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES) for major elements (Ca, Fe, K, Mg, Na, P).

E. Quality assurance and quality control

Analytical determinations were performed in compliance with the quality assurance requirements of the NP EN ISO/IEC 17025 standard (ISO/IEC, 2017) for all matrices. Analytical quality control was ensured in each digestion cycle by including at least one procedural blank to assess potential contamination and one spiked sample or certified reference material (CRM) to evaluate analytical accuracy. Additionally, the laboratory regularly participated in proficiency testing (PT) schemes (FAPAS® and BIPEA) and consistently obtained satisfactory results.

F. Statistical analysis

Data processing and statistical analysis were performed using Microsoft Excel and Statistica software (version 14.1, StatSoft Inc., Portugal).

Values below the limit of detection (LOD) were replaced by $LOD/\sqrt{2}$ to minimize bias in subsequent statistical analyses. To ensure comparability between variables with different units and magnitudes, the data were standardized using Z-score transformation (mean equal to zero and standard deviation equal to one). Descriptive statistics, including median, quartiles, minimum, and maximum values, were calculated for each variable and matrix.

Correlation analysis was carried out using Pearson's correlation coefficient, allowing the assessment of relationships between elements and between different environmental matrices. The statistical significance of correlations was evaluated using the corresponding t-test. Differences between sampling groups, namely in relation to proximity to mining areas, were assessed using two-way analysis of variance (two-way ANOVA), considering a significance level of $p < 0.05$. Additionally, Principal Component Analysis (PCA) was applied to explore patterns in the dataset and identify possible associations between variables, using orthogonal varimax rotation to facilitate component interpretation.

III. Results and Discussion

This section presents the results for the soil–water–plant system in lithium-bearing agricultural regions, with particular focus on the Guarda and Vila Real areas. Considering the complexity of environmental systems, the results are interpreted using an integrated statistical framework combining correlation analysis and PCA. Analyses were conducted separately for soils, irrigation waters, and crops with contrasting physiological characteristics, allowing both matrix-specific responses and system interactions to be evaluated.

A. Cabbages

The integrated analysis of correlation results and PCA highlights the high complexity of the soil–water–plant system and demonstrates that Li behaviour in cabbages results from the interaction of multiple environmental and physiological processes. This framework is particularly relevant in studies conducted in regions with potential mining interest.

Correlation analysis revealed statistically significant associations between Li concentrations in cabbage leaves and the different environmental matrices, although with contrasting magnitudes. A very strong correlation was observed

between Li concentrations in the leaves and in irrigation water (ρ = 0.92), as well as strong associations with Na in irrigation water (ρ = 0.89) and in the leaves (ρ = 0.82). These results suggest that irrigation water chemistry can be a relevant factor influencing the observed variability in Li concentrations in cabbage, given the high mobility of this element in aqueous solution. The observed correlations are also consistent with the similar geochemical behaviour of Li^+ and Na^+ ions, which share comparable physicochemical properties and uptake pathways in plants. The associations of Li and Na in irrigation water and in leaf tissues support the interpretation that both elements may be transported and absorbed through related mechanisms (Garcia-Daga et al., 2025). It is important to emphasize, however, that the presence of strong correlations does not imply exclusive or dominant control. Li in cabbage also showed associations with the soil, albeit more moderately, indicating that soil–plant transfer is conditioned and dependent on additional factors such as element bioavailability, ionic competition, and crop-specific physiological mechanisms.

Li in soils showed positive correlations with B (ρ = 0.70), K (ρ = 0.64), and Mg (ρ = 0.47). In addition, very strong correlations were observed between B and Sr (ρ = 0.92), B and Ca (ρ = 0.89), Sr and Ca (ρ = 0.93), and between Co and Ni (ρ = 0.86). These results appear to reflect the geochemical patterns of soils in both lithium-bearing regions (Guarda and Vila Real) (Pinheiro, 2020). Despite the strong coherence of the soil, these relationships are not directly reflected in the elemental composition of cabbage leaves, demonstrating the selectivity of root uptake processes and with the simultaneous influence of other environmental factors, such as irrigation water. Correlations between water and soil were only weak to moderate, such as between Li in soil and As in irrigation water (ρ = 0.44) or between Li in irrigation water and Zn in soil (ρ = 0.45).

The results obtained from correlation analysis are consistent with those derived from PCA, reinforcing the interpretation of a statistically complex system controlled by multiple independent factors. To corroborate the previously discussed correlations, the PCA loading plot (Factor 3 vs. Factor 6) provides a representation of this specific accumulation pattern, as illustrated in Fig 1. A robust association between Li and Na is highlighted, as both elements dominate Factor 6 with loadings > 0.7. Furthermore, this group is entirely separated from contaminants such as As and Be, which dominate Factor 3. This spatial separation in the multivariate model substantiates the premise that Li accumulation in cabbage follows a distinct physiological or environmental pathway, acting independently of the uptake of other heavy metals.

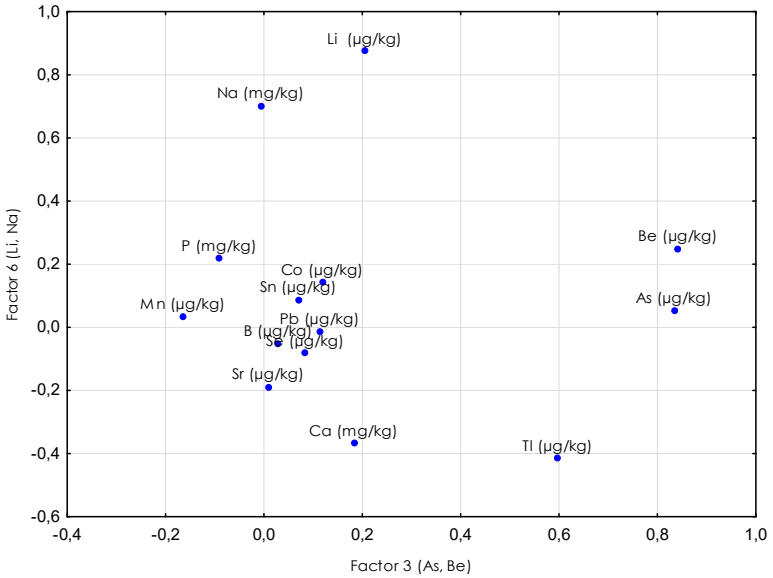


Fig 1. Principal component loading projection (Factor 3 vs. Factor 6) for the cabbage matrix.

PCA applied exclusively to cabbage leaf data identified six principal components with eigenvalues greater than 1, collectively explaining approximately 82% of the total variance. Variance is relatively evenly distributed across components, with the first principal component accounting for only about 22% of the total variability and no clearly dominant axis emerging.

The different principal components are structured by distinct groups of variables, including essential nutrients, transition metals, and potentially toxic elements, indicating that cabbage elemental composition is influenced by several independent and overlapping processes rather than by a single controlling factor. In this context, Li appears to be associated only with lower-order components, confirming its non-structuring role in the multivariate variability of the plant matrix.

A similar pattern was observed in the PCA applied to cabbage soils, which explained approximately 86% of the total variance and was dominated by gradients related to soil mineralogy and natural geochemical composition. In this case, Li is associated with an intermediate component, together with elements such as Tl, Be and K, reflecting a consistent geochemical behaviour within the soil compartment but without a dominant influence on the overall variance structure.

Of importance is that, in both the cabbage and soil PCA, samples do not show consistent separation by proximity to mining areas or between the Guarda and Vila Real regions, as illustrated in Fig 2. The continuous distribution of samples in multivariate space and the absence of well-defined clusters suggest that spatial differences do not exert a dominant control on the observed elemental variability.

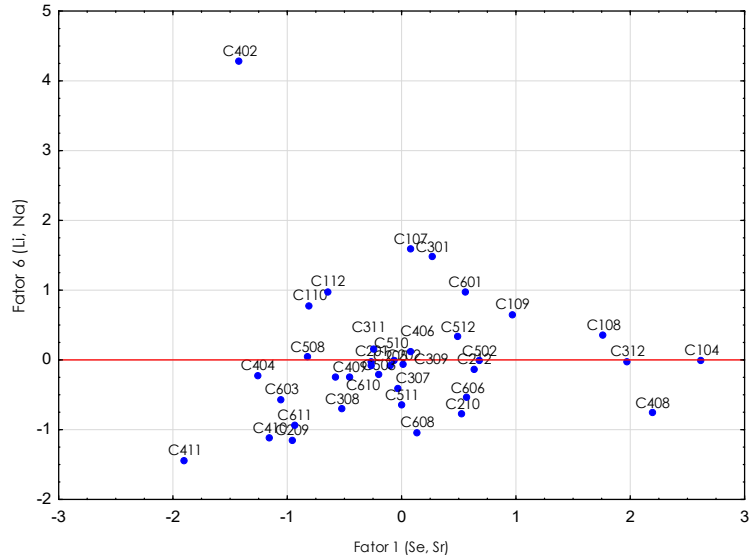


Fig 2. Principal component score projection (Factor 1 vs. Factor 6) for the cabbage matrix.

B. Potatoes

The integrated analysis of correlation results and PCA applied to potatoes (tuber) reveals a behaviour distinct from that observed in cabbages, reflecting physiological differences associated with the analysed plant organ (Tanveer et al., 2019).

Correlation analysis did not reveal a direct association between Li in the tuber and Li in soils or irrigation waters, indicating that Li transfer to the edible part is limited. Nevertheless, moderate correlations were observed between Li in the tuber and certain metals in soils, particularly Sn ($\rho = 0.75$) and Pb ($\rho = 0.49$). These associations primarily suggest geochemical co-occurrence within the environmental system, rather than a mechanism of proportional Li uptake by the tuber.

In contrast, Li in potato soils showed negative correlations with several elements measured in the tuber, including As ($\rho = -0.47$), Se ($\rho = -0.56$), Fe ($\rho = -0.47$), Mg ($\rho = -0.51$) and K ($\rho = -0.34$). This set of correlations indicates that higher soil Li concentrations may indirectly influence tuber elemental composition, possibly through ionic competition or selective exclusion mechanisms at the root level.

Correlation analysis within the potato matrix revealed strong associations between specific element pairs, such as As-Se ($\rho = 0.88$) and Mn-Zn ($\rho = 0.81$), reflecting internal physiological processes. Li showed only a moderate correlation with Na in the tuber ($\rho = 0.68$), consistent with similar ionic behaviour, but there was no evidence of a dominant role in structuring the plant matrix.

Soils associated with potatoes exhibited high internal geochemical coherence. The positive associations between Li in soil and elements such as Tl ($\rho = 0.80$), Be ($\rho = 0.68$), and K ($\rho = 0.66$) are compatible with the geochemical behaviour of these elements in regions characterized by lithium-bearing lithologies, where Li commonly occurs associated with granitic and pegmatitic systems enriched in alkali elements. In such geological contexts, Li, Tl and Be may share similar mineral hosts, including micas, feldspars and accessory silicate phases (Errandonea-Martin et al., 2026).

Irrigation waters associated with potatoes also showed strong correlations among transition metals, such as Mn-Ni ($\rho = 0.93$), Mn-Zn ($\rho = 0.94$), Mn-Pb ($\rho = 0.96$), Ni-Pb ($\rho = 0.98$), and Zn-Pb ($\rho = 0.97$). Strong correlations reflecting water mineralization were also observed, including Sr-Ca ($\rho = 0.98$), Mg-Ca ($\rho = 0.83$), Mg-Na ($\rho = 0.81$), Mg-K ($\rho = 0.83$), B-Mg ($\rho = 0.82$), and Se-Na ($\rho = 0.85$). Li showed only moderate correlations in water, particularly with Na ($\rho = 0.55$), B ($\rho = 0.42$), and Cu ($\rho = 0.46$), suggesting that this matrix does not exert a direct or dominant control on Li variability in the tuber.

The PCA applied exclusively to potato tubers revealed six principal components with eigenvalues greater than 1, explaining approximately 80.7% of the total variance, with variance distributed across several axes (PC1: 24.9%; PC2: 16.9%; PC3: 14.3%) and no clearly dominant component. The components are mainly associated with essential nutrients and transition metals, reflecting strong physiological control over tuber composition. Li showed its highest loading only on the fourth principal component, accounting for approximately 9.5% of the total variance, in association with elements such as Pb and Be. This position indicates that Li contributes only secondarily to overall tuber variability and does not follow the main compositional gradients.

PCA applied exclusively to potato soils explained approximately 85.1% of the total variance, with the first three components accounting for about 66.8% of observed variability. The first component is dominated by elements related to soil mineralogy, whereas the second component is strongly influenced by Li in association with Tl, Be, and K. Despite this clear geochemical structure in soils, factor score plots did not reveal consistent clustering or separation by sampling location. As observed for the PCA potato, variability is continuously distributed in multivariate space, with only a small grouping associated with a component dominated by B, As, Se, Mg, and K, involving 7 of the 24 samples, representing a weak and localized trend.

However, the PCA score plot for potato tubers (Factor 3 vs. Factor 4) effectively discriminates the samples from Barroso (Vila Real) from those of the Guarda district, as illustrate Fig 3. This clustering pattern suggests that the mineral profile of the tubers is primarily determined by the regional soil composition and the specific geological background of each area. Even with limited Li uptake, the potato matrix successfully captures a distinct geochemical signature, allowing a clear differentiation between the pre-exploitation site in Vila Real and the active mining contexts in Guarda.

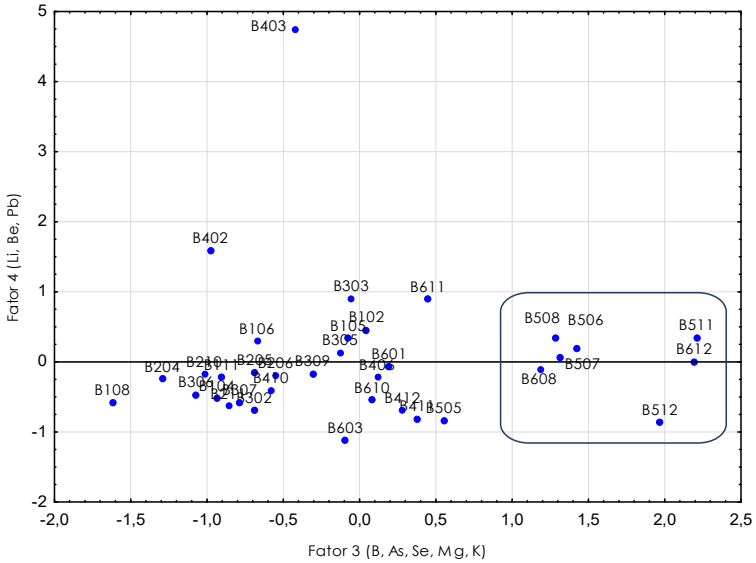


Fig 3. Principal component score projection (Factor 3 vs. Factor 4) for the potato matrix.

Overall, the results indicate that in potatoes, the relationship between Li and the composition of the edible part is strongly modulated by physiological regulation. The absence of dominant multivariate patterns and consistent spatial separation suggests that the observed variability is consistent with natural geochemical controls on soils and waters in the studied areas, with no clear multivariate statistical signal associated with recent anthropogenic influences.

C. Irrigation waters

Correlation analysis identified well-defined element groups, demonstrating high internal coherence of the water matrix. Strong correlations were observed between Sr and Ca ($\rho = 0.97$), Sr and Mg ($\rho = 0.83$), Mg and Ca ($\rho = 0.80$), Mg and Na ($\rho = 0.83$), and Mg and K ($\rho = 0.82$), which are consistent with the Guarda and Barroso lithium-bearing regions (Pinheiro, 2020). Li displayed only moderate correlations within the water matrix, notably with Na ($\rho = 0.53$), Be ($\rho = 0.45$), Cu ($\rho = 0.32$), and Ca ($\rho = 0.31$). The lack of strong correlations between Li and transition metals indicates that this element is not part of the dominant metal group and exhibits distinct geochemical behaviour in irrigation waters.

PCA applied to irrigation waters identified four principal components with eigenvalues greater than 1, explaining approximately 83.3% of the total variance. Variability is dominated by the first two components, with PC1 explaining approximately 36.6% and PC2 about 27.3% of the total variance, corresponding, respectively, to water mineralization and the grouping of transition metals and lead. Components of lower order explain smaller proportions of variance, with PC3 accounting for approximately 11.0% (dominated by As and Fe) and PC4 explaining about 8.5%, where Li loads together with Be.

The position of Li in a lower-order component confirms that this element contributes only secondarily to overall water variability and does not follow the dominant gradients defined by major ions or transition metals. Factor score plots show a continuous sample distribution, with no well-defined clustering or consistent separation by sampling location or region. Occasional extreme values, such as AR410 (borehole water) and AR210 (spring water), reflect localized hydrogeochemical conditions and do not influence the dominant multivariate patterns, as illustrate Fig 4.

Within this framework, the statistical structure observed in irrigation waters is compatible with a system predominantly ruled by natural hydrogeochemical gradients. The absence of consistent multivariate separation or clustering associated with specific areas suggests that potential local influences do not manifest, at the time of sampling, as a dominant statistical signal in water composition.

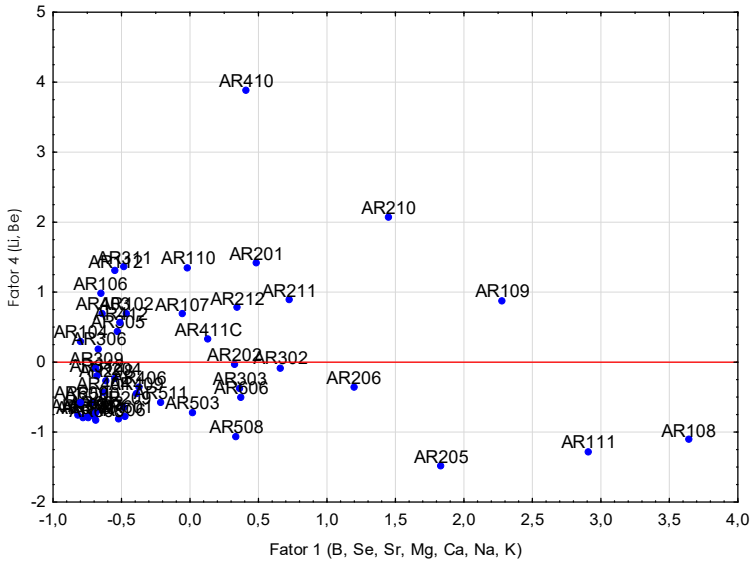


Fig 4. Principal component score projection (Factor 1 vs. Factor 4) for the irrigation water matrix.

IV. Conclusions

This study characterized the soil–water–plant system in agricultural contexts of the Li bearing regions, particularly in Guarda and Vila Real regions, using an integrated statistical approach based on correlation analysis and principal component analysis, applied separately to soils, irrigation waters, and agricultural crops with contrasting physiological characteristics.

The results showed that crop responses varied according to the physiological characteristics of the edible plant organ. In cabbages, significant associations were observed between Li concentrations in leaves and irrigation water, indicating that this matrix can play a role in Li variability. In contrast, potato tubers exhibited stronger limiting Li transfer to the edible part and showed mainly indirect effects related to nutrient redistribution.

Across all matrices, multivariate analyses did not reveal well-defined clusters or systematic separation of samples according to proximity to current mining areas or regional origin. The continuous distribution of samples in factor space, together with the absence of dominant components in principal component analysis, suggests that the observed variability is largely explained by the interaction of natural geochemical and physiological processes. At the time of sampling, no clear multivariate statistical signal indicative of recent anthropogenic influence was identified.

This framework is particularly relevant for studies conducted in lithium-bearing regions with potential mining interest, where distinguishing natural geochemical variability from localized anthropogenic signals requires robust multivariate analysis.

Acknowledgment

Ana Sofia Matos acknowledge financial support via National funds from FCT, I.P. – Fundação para a Ciência e a Tecnologia, financed this work in the scope of the project UID/00667/2025 (https://doi.org/10.54499/UID/00667/2025) (UNIDEMI).

This work is financed by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., within the scope of the project EXPL/CTA-AMB/0977/2021 (http://doi.org/10.54499/EXPL/CTA-AMB/0977/2021). The Authors acknowledge the strategic projects UID/4292/2025 and UID/PRR/4292/2025 granted to MARE - Marine and Environmental Sciences Centre, project LA/P/0069/2020 (https://doi.org/10.54499/LA/P/0069/2020) granted to the Associate Laboratory ARNET - Aquatic Research Network, and to the Portuguese Mass Spectrometry Network (RNEM - Rede Nacional de Espectrometria de Massa - LISBOA2030-FEDER-01319300).

References

Bolan, N., Hoang, S. A., Tanveer, M., Wang, L., Bolan, S., Sooriyakumar, P., Robinson, B., Wijesekara, H., Wijesooriya, M., Keerthanan, S., Vithanage, M., Markert, B., Fränzle, S., Wünschmann, S., Sarkar, B., Vinu, A., Kirkham, M. B., Siddique, K. H. M., & Rinklebe, J. (2021). From mine to mind and mobiles – Lithium contamination and its risk management. Environmental Pollution, 290, 118067. https://doi.org/10.1016/j.envpol.2021.118067

Bollaert, Q., Vassilieva, E., Lavikko, S., Araujo, F. P., Dehaine, Q., & Cappuyns, V. (2025). Metal(loid) mobility, solid-phase speciation and in vitro bioaccessibility in European hard-rock lithium (Li) deposits. Science of The Total Environment, 970, 178992. https://doi.org/10.1016/j.scitotenv.2025.178992

Chaves, C., Pereira, E., Ferreira, P., & Guerner Dias, A. (2021). Concerns about lithium extraction: A review and application for Portugal☆. The Extractive Industries and Society, 8(3), 100928. https://doi.org/10.1016/j.exis.2021.100928

Dunlap, A., & Riquito, M. (2023). Social warfare for lithium extraction? Open-pit lithium mining, counterinsurgency tactics and enforcing green extractivism in northern Portugal. Energy Research & Social Science, 95, 102912. https://doi.org/10.1016/j.erss.2022.102912

Eniowo, O. D., Grobler, H., Mulaba-Bafubiandi, A. F., Onifade, M., Makinde, O., & Daramola, S. O. (2026). Plugging the gaps: Sustainable resource policy and revenue leakages in Nigeria's small-scale lithium mining. The Extractive Industries and Society, 25, 101788. https://doi.org/10.1016/j.exis.2025.101788

Errandonea-Martin, J., Roda-Robles, E., Garate-Olave, I., García De Madinabeitia, S., Sarrionandia, F., Santos-Loyola, N., & Lima, A. (2026). Deciphering geochemical variations driven by metasomatism around a lithium pegmatite at the Gonçalo deposit, Portugal: Implications for rare-metal exploration in granitic terranes of the European Variscan Belt. Mineralium Deposita. https://doi.org/10.1007/s00126-026-01440-9

Garcia-Daga, S., Fischer, S., & Gilliham, M. (2025). Lithium in plants. New Phytologist, 248(4), 1639–1654. https://doi.org/10.1111/nph.70523

Gbondo, D., Cerpa-Perez, V., Pham, N. M., Zhao, Y., & Rumchev, K. (2025). Exposure to Airborne Contaminants and Respiratory Health Among Lithium Mine Workers in Western Australia. Environments, 12(6), 206. https://doi.org/10.3390/environments12060206

Jesus, S., Ventura, M., Assunção, R., Gueifão, S., Delgado, I., Rego, A., Ribeiro, M., Martins, M., Neves, O., & Coelho, I. (2024). Study around the Barroso mine (Portugal): Baseline levels of lithium for assessing future exposure and risks from Li mining activity. Food and Chemical Toxicology, 190, 114825. https://doi.org/10.1016/j.fct.2024.114825

Pinheiro, D. A. C. (2020). Concentração e distribuição do lítio em solos portugueses. Master Tesis.

Sadik-Zada, E. R., Gatto, A., & Scharfenstein, M. (2023). Sustainable management of lithium and green hydrogen and long-run perspectives of electromobility. Technological Forecasting and Social Change, 186, 121992. https://doi.org/10.1016/j.techfore.2022.121992

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

From Validated Problem to First Revenue: A Quality-Grounded Design Science Research Methodology for Early-Stage Startups

Luis Alberto González [1][0009-0006-7408-3015], **Valeria de Castro** [1][0000-0003-2193-7807] **and Esperanza de Marcos** [1][0000-0003-2666-9995]

[1] Universidad Rey Juan Carlos, Calle Tulipán S/N, 28933, Madrid, España
Luis.gonzalezc@urjc.es

***Abstract*—**Early-stage startups continue to face very high failure rates despite the wide use of validation frameworks such as Lean Startup, Customer Development, and Business Model Canvas. A clear gap remains: none of these frameworks operationally integrates the transition from problem–solution fit to first sustainable sales and basic financial discipline within a quality management structure. This paper addresses that gap through a Design Science Research approach, triangulating three different evidence sources: an ongoing systematic literature review (37 primary studies), 30 semi-structured interviews with founders recorded in podcast format (2020–2026), and two descriptive surveys of early-stage founders (n=17 each, 2024 and 2026). The three sources converge on four critical axes for the early stage: validation of severe and monetisable problems, team configuration and social capital, financial discipline and cash management, and early sales with commercial legitimacy building. In response, a three-stage methodology is proposed, structured as an applied PDCA cycle: problem discovery and validation; problem–solution fit and early product–market fit (PMF) signals; and first sales with basic financial discipline. Each stage is bounded by explicit quality gates that put the PDCA Check logic into practice. A comparative analysis against five existing frameworks shows that no prior approach integrates the full set of early-stage dimensions within a single sequential and iterative roadmap. The proposed methodology gives founders and incubators an actionable, quality-grounded roadmap that aligns market validation with organisational configuration and financial control, designed to help ventures navigate the valley of death under limited resources.

Keywords – *Early-Stage Startups; Problem Validation; Financial Discipline; Design Science Research*

I. Introduction

Early-stage startups have established themselves as key drivers of innovation and employment (Shanbhag and Pardede, 2022);(Jing, 2023), However, they continue to face very high failure rates in their first years of operation (Hejazi *et al.*, 2024). This challenge is particularly acute in the Spanish entrepreneurial ecosystem, where the Total Early-stage Entrepreneurial Activity (TEA) rate reached 7.2% of the adult population in 2024 (four consecutive years of growth) yet the abandonment rate rose to 3.5%, with lack of profitability as the primary cause in 31.5% of cases (Fernández-Laviada *et al.*, 2025). More than half of recent ventures are bootstrapped with under €15,000 in seed capital and 53% operate without employees (Fernández-Laviada *et al.*, 2025), making lean resource management and early revenue generation essential for survival.

In response to this challenge, the literature has developed frameworks such as Lean Startup (Ries, 2011), Customer Development (Blank, 2007), and Business Model Canvas (Osterwalder and Pigneur, 2010), which emphasise rapid experimentation, minimum viable products, and the iterative validation of market hypotheses. More recent frameworks, such as the Blitz Canvas (Shanbhag and Pardede, 2022) and the Scaling Wheel Framework (Font-Cot *et al.*, 2025), have extended the focus toward scalability and governance. However, none of these approaches operationally integrates the transition from problem–solution fit to first sustainable sales and the basic financial discipline required to prevent cash exhaustion, nor do they embed this transition within a quality management structure.

In parallel, research on startup survival has identified a broad set of critical factors that transcend product validation: founding team configuration, quality of social capital, organisational learning, financial discipline, and the capacity to build legitimacy in ecosystems where the venture is still unknown (Hejazi *et al.*, 2024); (Spiegel *et al.*, 2016). These elements appear in fragmented form across the literature and are rarely integrated into operational methodologies that can guide entrepreneurs through the idea, validation, and launch phases. Recent reviews by Nosella et al. (2025) and Argaw and Liu (2024) identify this gap explicitly, highlighting the absence of dynamic frameworks that connect survival factors with tools and lifecycle phases (Argaw and Liu, 2024); (Nosella *et al.*, 2025).

This paper addresses that gap by proposing a methodology for early-stage startups that integrates three components frequently treated in isolation: the validation of a severe and monetisable problem, the design and testing of minimal solutions with strong demand evidence, and the generation of first sales under basic financial discipline. Previous frameworks focus mostly on product validation (Pattyn, 2023) or managing uncertainty (Sosa Gómez, 2025). In contrast, our methodology explicitly integrates sales, financial control, and quality management into a single iterative roadmap.

To achieve this, the proposed methodology is built on three quality management principles. The three-stage process follows the PDCA (Plan-Do-Check-Act) cycle. Each stage represents a key quality management function. Stage 1 focuses on quality planning by defining the problem and the validation criteria. Quality control happens at the transitions, where quality gates check the evidence against specific thresholds. Finally, quality assurance is ensured by the iteration logic, which prevents startups from moving forward without validated demand. Second, the methodology applies the stage-gate logic widely used in product development quality management (Cooper, 1990) to the pre-revenue startup context, where limited resources make process discipline especially important; each quality gate functions as a go/kill decision point analogous to a stage-gate review. Third, the financial discipline routines in Stage 3 act as an early form of operational process control for resource-constrained founders. Cash monitoring and runway tracking serve as concrete methods to manage performance under

uncertainty, adapting process-control logic from quality management to the resource-constrained startup context. A synthesis table in Section V makes this correspondence explicit, aligning each stage with its PDCA phase and QMS function.

The proposal is grounded in an ongoing systematic literature review, 30 semi-structured founder interviews, and two surveys conducted in 2024 and 2026. The paper is structured as follows. Section II presents the theoretical and empirical background. Section III describes the Design Science Research methodology. Section IV synthesises the key findings for the early stage. Section V presents the proposed stage-based methodology. Section VI discusses implications. Section VII presents the conclusions, limitations, and future research directions.

II. Theoretical and empirical background

A. Existing Frameworks for Early-Stage Startups

Among the most influential frameworks are Lean Startup (Ries, 2011), Customer Development (Blank, 2007), and Business Model Design (Osterwalder and Pigneur, 2010). Lean Startup proposes iterative build–measure–learn cycles supported by MVPs to speed up validated learning and reduce resource waste, a logic close to PDCA, though without explicit quality gates or financial discipline. Lean Startup's "Innovation Accounting" focuses on product engagement metrics to validate learning, not on operational financial discipline such as runway and cash-burn tracking, which is the concern of Stage 3. Customer Development (Blank, 2007) structures customer discovery and validation through the formulation and testing of hypotheses about the problem, solution, channel, and revenue model, and its Customer Validation phase explicitly aims to build a repeatable sales roadmap. However, its structural assumptions cater more to heavily resourced or enterprise-level environments, leaving bootstrapped founders without granular tactical guidance for zero-budget prospecting and strict financial runway gates, precisely the context that characterises the Spanish early-stage ecosystem described in Section I. The Business Model Canvas gives a synthetic view of how a startup creates, delivers, and captures value (Osterwalder and Pigneur, 2010), but it is an ontology rather than a methodology: it provides a snapshot of a business model without any sequential logic for testing it or managing the cash required to do so, as several studies have noted (Font-Cot *et al.*, 2025);(Shanbhag and Pardede, 2022).

The problem discovery activities behind Stage 1 of the proposed methodology are also close to the Empathise and Define phases of Design Thinking (Brown, 2008): deep customer interviews, friction mapping, and workaround identification serve a similar purpose to design thinking's empathy-driven problem framing. This link is acknowledged, but the proposed methodology goes further than design thinking by placing the whole process (from problem discovery to first sales) within a PDCA quality structure with explicit gate criteria.

The ongoing systematic review has also identified more recent integrative frameworks, such as the Blitz Canvas (Shanbhag and Pardede, 2022) and the Scaling Wheel Framework (Font-Cot *et al.*, 2025), which extend the focus toward scalability and governance. Nevertheless, the dominant pattern remains an emphasis on ideation and early validation, with comparatively little attention to the launch phase and first sales, and no framework embeds this process within an explicit quality management cycle.

B. Key Findings from the Systematic Literature Review (in progress)

The authors conducted a systematic literature review (Kitchenham, 2004) on frameworks, models, and methodologies oriented toward startup survival, following the PRISMA 2020 protocol (Page *et al.*, 2021) and using Web of Science and Scopus as primary databases. The search was completed in January 2025, yielding 196 records; 37 primary studies were included after full-text screening, quality appraisal, and data extraction. Synthesis and write-up of the full SLR are ongoing and will be reported in a dedicated publication; all findings cited here are stable as of the search cutoff date.

Frameworks and dimensions covered. The frequency distributions reveal two principal patterns. First, Business Model Design (n=10) and Lean Startup (n=8) dominate the literature, followed by Agile (n=5) and Customer Development (n=3). This distribution reflects a predominant emphasis on early validation and business model design. Second, the most recurrent dimensions covered are team (n=14), processes and operations (n=13), and resources and capabilities (n=12), followed by strategic alliances and networks (n=11) and financial management (n=11). Strikingly, sales (n=3) appears at the bottom of the dimension frequency list — a structural bias toward internal organisation, governance, and networks at the expense of revenue generation and commercial capability.

TABLE I. Frameworks and Dimensions Covered in the SLR (n=37 studies)

Category	Item	Frequency
Frameworks	Business Model Design	10
	Lean Startup	8
	Agile	5
	Customer Development	3
Dimensions	Team	14
	Processes and Operations	13
	Resources and Capabilities	12
	Strategic Alliances and Networks	11
	Financial Management	11
	Product / Value Proposition / Customers / Business Model	7
	Marketing and Communication	6
	Technological Infrastructure	4
	Sales	3

Critical survival factors. The founding team (n=19 studies) and market adaptation (n=12) emerge with greatest consistency, followed closely by social capital and networks (n=11), and financial management and organisational learning (n=10 each). The financial dimension is particularly revealing: despite appearing both as a critical success factor and as a recurring obstacle (cash flow problems are identified in n=8 studies (Giardino *et al.*, 2016);(Pattyn, 2023)) none of the dominant methodologies incorporates financial discipline tools in explicit and operational form.

TABLE II. CRITICAL SURVIVAL FACTORS IDENTIFIED IN THE SLR (N=37 STUDIES)

Survival Factor	Frequency	Notes
Founding team	19	Most frequently cited; limited managerial capability identified as barrier in n=16 studies
Market adaptation	12	Agility in responding to market feedback
Social capital and networks	11	Access to resources and organisational learning
Financial management	10	Financial discipline and cash flow management
Organisational learning	10	Ability to iterate and adapt based on evidence
PMF difficulty	9	Agile validation of PMF identified as principal obstacle
Commercial capability	8	Marketing and sales capabilities ranked as critical success factor
Cash flow problems	8	Identified as recurring obstacle to survival

Dynamic Integrative Framework. From this synthesis, the authors formulated a Dynamic Integrative Framework that organises the recurrent factors into four strategic domains: (1) human and social capital; (2) opportunity and competitive environment; (3) business model design and evolution; and (4) strategic capital and resources. These domains are deployed across five maturity phases: idea, validation, launch, scaling, and consolidation. The stage-based methodology proposed in this paper constitutes a specialisation of that framework, operationalising the first three phases within a PDCA–quality-gate structure.

C. Complementary Evidence: Podcast Interviews and Surveys

Two complementary evidence sources triangulate the systematic review. Both are treated as independent cross-sections that corroborate the SLR at two points in time; they do not track the same entrepreneurs and no claim of longitudinal evolution between waves is made. Their methodological role is to test whether the structural gaps identified in the literature also surface in practitioner accounts and in founders' self-reported challenges.

Podcast interviews. The primary qualitative source comprised 30 semi-structured interviews with founders and co-founders, conducted in podcast format between 2020 and 2026 within two series focused on impact-oriented entrepreneurship: Socialpreneurs Podcast (González, 2023)and De Idea a Impacto Podcast (González, 2026). The 30 interviewed founders (averaging 3 to 5 years of sector experience) represented impact-driven and tech-enabled sectors (e.g., EdTech, AgTech, FinTech) within the Spanish ecosystem, primarily operating B2B or SaaS models. Founding teams typically comprised 2 to 3 members. These interviews directly informed the design of the roadmap's practical stages, translating the theoretical SLR gaps into actionable steps. At the time of the interview, these ventures had already passed through the early-stage "valley of death" and were mostly in the scaling and consolidation phases, with proven market traction, operational break-even, or profitability. Early financing relied heavily on bootstrapping and organic revenue from manual MVPs, and was often complemented at later stages by equity crowdfunding, public or European grants, and seed venture capital.

Because the sample is made up of ventures that survived and reached operational stability, full episode transcriptions were analysed through inductive thematic analysis to capture retrospective, survival-tested insights about their early-stage transitions. The analysis produced consistent patterns that directly shaped the validation requirements of the proposed roadmap: manual, founder-led problem validation before any significant engineering spend; first sales carried out personally by the founding team through direct niche prospecting; strict bootstrapping routines focused on early cash flow control; and the use of incubator affiliation to reduce the liability of newness. The data also revealed a useful tension: several founders claimed, in hindsight, to have followed a problem-first approach, while their own accounts described building before validating, which then led to costly pivots. This gap between what founders said and what they actually did reinforces the need for the explicit, threshold-based quality gates built into the proposed methodology.

Surveys. The second source comprised two descriptive surveys, each administered to 17 early-stage founders recruited through the authors' professional networks. Survey 1 (2024) included founders from social impact, mobility, agritech, and software development. Survey 2 (2026) covered fintech, proptech, edtech, sustainability, and consumer goods. Consistent with the formative DSR design of this study, the surveys are deployed deliberately as illustrative triangulation instruments rather than as a basis for inference: with n=17 and non-probability sampling, they serve to corroborate, at the level of founder-perceived priorities, the dimensions identified by the systematic review and the interviews. No claim of statistical representativeness is made, and survey results are reported as ranked priorities rather than population estimates.

Survey 1 (2024) found that a majority of founders were unfamiliar with process management certifications, and most had not considered structured process methods, indicating that process management is not an explicit priority in early stages. Survey 2 (2026) revealed that 9 of 17 respondents (53%) perceived ineffective marketing and sales strategies as a serious or critical problem; access to financing, connections with corporations or potential clients, and marketing and sales support were the most valued forms of support; and 14 of 17 respondents (82%) had participated in incubator programmes, with around 10 of those 14 (72%) rating their impact as highly positive, consistent with the literature highlighting the role of such programmes in providing resources, networks, and legitimacy (Bellavitis *et al.*, 2020).

III. Research Methodology (Design Science Research Process)

This study adopts Design Science Research (DSR) as its methodological framework, following the process model of Peffers et al. (2007) and the conceptual framework of Hevner et al. (2004). DSR is the appropriate paradigm when the research objective is to design and evaluate an artefact that resolves a relevant practical problem while contributing to the knowledge base of a field (Hevner *et al.*, 2004). The research process followed the six activities of the Peffers et al. (2007) model (Peffers *et al.*, 2007).

To clarify the methodology: this research follows a six-step DSR process (problem identification, objective definition, design, demonstration, evaluation, and communication) to construct the artefact. The resulting artefact itself (detailed in Section V) is a distinct three-stage quality management roadmap for startups. The six DSR steps represent the research method, while the three stages represent the proposed practical solution.

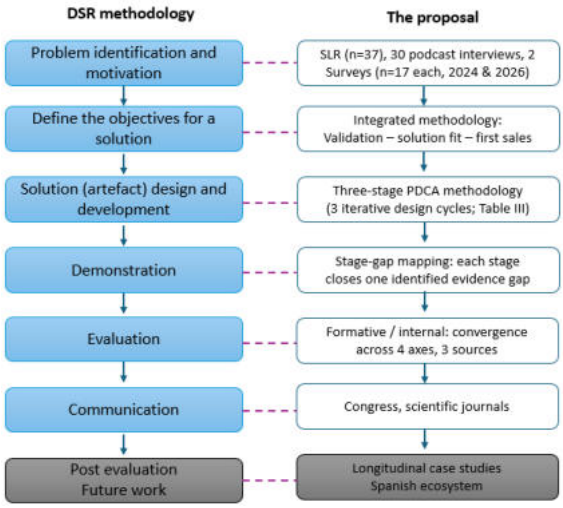


Fig. 1. DSR process model (Peffers *et al.*, 2007) mapped to the artefact outputs of this study.

Problem identification and motivation. Three evidence sources, described in Section II, played distinct roles: the SLR (37 primary studies) gave the theoretical anchoring and identified the structural gaps; the 30 interviews gave qualitative evidence of how relevant those gaps are in practice; and the two surveys gave descriptive quantitative evidence used to prioritise the elements of the artefact. The three sources point to the same structural gap (the lack of an operational, quality-grounded framework that integrates problem validation, solution fit, and first sales with basic financial discipline) and this is the motivation for the artefact design.

Definition of solution objectives. The artefact was designed as a stage-based methodology that integrates, within a sequential and iterative PDCA-structured roadmap, the critical survival factors identified — in a format that founders and incubators can understand and apply directly during the idea, validation, and launch phases. Kernel theories include Customer Development (Blank, 2007), Lean Startup (Ries, 2011), problem and solution validation methodologies (Fitzpatrick, 2013);(Maurya, 2012), founder-led sales frameworks (Kazanjy, 2020), and financial management for early-stage startups (Bhimani, 2017).

Artefact design and development. The design went through three iterative cycles. In the first, the four critical axes from the SLR were mapped to candidate stage objectives, giving an initial two-stage structure that separated validation from launch. In the second, qualitative evidence from the 30 interviews showed that founders consistently described the transition from validated demand to first revenue as a distinct and difficult phase (needing specific practices (founder-led selling, cash monitoring) that were neither product development nor full-scale commercial activity) which led to separating Stage 3. In the third, the survey findings (priority given to financing access, commercial connections, and sales support in Survey 2) were used to weight the practices within each stage. Table III (presented in Section V) shows the traceability from each evidence source to each design decision.

Demonstration. The methodology was demonstrated by applying it to the structural gap documented in the SLR: Stage 1 responds to the problem validation deficit; Stage 2 responds to the lack of operational problem–solution fit guidance; Stage 3 responds to the sales and financial discipline gap. Each stage defines at least one objective, one set of recommended practices, and one quality gate, showing that the structure of the artefact maps onto the gaps documented in the evidence (Peffers *et al.*, 2007).

Evaluation. Artefact evaluation is formative and internal, in line with the early-stage DSR contribution type described by Gregor and Hevner (2013). The fact that three different sources converge on the same four critical axes is evidence of internal coherence and content validity. No summative evaluation is claimed; this paper is a Type 3 contribution (Improvement) in Gregor and Hevner's (2013) typology (Gregor and Hevner, 2013).

IV. Synthesis of Key Findings for the Early Stage

This section presents the results of the triangulation as the empirical basis of the artefact. The integration of findings from the three evidence sources (the systematic review, the 30 founder interviews, and the two surveys) converges on four axes especially critical during the idea, validation, and launch phases. Each axis is reported below with its supporting evidence from all three sources, so that the contribution of the interviews to the roadmap is explicit rather than implied. These axes correspond to the dimensions with highest combined frequency across the three sources: team and social capital (SLR: n=19 studies for founding team; n=11 for social capital), financial discipline (SLR: n=10 for financial management; n=8 for cash flow problems), problem validation (SLR: n=9 for PMF difficulty), and sales (SLR: n=3 frameworks addressing sales, despite n=8 studies identifying commercial capability as a critical success factor).

Validation of severe and monetisable problems. Both the reviewed literature and practitioner testimonies highlight the importance of addressing critical problems (those associated with economic losses, serious inefficiencies, or operational risks) rather than incremental improvements (Chorev and Anderson, 2006). The SLR identifies difficulty in agile product–market fit validation as one of the principal obstacles in early stages (n=9 studies), and notes that many startups fail because of superficial validation that conflates declared interest with genuine willingness to pay (Giardino *et al.*, 2016). This emphasis is consistent with Customer Development methodology (Blank, 2007) and with evidence highlighting the diagnostic value of exploring past behaviours and prior attempts to solve the problem (Fitzpatrick, 2013). Interview evidence consistently identified premature solution development as a recurring cause of early difficulty.

Team configuration, social capital, and liability of newness. The SLR consistently places the founding team (n=19) and social capital (n=11) among the most frequently cited critical survival factors. Prior experience, complementarity between technical and business profiles, and the capacity to mobilise networks are associated with better resource access and faster organisational learning (Spiegel *et al.*, 2016). Limited managerial capability emerges as the most frequent barrier to growth (n=16). The literature on liability of newness establishes that startups begin with a legitimacy deficit that impedes customer acquisition and financing (Chorev and Anderson, 2006), addressed through prior contacts, reference customers, and incubator participation (Aldrich, 1999).

Financial discipline and cash management. The SLR identifies financial discipline as a survival factor (Hejazi *et al.*, 2024); (Shi *et al.*, 2024) and documents recurring cash flow problems (n=8 studies). Premature scaling (before PMF) ranks among the most frequent barriers (n=6), amplifying financial and operational risks. The 2026 survey confirms access to financing as a significant challenge.

Sales, early traction, and commercial legitimacy. The literature and practitioner evidence converge on the view that early selling is less a mass marketing exercise and more a learning process grounded in direct customer interaction, founder-led sales, and a focus on early adopter niches with intense pain (Asmoro *et al.*, 2018). Sales (n=3) is the least addressed dimension in existing frameworks, despite commercial difficulties appearing as a growth barrier (n=5 studies) and marketing and sales capabilities ranking among the critical internal success factors (n=8). The 2026 survey reinforces this: 53% regard the absence of effective sales strategies as a serious or critical problem.

V. Proposed Three-Stage Methodology (The Artefact)

The methodology in this section is the main result of the Design Science Research process: it is the artefact produced by triangulating the three evidence sources. The systematic review provided the dimensions and their relative weight; the interviews provided the practitioner-grounded observations that shaped how those dimensions were sequenced and put into practice; and the surveys provided the founder-perceived prioritisation used to weight practices within stages. Three design decisions show the generative role of the interviews. First, the repeated description of the validated-demand-to-first-revenue transition as a distinct and poorly supported phase led directly to separating Stage 3 from Stage 2. Second, the observation that degraded prototypes produced unreliable demand signals in experience-intensive ventures motivated adding the MXP option alongside the conventional MVP. Third, the consistent description of first sales as personal, learning-oriented prospecting (rather than delegated commercial activity) established founder-led sales as the core practice of Stage 3. Table III sets out this evidence-to-artefact traceability in full.

TABLE III. Design Decision Traceability

Design Decision	SLR Evidence	Interview Evidence	Survey Evidence
Three-stage structure (separating validation from launch)	Cash flow problems identified as critical obstacle (n=8); gap in sales frameworks (n=3).	Transition to revenue was consistently reported as a distinct, difficult phase requiring targeted practices.	53% rated sales as a serious/critical problem; sales support ranked 3rd most valued.
Problem-first validation (Stage 1)	PMF difficulty (n=9); superficial validation cited as a primary failure cause (Giardino *et al.*, 2016)	Tension surfaced: founders espoused problem-first validation but several built before validating and pivoted thereafter	Majority had not considered alternative process-management approaches
Founder-led sales (Stage 3 core practice)	Scarcity of sales frameworks (n=3); commercial capability identified as key success factor (n=8).	First sales were executed personally via direct prospecting, framed as a continuous learning process rather than scaling.	53% perceived ineffective sales strategies as serious/critical problem
Cash and runway monitoring (Quality gate)	Financial management n=10; cash flow problems n=8 (Hejazi *et al.*, 2024); (Shi *et al.*, 2024)	Founders described bootstrapping discipline oriented toward early break-even	Access to financing among most valued forms of support
Incubator participation (Stage 3 practice)	Social capital n=11; liability of newness (Aldrich, 1999); (Chorev and Anderson, 2006)	Founders described mobilising assembling reputation and incubator affiliation to mitigate liability of newness	72% of incubator participants rated impact as highly positive (Bellavitis *et al.*, 2020)

PDCA framing of the three stages	Processes and operations (n=13)emerged as the most recurrent dimension after team.	—	Most founders unfamiliar with process management certifications

The proposed methodology guides the transition from identifying a severe problem to making first sales under basic financial discipline. It is structured as an applied PDCA cycle: Stage 1 follows the *Plan* logic (defining the problem and the validation protocol), Stage 2 follows the *Do* logic (building and testing the minimal solution), and Stage 3 follows the *Act* logic (putting the validated solution into operation and feeding learning back into a new cycle). The *Check* logic is not limited to a single stage; it takes place at every stage transition through an explicit quality gate. Each stage can also contain its own internal Plan–Do–Check–Act micro-iterations, so the stage-level mapping describes the main orientation of each stage rather than a strict once-through sequence. The quality gates (observable, threshold-based criteria that decide whether a venture should advance, iterate, or pivot) work like the stage-gate decision points used in established quality management systems (Cooper, 1990). Fig. 2 illustrates this structure, and Table IV below presents the complete methodology with its PDCA and quality management correspondence.

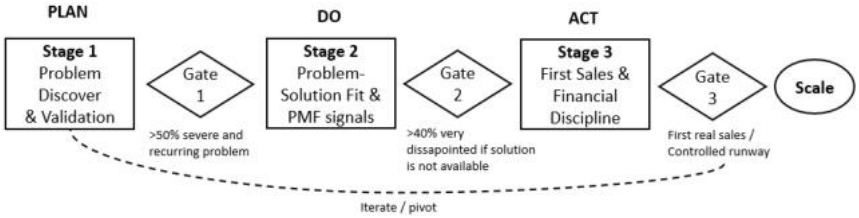


Fig. 2. PDCA–stage-gate structure of the proposed methodology. Each gate enacts the Check logic; a failed gate triggers Act (iterate or pivot).

TABLE IV. STAGE-BASED METHODOLOGY FOR EARLY-STAGE STARTUPS

	Stage 1	**Stage 2**	**Stage 3**
PDCA Phase	Plan	Do	Act
QM Function	Quality planning	Process execution	Process control and assurance
Name	Problem discovery and validation	Problem–solution fit and early PMF signals	First sales and basic financial discipline
Objective	Identify a severe and monetisable problem in a specific segment (Furr and Ahlstrom, 2011)	Design and validate a minimal solution with strong demand evidence	Convert validated fit into first recurring sales with cash control
Key Practices	Interviews with prospective buyers; hypothesis mapping on the BMC (Osterwalder and Pigneur, 2010); identification of workarounds and real friction points (Fitzpatrick, 2013)	Low-fidelity prototype / MVP / MXP — Minimum Experienceable Product (Hokkanen *et al.*, 2016), pursuit of paid pilots, pre-sales, or letters of intent	Founder-led sales(Kazanjy, 2020); direct prospecting in early adopter niche; reference customers (Furr and Ahlstrom, 2011); incubator participation (Bellavitis *et al.*, 2020); cash and runway monitoring (Bhimani, 2017)
Quality Gate	>50% of interviewees describe the problem as severe and recurring (Furr and Ahlstrom, 2011); initial willingness-to-pay signals identified; priority segment clearly defined (Maurya, 2012)	At least 40% of users would feel "very disappointed" if the solution were no longer available (Maurya, 2012); evidence of repeated use or early renewal; pilot payments or small-scale contractual commitments (Furr and Ahlstrom, 2011)	First real sales, as distinct from free pilots (Kaulio, 2003); runway monitored regularly with current figures known at all times (Bhimani, 2017); at least 70% of discretionary spending allocated to validation and customer acquisition activities, monitored monthly (Sethi, 2024)
Gate Decision	Advance to Stage 2 / Iterate (reframe segment or problem) / Pivot	Advance to Stage 3 / Iterate (redesign MVP/MXP) / Pivot	Advance to scaling phase / Iterate (refine sales process) / Pivot to Stage 1

A. Stage 1: Problem Discovery and Validation (PDCA: Plan)

Stage 1 follows the Plan phase of PDCA: it defines the problem hypothesis, designs the evidence collection protocol, and sets the quality gate thresholds that decide whether the venture is building on a real and monetisable foundation.

The stage calls for a minimum of twenty interviews (Ross and Lemkin, 2016) with genuine prospective buyers, following the principles of The Mom Test (Fitzpatrick, 2013): focusing on past behaviours rather than hypothetical futures, not presenting the idea in early conversations, and listening much more than speaking. The problem discovery activities at this stage (deep customer interviews, friction mapping, workaround identification) are close to the Empathise and Define phases of Design Thinking (Brown, 2008), grounding the planning process in real customer evidence before any solution is designed. The goal is to find pain points strong enough that customers would interrupt their routine to look for alternatives, and to map the resulting hypotheses onto the Business Model Canvas (Osterwalder and Pigneur, 2010).

Interview evidence motivating this stage: Founders' accounts consistently pointed to premature solution development (building before fully understanding the problem) as a recurring cause of early difficulty. This pattern was instructive because it appeared as a tension: several founders described a problem-first approach in principle, while their accounts showed building before validating, followed by a costly pivot. This gap between stated principle and actual practice supports the stage's focus on structured, gated problem validation rather than discretionary validation effort.

Quality Gate 1 requires: (1) more than 50% of interviewees describing the problem as severe and recurring (Furr and Ahlstrom, 2011), a threshold subject to sectoral calibration; (2) initial signals of willingness to pay (Maurya, 2012); (3) a clear priority

segment where the pain is most concentrated (Kaulio, 2003). Not clearing this gate triggers iteration or a pivot, following the PDCA logic of not moving to Do without a validated Plan.

B. Stage 2: Problem–Solution Fit and Early PMF Signals (PDCA: Do)

Stage 2 follows the Do phase: the validated problem hypothesis from Stage 1 is actively tested by building and deploying a minimal solution.

The stage uses either a Minimum Viable Product (MVP) or a Minimum Experienceable Product (MXP) (Hokkanen *et al.*, 2016). The MVP minimises both functionality and experience to maximise learning speed, while the MXP keeps the quality of the core user experience and reduces feature scope instead, applying quality management thinking to the product design decision. In segments where a degraded experience would hide honest willingness-to-pay signals, the MXP logic is preferable. In technical or infrastructure contexts where speed outweighs experience fidelity, the conventional MVP approach remains appropriate.

Interview evidence motivating this stage: Founders' accounts showed that prototype fidelity affected the reliability of early demand signals: in experience-intensive ventures, a degraded prototype tended to produce discouraging responses that reflected the prototype rather than the underlying value proposition. This observation motivated adding the MXP option alongside the conventional MVP, together with its quality rationale.

Key activities include building low-fidelity prototypes and looking for strong demand evidence (paid pilots, pre-sales, or letters of intent) rather than positive opinions or survey answers (Furr and Ahlstrom, 2011). The aim is to close real commitments, in line with the logic of Customer Development (Blank, 2007).

Quality Gate 2 requires: (1) at least 40% of users saying they would feel "very disappointed" if the solution were no longer available(Maurya, 2012); (2) evidence of repeated use or early renewal within a 30-day window, or at least two referrals from pilot users (Maurya, 2012); (3) pilot payments or small contractual commitments (Kaulio, 2003); (Furr and Ahlstrom, 2011). The 40% figure was developed in a SaaS context and should be treated as a sectoral starting reference rather than a universal gate (Maurya, 2012). This gate is the most important Check point in the PDCA cycle: it separates genuine validated demand from declared interest, and prevents an early move to Stage 3 and the cash burn that would follow.

C. Stage 3: First Sales and Basic Financial Discipline (PDCA: Act)

Stage 3 follows the Act phase: the validated fit is turned into operational reality, first recurring sales and the basic financial discipline needed to navigate the valley of death. At the same time, the stage closes the PDCA loop: cash monitoring data and sales learning feed back into the next Plan cycle if market evidence calls for going back to an earlier stage.

Interview evidence motivating this stage: Separating Stage 3 as a distinct phase came directly from interview evidence. Founders consistently described the transition from validated demand to first revenue as the most difficult and least-supported phase, requiring specific practices (direct prospecting, objection handling, cash tracking) that were neither product development nor full-scale commercial activity. No existing framework addressed this transition in operational terms.

The stage uses the principle of founder-led sales (Kazanjy, 2020): the founding team takes responsibility for early selling as a discovery process, collecting objections, usage patterns, and adoption factors directly from customers. The focus is a well-defined early adopter niche (the segment with the most acute pain and the greatest willingness to tolerate the uncertainty of a new solution), approached through direct prospecting before any attempt to move into broader segments (Furr and Ahlstrom, 2011).

Commercial legitimacy is built on the founding team's social capital: prior contacts, reference customers whose adoption serves as a market signal to other actors, and participation in incubator and accelerator programmes (Spiegel *et al.*, 2016). These programmes serve two functions: they give access to networks and training, and they help reduce the liability of newness by giving the venture legitimacy that lowers information asymmetries with potential investors (Bellavitis *et al.*, 2020).

In parallel, basic financial management routines are introduced — among the earliest forms of process control available to a resource-constrained founder: cash and runway monitoring, and linking bootstrapping versus external financing decisions to concrete validation milestones rather than arbitrary calendar dates (Maurya, 2012); (Bhimani, 2017)

Quality Gate 3 requires: (1) first real sales, separate from free pilots (Kaulio, 2003); (2) regular runway monitoring with current figures known at all times (Bhimani, 2017); (3) at least 70% of discretionary spending going to validation and sales activities, reviewed monthly (Sethi, 2024).

D. Positioning Relative to Existing Frameworks

Table V maps the coverage of each existing framework against the critical early-stage dimensions. The coding scheme uses three levels. A full coverage mark (✓) is assigned when a framework provides explicit operational guidance for the dimension in question, that is, it specifies recommended practices, tools, or activities and includes at least one observable progress indicator or decision criterion. A partial coverage mark (~) is assigned when a framework addresses the dimension conceptually or implicitly but does not provide operational tools or thresholds. An absence mark (✗) is assigned when the dimension is neither addressed conceptually nor operationally within the framework's published scope. Coding was performed against the primary sources for each framework. (Ries, 2011); (Blank, 2007); (Osterwalder and Pigneur, 2010); (Shanbhag and Pardede, 2022); (Font-Cot *et al.*, 2025).

TABLE V. COVERAGE OF EXISTING FRAMEWORKS ACROSS CRITICAL EARLY-STAGE DIMENSIONS

Dimension	Lean Startup	Customer Development	Business Model Canvas	Blitz Canvas	Scaling Wheel	This study
Problem validation	✓	✓	~	✓	~	✓
Problem–solution fit	✓	✓	~	✓	~	✓
Team configuration and liability of newness	~	~	✗	~	✓	✓
Transition to first sales	✗	~	✗	✗	✗	✓
Financial discipline and cash management	✗	✗	✗	✗	~	✓
Commercial legitimacy building	~	~	✗	✗	~	✓
Sequential and iterative integration	~	~	✗	✓	✗	✓
PDCA / QM structure	~	✗	✗	✗	✗	✓

Four codings require explicit justification based on close reading of the primary sources.

(1) Blitz Canvas, Problem validation: ✓. Studying the User, Defining the Solution and 5 User's Feedback specify concrete operational activities (user personas, user stories, low-fidelity wireframe prototyping, and customer feedback collection) that constitute an operational problem validation process (Shanbhag and Pardede, 2022).

(2) Blitz Canvas, Sequential and iterative integration: ✓. The Blitz Canvas is explicitly process-driven, structured around ten sequential stages with defined inputs and outputs at each transition, and designed to iterate through feedback loop (Shanbhag and Pardede, 2022).

(3) Scaling Wheel, Problem validation: ~. Competitive Environment evaluates the urgency and scope of the problem a startup addresses, and Product/Service assesses whether a critical market need is met (Font-Cot *et al.*, 2025). However, these are diagnostic tools applied to existing ventures; the framework does not specify how to discover or validate a problem from scratch.

(4) Scaling Wheel, Sequential and iterative integration: ✗. The Scaling Wheel is designed as a simultaneous five-module diagnostic wheel, not a sequential process (Font-Cot *et al.*, 2025). The framework's own documentation states it is a holistic assessment tool rather than a step-by-step process.

Two patterns emerge from Table V. Lean Startup and Customer Development give the strongest validation coverage but leave the transition to first sales and financial discipline unaddressed, and Customer Development's assumptions about team size and resources do not fit the bootstrapped context studied here. The Business Model Canvas maps hypotheses without testing them, while the process-oriented frameworks (Blitz Canvas and Scaling Wheel) cover sequencing or diagnosis but not the path from validation to first sales under cash constraints

VI. DISCUSSION AND IMPLICATIONS

The proposed stage-based methodology helps address several gaps found in the literature and in quality management practice:

First, it complements approaches such as Lean Startup (Ries, 2011) and Customer Development (Blank, 2007) by keeping their focus on validated learning while adding dimensions these frameworks tend to leave aside: financial discipline, the strategic use of social capital, the building of commercial legitimacy, and a quality management structure. With this integration, early-stage validation becomes a wider process of organisational configuration aimed at navigating the valley of death, going beyond the testing of product and market hypotheses. The proposed methodology also differs in kind from tools like the Business Model Canvas (Osterwalder and Pigneur, 2010) or the Scaling Wheel (Font-Cot *et al.*, 2025). Those are diagnostic tools: they help assess an existing model. Our methodology is prescriptive: it guides founders step by step from problem discovery to first revenue. The absence of such a process/prescriptive artefact that bridges validated learning and first sales under cash constraints is precisely the gap this study addresses.

Second, the PDCA framing makes the quality management contribution of the methodology clear and verifiable. The methodology is grounded in quality-management logic by adapting three core functions (quality planning, quality control and quality assurance) to the early-stage startup context. Quality planning occurs in Stage 1, where the problem and its validation criteria are defined before any resource is committed. Quality control occurs at each quality gate, where observable evidence is inspected against predefined thresholds) the startup equivalent of an in-process inspection point. Quality assurance comes from the iterative pivot logic. It stops ventures built on weak demand evidence from moving to the next stage, where failure would be more costly (Cooper, 1990). Survey 1's finding that most founders are unfamiliar with process management certifications shows that this quality discipline is both needed and missing in current practice.

Third, the methodology works as a specialisation of the Dynamic Integrative Framework derived from the systematic review, turning its four strategic domains into concrete practices linked to three key decisions: which problem to solve, which minimal solution to build, and how to sell it without running out of cash. The systematic review identifies the development of empirically validated holistic frameworks as the second most frequently proposed future research direction (n=17 studies), behind only replication across different contexts (n=21); this work moves in that direction by offering a structured proposal based on triangulated evidence.

For incubators and accelerators, the methodology suggests early-phase itineraries that combine problem discovery, MVP/MXP design, founder-led sales, and basic financial management. The quality gate structure also gives incubators a standard tool for tracking several ventures at the same time.

VII. Conclusions

This paper has addressed the gap between the dominant frameworks for early-stage startups and the factors that the evidence links to effective survival, and has done so within an explicit quality management structure. Based on the triangulation of an ongoing systematic literature review (37 primary studies), 30 semi-structured founder interviews, and two surveys conducted in 2024 and 2026, a stage-based methodology has been proposed for the idea, validation, and launch phases.

The methodology has three stages, structured as an applied PDCA cycle (discovery and validation of a severe and monetisable problem (Plan); design and testing of a minimal solution with strong demand evidence (Do); and generation of first sales with basic financial discipline and commercial legitimacy building (Act)) each bounded by explicit quality gates that apply the Check logic and prevent moving forward too early. *Table III* shows how each methodological choice is grounded in the three evidence sources, with the interview data playing a direct and traceable role in separating Stage 3 and in adding the MXP option and founder-led sales. The main contribution is turning fragmented academic knowledge into a quality-grounded, actionable roadmap for founders and incubators, aligning market validation with organisational and financial configuration within a PDCA–stage-gate structure.

Limitations

While this study provides an actionable initial framework, several limitations must be acknowledged. First, the methodology relies on a formative DSR design with data restricted to the Spanish entrepreneurial ecosystem, utilizing small, non-probabilistic survey samples (n = 17) as illustrative cross-sections. Second, evaluating the artefact using the same data sources employed for its design introduces potential bias, although it maintains the internal coherence necessary for early-stage design science.

On the other hand, the interview sample is predominantly composed of ventures that survived the earliest stages and achieved traction or operational stability. Their retrospective accounts may therefore underrepresent failure pathways and introduce survivorship bias. Future longitudinal validation should include ventures from inception, including those that pivot, stall or cease operations.

Future Research Directions

Future research should undertake a summative longitudinal evaluation of the proposed artefact through multiple case studies of early-stage startups. The initial evaluation should be conducted in the Spanish entrepreneurial ecosystem, where resource scarcity and early abandonment make the proposed quality-gate logic especially relevant. Using the indicators defined in Table IV, the study should examine whether ventures applying the roadmap demonstrate clearer stage-progression decisions, earlier paid sales, more systematic runway monitoring and more evidence-based pivot decisions than comparable ventures not explicitly applying the roadmap. Subsequent studies should evaluate transferability across sectors and across European and Latin American entrepreneurial ecosystems.

Acknowledgment

This work is part of the TranSformIT project (PID2024-157982OB-I00), funded by MICIU/AEI/10.13039/501100011033, and the DESAFÍO-CM project (TEC-2024/COM-235), funded by the Comunidad de Madrid.

References

Aldrich, H.E. (1999), Organizations Evolving., Sage, London.

Argaw, Y.M. and Liu, Y. (2024), "The Pathway to Startup Success: A Comprehensive Systematic Review of Critical Factors and the Future Research Agenda in Developed and Emerging Markets", Systems, 1 December, doi: 10.3390/systems12120541.

Asmoro, A., Nugroho, L.E. and Sulistyo, S. (2018), "Preliminary Comparative Study of Innovation Process Model for Startup in Indonesia", IOP Conference Series: Materials Science and Engineering, Vol. 384, doi: 10.1088/1757-899X/384/1/012057.

Bellavitis, C., Hillenbrand, C., Isaksson, R. and McNaughton, R.B. (2020), New Frontiers in Entrepreneurial Fundraising., Springer Nature.

Bhimani, A. (2017), Financial Management for Technology Start-Ups: A-Z Guidance for the Innovator., Kogan Page.

Blank, S.G. (2007), "The Four Steps to the Epiphany", Cafepress.

Brown, T. (2008), "Design thinking", Harvard Business Review, Vol. 86 No. 6, doi: 10.47381/aijre.v22i2.611.

Chorev, S. and Anderson, A.R. (2006), "Success in Israeli high-tech start-ups; Critical factors and process", Technovation, Vol. 26 No. 2, pp. 162–174, doi: 10.1016/j.technovation.2005.06.014.

Cooper, R.G. (1990), "Stage-gate systems: A new tool for managing new products", Business Horizons, doi: 10.1016/0007-6813(90)90040-I.

Fernández-Laviada, A., Calvo, N., Monje-Amor, A. and Atrio Lema, Y. (2025), Global Entrepreneurship Monitor. Informe GEM España 2024-2025, Global Entrepreneurship Monitor. Informe GEM España 2024-2025, Editorial de la Universidad de Cantabria, doi: 10.22429/euc2025.012.

Fitzpatrick, R. (2013), The Mom Test: How to Talk to Customers & Learn If Your Business Is a Good Idea When Everyone Is Lying to You, CreateSpace Independent Publishing Platform.

Font-Cot, F., Lara-Navarra, P., Arnau, C.S. and Serradell-Lopez, E. (2025), "Scaling Wheel Framework: A holistic approach to startup scalability, governance models, and AI-driven innovation ecosystem competitiveness", Journal of Open Innovation: Technology, Market, and Complexity, Vol. 11 No. 3, doi: 10.1016/j.joitmc.2025.100619.

Furr, N. and Ahlstrom, P. (2011), Nail It Then Scale It: The Entrepreneur's Guide to Creating and Managing Breakthrough Innovations., NISI Institute.

Giardino, C., Paternoster, N., Unterkalmsteiner, M., Gorschek, T. and Abrahamsson, P. (2016), "Software Development in Startup Companies: The Greenfield Startup Model", IEEE Transactions on Software Engineering, Vol. 42 No. 6, pp. 585–604, doi: 10.1109/TSE.2015.2509970.

González, L. (2023), "Socialpreneurs Podcast".

González, L. (2026), "De Idea a Impacto Podcast".

Gregor, S. and Hevner, A.R. (2013), Positioning and Presenting Design Science Research for Maximum Impact., doi: https://doi.org/10.25300/MISQ/2013/37.2.01.

Hejazi, S.R., Seyyedamiri, N., Shokouhi Bidhandi, M. and Rasoulian, P. (2024), "Identifying Factors Affecting Startups Survival: A Systematic Literature Review", Journal of Organisational Studies and Innovation, Management & Business Academy, Vol. 19 No. 30, pp. 31–64, doi: 10.51659/josi.23.196.

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004), "Design science in information systems research. ", MIS Quarterly, Vol. 28 No. 1, pp. 75–105.

Hokkanen, L., Kuusinen, K. and Väänänen, K.A. (2016), Minimum Viable User Experience: A Framework for Supporting Product Design in Startups, Lecture Notes in Business Information Processing, Vol. 251, doi: 10.1007/978-3-319-33515-5_6.

Jing, R. (2023), "Analyzing How Competitor Alliances Affect New Venture Milestones using a Cox Model", Journal of Enterprise and Business Intelligence, pp. 115–125, doi: 10.53759/5181/jebi202303012.

Kaulio, M.A. (2003), "Initial conditions or process of development? Critical incidents in the early stages of new ventures", R and D Management, Vol. 33 No. 2, pp. 165–175, doi: 10.1111/1467-9310.00290.

Kazanjy, P. (2020), Founding Sales: The Early Stage Go-to-Market Handbook, Holloway.

Kitchenham, B. (2004), "Procedures for Performing Systematic Reviews Kitchenham, B., 2004.", Keele, UK, Keele University, Vol. 33 No. 2004.

Maurya, A. (2012), Running Lean: Iterate from Plan A to a Plan That Works., O'Reilly Media.

Nosella, A., Forza, C., Workalemahu, R.N. and Molaro, G.I. (2025), "Framing startups survival research at the crossroad of entrepreneurship, strategy and business model: a literature review and a research agenda", International Journal of Innovation Studies, KeAi Publishing Communications Ltd., 1 June, doi: 10.1016/j.ijis.2025.05.003.

Osterwalder, A. and Pigneur, Y. (2010), Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers, Booksgooglecom, Vol. 30, Wiley.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., et al. (2021), "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews", BMJ, BMJ Publishing Group, 29 March, doi: 10.1136/bmj.n71.

Pattyn, F. (2023), "Enhancing Startup Success Rates: Towards a Pragmatic Framework for Product Managers (PFPM)", Proceedings of the IEEE International Conference on Requirements Engineering, Vol. 2023-September, pp. 381–385, doi: 10.1109/RE57278.2023.00059.

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S. (2007), "A design science research methodology for information systems research. ", Journal of Management Information Systems, Vol. 24 No. 3, pp. 45–77.

Ries, E. (2011), The Lean Startup : How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses, The Lean Startup : How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses, Crown Business.

Ross, A. and Lemkin, J. (2016), From Impossible To Inevitable. How HyperGrowth Companies Create Predictable Revenue, edited by John Wiley & Sons Inc.

Sethi, A. (2024), From Startup to Unicorn: An Essential Guide to Build, Scale and Sustain Value for Platform and Tech Startups, From Startup to Unicorn: An Essential Guide to Build, Scale and Sustain Value for Platform and Tech Startups, Springer Nature, doi: 10.1007/978-3-031-53894-0.

Shanbhag, N. and Pardede, E. (2022), "The Blitz Canvas: A Business Model Innovation Framework for Software Startups", Systems, Vol. 10 No. 3, doi: 10.3390/systems10030058.

Shi, Y., Eremina, E. and Long, W. (2024), "Machine learning models for early-stage investment decision making in startups", Managerial and Decision Economics, Vol. 45 No. 3, pp. 1259–1279, doi: 10.1002/mde.4072.

Sosa Gómez, I. (2025), "Navigating uncertainty: a strategic framework for early-stage entrepreneurial; Marco estratégico de emprendimientos para la creación de startups en la fase inicial", European Public and Social Innovation Review, Vol. 10, doi: 10.31637/epsir-2025-1115.

Spiegel, O., Abbassi, P., Zylka, M.P., Schlagwein, D., Fischbach, K. and Schoder, D. (2016), "Business model development, founders' social capital and the success of early stage internet start-ups: a mixed-method study", Information Systems Journal, Vol. 26 No. 5, pp. 421–449, doi: 10.1111/isj.12073.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Governing the Angenic Shift: A Quality Governance Framework for Sovereign AI in Industrial Organizations

Eduardo L. García-Garcia[1]

[1] Factory Management Institute, NPO | EMPWR
Corresponding author: eduardo.garcia@factorymanagementinstitute.com

***Abstract*— The dominant narrative surrounding industrial AI deployment frames the human role as threatened by automation. This paper argues the opposite: the organizations that will lead in the autonomous AI era are not those that deploy the most capable models, but those that develop the most rigorous internal governance capacity — elevating human practitioners from task supervisors to Sovereign Architects of autonomous systems.**

We introduce the Jishu-ken G Framework, a three-phase sociotechnical maturity model providing a structured governance path from AI-assisted operations (Lean State) through bounded automation (Agentic State) to self-governing agent ecosystems (Angenic State). The framework is grounded in the Jishu-ken study group — the same self-directed cognitive mechanism through which post-war Japan generated the Toyota Production System from within — as the organizational engine capable of building sovereign AI governance architecture without external dependency. Organizations that outsource this capability accumulate Sovereignty Debt: the progressive erosion of the capacity to govern their own automated systems — the structural risk the framework is designed to eliminate.

A Design Science Research methodology produces two governance instruments of increasing scope, preceded by a Lean prerequisite. The Lean State requires no new artifact: classical Lean methodology — value stream mapping, waste identification, standardization — applied rigorously to the data and process layer produces the Golden Record, the deterministic data substrate that grounds probabilistic AI in empirical truth before any autonomous delegation begins. The AI Agent Integrity Canvas then governs the individual agent at the Agentic State, encoding human-defined sovereign boundaries as deterministic Poka-Yoke constraints. The Angenic Governance Blueprint governs the ecosystem at the Angenic State: a constitutional document encoding sovereign intent in a form simultaneously human-readable and machine-verifiable.

Keywords – Quality 4.0; Quality 4.0; AI Governance; Jishu-ken; Sovereign Architect; Design Science Research; Industry 5.0; Operational Excellence.

VIII. Introduction: The Governance Crisis Behind the AI Productivity Paradox

In 2025, McKinsey estimated that fewer than 5% of companies deploying generative AI had captured meaningful returns on their investment. The paradox is striking: organizations have access to the most powerful automation tools in history, yet most report that their processes are not materially faster, their decisions not materially better, and their people not materially more productive.

The standard diagnosis blames change management, data quality, or organizational culture. These factors are real, but they miss the structural root cause: organizations are deploying autonomous AI agents inside governance architectures designed for deterministic software tools. A tool that executes precisely what it is told requires a supervisor. An agent that reasons, decides, and acts autonomously requires an **architect**. Most organizations still have supervisors where they need architects — and no map for how to make the transition.

This paper provides that map. We introduce the **Jishu-ken G Framework**, a three-phase evolutionary model that takes organizations from their current state — what we call the Lean State, where AI assists human decision-making — through the Agentic State, where AI executes bounded autonomous tasks, to the **Angenic State**, where ecosystems of AI agents self-orchestrate within human-defined sovereign boundaries. The word "Angenic" is a deliberate coinage from "Agentic" + "Genic" (originating from), signalling the shift from AI that performs tasks to AI that generates the conditions for its own coordinated evolution.

A. 1.1 The Governance Gap

The Governance Gap is the organizational inability to translate strategic intent into machine-readable constraints that remain valid as AI autonomy increases. It is fundamentally a crisis of human abstraction capacity: moving from Human-in-the-Loop (execution oversight) to Human-on-the-Loop (architectural oversight) requires cognitive capabilities that most organizations have never been asked to develop. When they lack this capacity internally, they outsource it — creating what we call **Sovereignty Debt**: the progressive erosion of an organization's ability to govern its own automated systems from within.

Sovereignty Debt, in this context, is not merely a governance failure. It is a strategic position: the difference between an organization that constitutes its own autonomous systems and one that consumes them.

B. 1.2 Why Existing Frameworks Are Insufficient

Existing AI governance frameworks are designed to govern AI as a tool — a discrete system with defined inputs and outputs — rather than as an agent capable of autonomous, emergent behavior. Sociotechnical Systems (STS) theory provides the

closest conceptual foundation, but STS was developed for deterministic industrial systems, not probabilistic AI agents. Our framework extends STS theory in three directions: (1) a maturity model mapping evolving governance requirements at each stage of AI autonomy; (2) operational artifacts that translate strategic intent into machine-executable constraints; and (3) redefining the human role from task supervisor to **Sovereign Architect**.

The framework reconceptualises AI governance not as problem resolution but as value-space constitution.

A further limitation of existing frameworks is temporal: they govern AI systems as they exist today, not as they will exist in three or five years. The Jishu-ken G Framework addresses this gap explicitly — its value is not contingent on the Angenic State being fully achievable in the near term, but on providing the architectural grammar that ensures today's infrastructure decisions do not foreclose tomorrow's governance options.

The framework thus reconceptualises AI governance not as problem resolution but as value-space constitution: defining what the organization's autonomous systems are permitted to generate, rather than merely constraining what they are prohibited from doing.

IX. 2. Theoretical Foundations: From Sociotechnical Misalignment to Sovereign Orchestration

The Jishu-ken G Framework is grounded in three converging theoretical traditions: Sociotechnical Systems theory, the Resource-Based View of competitive advantage, and cybernetic control theory.

A. 2.1 The Three Theoretical Propositions

TP1 — Sociotechnical Stability as a Prerequisite for Autonomy. Following STS theory (Trist & Bamforth, 1951) (Pasmore, Winby, Mohrman, & Vanasse, 2019), we posit that deploying agentic AI without first establishing a stable digital foundation induces systemic fragility. Viability in autonomous systems — in the sense formalized by Ivanov and Dolgui (2020) for intertwined supply networks — requires what we call a "Golden Record" — a deterministic data substrate that grounds probabilistic AI in empirical truth before autonomous delegation begins.

TP2 — Cognitive Sovereignty and Knowledge Encapsulation. According to the Resource-Based View (Barney, 1991) (Teece, Pisano, & Shuen, 1997), the transition from AI assistance to AI autonomy requires internalizing the capability to encode operational expertise as machine-readable logic. Organizations that outsource this capability create Sovereignty Debt.

TP3 — The Angenic Shift as Architectural Orchestration. The final evolution toward the Angenic State requires the simultaneous cultivation of *Monozukuri* (process excellence) and *Hitozukuri* (human capability development). This dual mandate, derived from the practitioner tradition of Jishu-ken as a self-directed cognitive laboratory (Garcia-Garcia, 2026), enables the Angenic Shift: the point at which human architects govern system integrity while AI agents govern system execution.

B. 2.2 The Jishu-ken G Roadmap: Three Evolutionary States

The framework maps three evolutionary states along two parallel trajectories: human cognitive capability (Operator → Supervisor → Architect) and AI autonomy (Co-pilot → Agent → Ecosystem). The core mechanism driving both simultaneously is **Reciprocal Cognitive Escalation**: machine autonomy cannot safely advance unless human governance capability advances in lockstep (see Table 1 and Figure 1).

Dimension	Lean State	Agentic State	Angenic State
Human Role	Operator / Co-pilot	Supervisor / Agent Owner	Sovereign Architect
AI Role	Assisted Intelligence	Encapsulated Autonomy	Recursive Ecosystem
Core Challenge	Data hygiene & Golden Record	Expertise encapsulation & Full Limits	Sovereign orchestration & Flywheel
Governance Artifact	Lean Tools: Process mapping, Standardization, Waste elimination, Problem solving, etc.	AI Agent Integrity Canvas	Angenic Governance Blueprint
Failure Mode	Data fragmentation / shadow AI	Hallucinating agents / Sovereignty Debt	Agentic silos / strategic drift

Table 1. The Cognitive Escalation Ladder — Roles, artifacts, and failure modes across the three states.

This Reciprocal Cognitive Escalation operationalizes what Kolbjørnsrud (2023) terms the substitution principle of human-AI collaboration: as AI assumes routine cognitive tasks, human capacity is progressively redirected toward architectural judgment — the governance of boundaries rather than the execution of tasks.

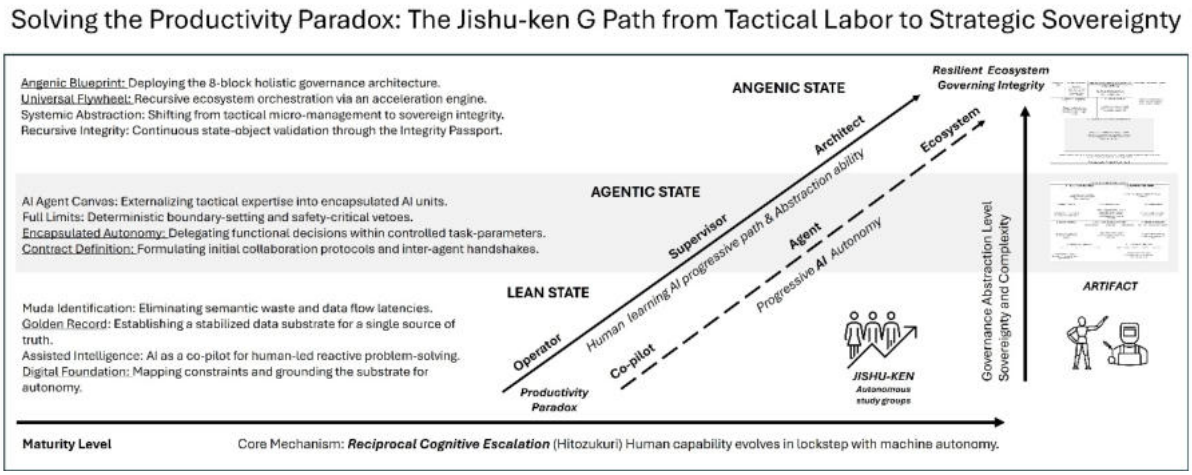


Figure 1: The Jishu-ken G Roadmap — A Decision Aid Model for scaling abstraction in autonomous systems.

This governance architecture directly addresses the capability gaps identified by Jackson et al. (2024) in their analysis of generative AI implementation in supply chain and operations management — specifically the absence of deterministic constraint frameworks for probabilistic agent behavior.

C. 2.3 Jishu-ken as Organizational Ambidexterity Engine

Traditional operations management literature treats Jishu-ken as either a knowledge-creation mechanism (Nonaka and Takeuchi 1995) or a management-led kaizen activity. A third, more radical interpretation — developed in full in a companion paper (Garcia-Garcia, 2026) — presents Jishu-ken as presented here: an amorphous, self-directed cognitive laboratory with no fixed form or externally imposed deadline, whose primary objective is to force practitioners to think for themselves — what Argyris and Schön (1997) would recognize as double-loop learning: not merely correcting agent outputs within existing governance rules, but developing the capacity to revise the rules themselves. This is not the first time that Jishu-ken study groups have served as the organizational engine of a technological paradigm shift. The self-directed, practitioner-led study group was the connective tissue through which post-war Japan transformed Deming's statistical quality principles into the Toyota Production System, Total Quality Control, and what the West would later label the Lean Ecosystem — in a nation economically devastated, operating under radical resource constraints, facing simultaneous pressure to rebuild industrial capacity and close a massive technological gap. The urgency was existential; the solution was not imported expertise but internalized capability. The parallel with the current moment of AI disruption is not decorative: the same organizational mechanism that generated the TPS, TQC, and TQM revolutions from within is precisely the mechanism capable of generating the governance architecture for the Angenic transition.

This model operationalizes organizational ambidexterity (March, 1991) by simultaneously advancing Exploitation (*Monozukuri*: process standardization) and Exploration (*Hitozukuri*: human capability development). Behavioral meta-analyses confirm that process goals — learning how to govern, rather than optimizing a specific outcome — significantly increase self-efficacy (d=1.11), precisely the cognitive conditions needed for the Human-on-the-Loop governance posture required in the Angenic State.

X. 3. Research Methodology: Design Science and Three-Shot Triangulation

This study adopts a Design Science Research (DSR) approach (Hevner, 2004) (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007), well-suited to the Jishu-ken G Framework because the research question is prescriptive: not "what is happening?" but "how should organizations govern the transition to autonomous AI?"

A. 3.1 The Governance Artifacts

The framework's two governance instruments are designed as "Boundary Objects" (Star & Griesemer, 1989): structured representations that bridge the semantic gap between domain experts and AI engineers, translating operational knowledge into machine-executable constraints without requiring technical expertise from domain practitioners.

The Lean Stage requires no new instrument. Classical Lean methodology — value stream mapping, waste identification, standardization — applied rigorously to the data and process layer already produces the prerequisite for autonomous delegation: the Golden Record, a deterministic data substrate that grounds probabilistic AI in empirical truth. Consider a company deploying an LLM-based copilot to assist engineers writing FMEAs — as ASML is currently doing for its Extreme Ultraviolet lithography machines using ChatGPT (Dillen & Spaan, 2025): if the underlying failure mode taxonomy is inconsistent across engineering teams, the copilot generates plausible but context-incorrect outputs. Lean process discipline eliminates that semantic waste before any agentic layer is deployed. Without a Golden Record, any agent built on top will hallucinate, because its probabilistic reasoning has no empirical ground truth to anchor against.

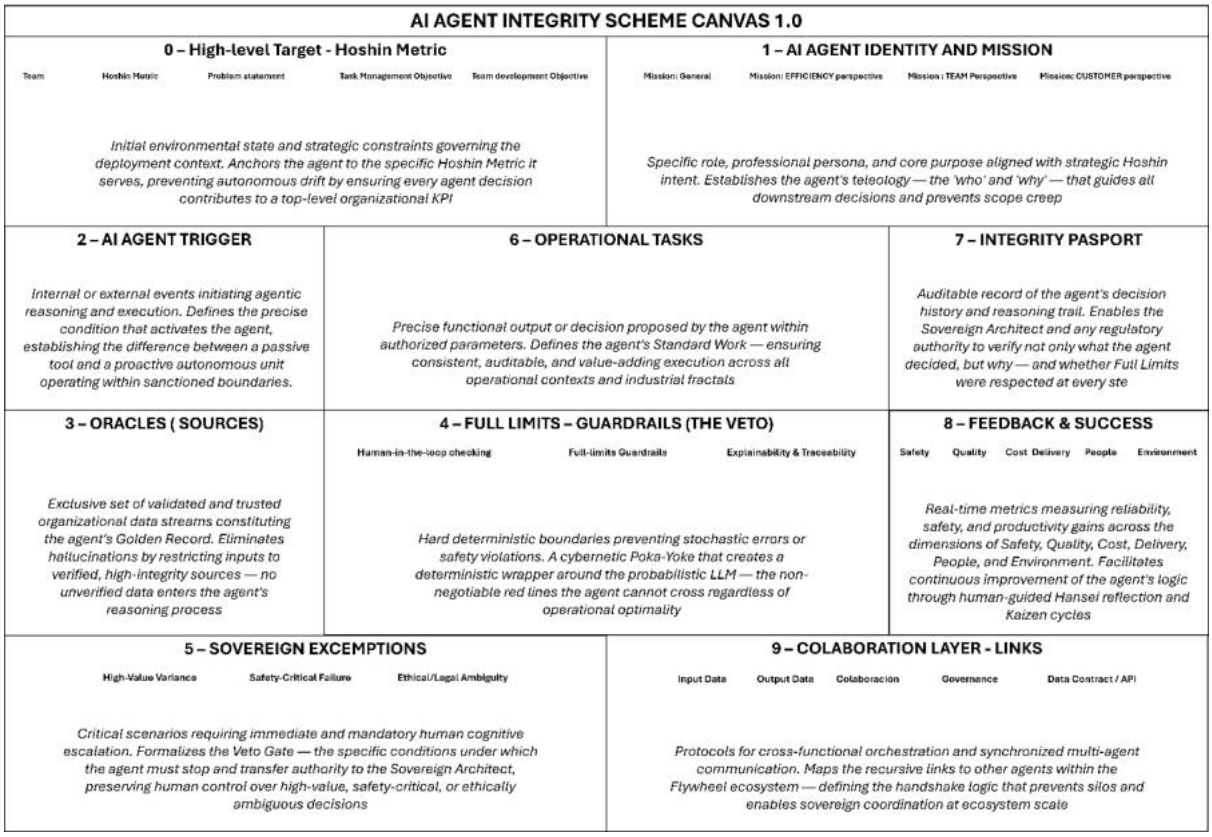


Figure 2: AI Agent Integrity Canvas 1.0 — Eight-block governance structure for autonomous agent deployment.

The AI Agent Integrity Canvas (Agentic Phase): A deterministic wrapper for probabilistic agents that operationalizes, at the individual agent level, the three pillars of Industry 5.0 identified by Ivanov (2023): human-centricity (Full Limits encoding human-defined sovereign boundaries), resilience (hard constraint gates preventing stochastic drift), and viability (the governance conditions under which autonomous delegation remains safe over time). Its central innovation is the "Full Limits" structure — a set of hard, non-negotiable constraints (Poka-Yoke logic gates) that bound an agent's reasoning regardless of what its underlying model might generate. The Canvas operates as both a design tool (defining an agent before deployment) and an audit tool (verifying that an agent in operation remains within its sovereign bounds) (See Figure 2).

The Angenic Governance Blueprint (Angenic Phase): A system-of-systems topology that maps the Universal Flywheel (Collins, 2001) (Bezos, 2001)— the self-reinforcing loop through which every agentic interaction feeds into a Structured Knowledge Core (SKC), progressively transforming operational volume into structural intelligence. The Blueprint's critical innovation is the Collaboration Contract (Block 6): formalized inter-agent handshake protocols.

In practice, Blocks 0–5 of the Blueprint require direct human expertise and strategic abstraction. Blocks 6 and 7, by contrast, are amenable to AI-assisted development. Current large language models — including those used in the preparation of this paper — demonstrate sufficient contextual reasoning capacity to interpret and populate Blocks 6 and 7 from a completed Flywheel specification. Block 3, however, consistently requires human causal judgment: no current LLM can substitute for the practitioner's direct knowledge of the operational ecosystem being governed.

B. 3.2 Validation Strategy: Empirical and Innovation Triangulation

The validation strategy distinguishes between two levels of evidence, reflecting the paper's dual contribution of an established framework (Lean and Agentic stages) and a prospective architectural model (Angenic stage).

Empirical Longitudinal Validation (TP1 and TP2): The Microsoft Commercial Operations case (2023–2025) provides longitudinal evidence that industrial AI ROI requires sequenced foundations: first, a Lean cleanup to establish the Golden Record; second, agentic encapsulation of complex workflows. Investigative reporting by Bousquette and Lin (2025) in The Wall Street Journal documents how Microsoft's AI transformation required precisely this sequenced approach.

Innovation Triangulation (TP3): To validate structural invariance of the Angenic architecture, we apply the Governance Blueprint across three divergent industrial substrates: (A) Microsoft's projected evolution toward a Zero-Latency / Zero-Onboarding ecosystem; (B) OECD/ITF Urban Last-Mile Logistics (ITF 2024); and (C) Aerospace Production Planning with integrated Plan+MES+PostSales+Maintenance.

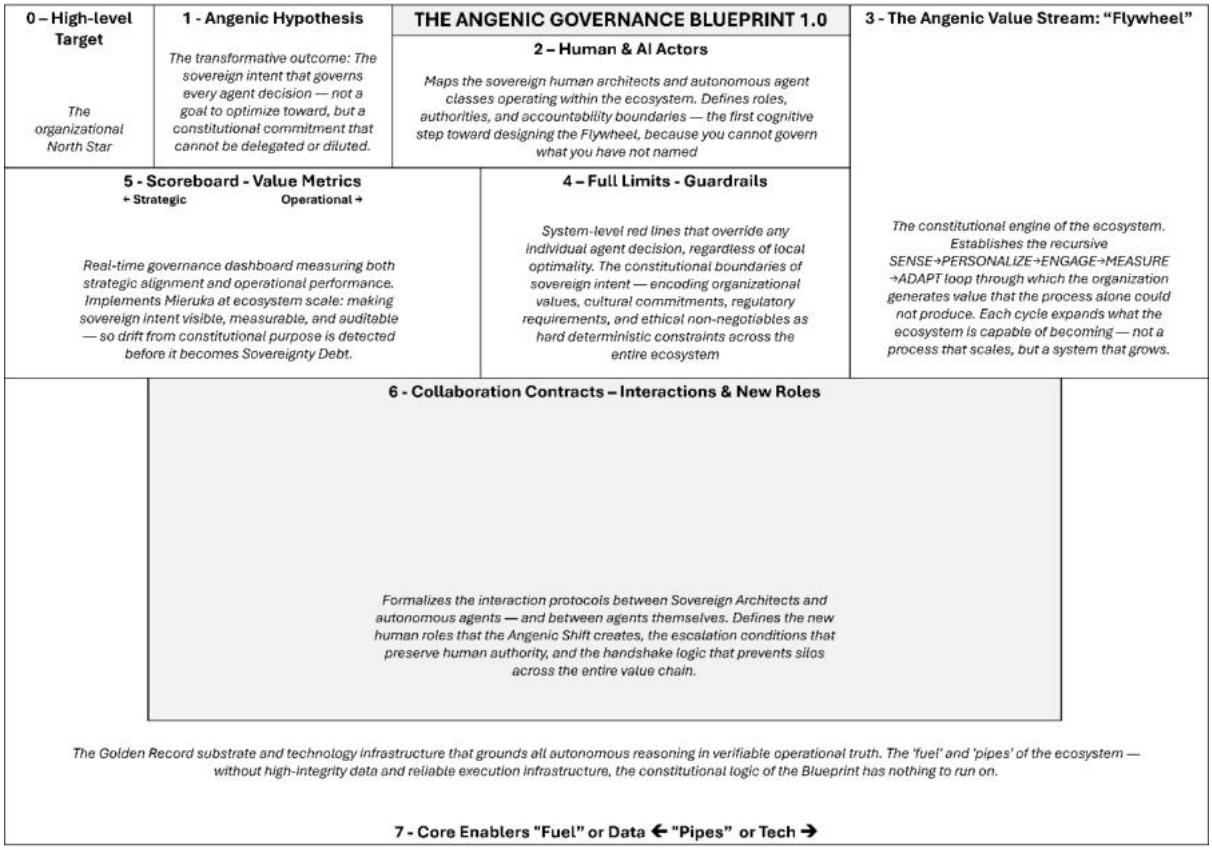


Figure 3: The Angenic Governance Blueprint 1.0 — Blank template.

C. 3.3 Artifact Genesis: An Inductive DSR Development Process

The governance artifacts emerged inductively through an iterative development process that constitutes an integral part of the DSR methodology. The AI Agent Integrity Canvas originated in an applied experiment automating a marketing agent using the KJ affinity method. Three successive Canvas versions were developed, each refined through an adversarial dialogue with an LLM asked to identify missing governance elements — a process of LLM-as-validator that proved methodologically significant: the model's ability to correctly interpret the Canvas and generate improvement suggestions served as an empirical test of artifact coherence and operational completeness.

The Angenic Governance Blueprint followed a parallel inductive path. Two Blueprint iterations were developed for the same autonomous marketing 24/7 system case before the framework achieved sufficient stability for more complex environments. The Flywheel — Block 3, the most cognitively demanding element — was conceived at the intersection of Amazon's value generation flywheel logic and the requirements of a fully autonomous 24/7 value creation system. Its structural invariance was not hypothesized in advance but discovered empirically through progressive validation across three environments of increasing entropy: ITF last-mile logistics, Microsoft commercial operations, and the Aerospace Plan+MES architecture. The fractal MVIT architecture emerged inductively from observing Tier-1/2/3 supplier structures in aerospace and automotive supply chains, rather than from theoretical design. All intermediate artifacts — Canvas versions, Blueprint iterations, and LLM validation dialogues — are retained by the author and available upon request.

The artifact development process combined three intellectual traditions: the KJ affinity method (Kawakita, 1967) governed the divergent brainstorming sessions; the Business Model Canvas (Osterwalder & Pigneur, 2010) provided the design principle of capturing a complex governance system in a single legible page; and the Amazon Flywheel logic — encountered as a conceptual convergence rather than a deliberate reference — informed the self-reinforcing loop at the heart of the Blueprint. Initial co-design sessions with Marco Cachi García, a software engineer with direct production experience building AI agents in Make and similar AI platforms, proved decisive: his practitioner knowledge of real agent behaviour in uncontrolled

environments shaped the Full Limits architecture and the Trigger taxonomy that no purely theoretical design process would have reached. The critical conceptual breakthrough — that the Angenic Flywheel must generate value that the process alone cannot produce, not merely optimize an existing process — emerged through sustained adversarial interaction with LLMs over approximately 47 days, in which the authors refused incremental solutions until a structurally disruptive answer was found. A final discovery confirmed the dual legibility property of the Blueprint: when presented with only the Flywheel block, current frontier LLMs were able to autonomously generate the remaining governance blocks with structural coherence — demonstrating that the Blueprint encodes sovereign intent in a form that is not merely human-readable but machine-generative. This property does not reduce the human authorship requirement: the Flywheel itself, particularly for high-entropy

environments such as the ITF and aerospace cases, requires cross-domain abstraction capacity that no LLM in this study generated independently.

XI. Multi-Domain Validation: Three Substrates, One Governance Logic

The following three cases demonstrate the structural invariance of the Jishu-ken G Framework. While the operational context, physical constraints, and strategic objectives differ radically, the governance logic — the sequencing of agents, the role of the Flywheel, the architecture of sovereign vetoes — remains consistent (see Table 2).

Dimension	Case A — Microsoft	Case B — ITF Logistics	Case C — Aerospace
Substrate	Service ("Bits")	Physical ("Atoms")	Industrial ("Fractals")
Strategic Intent	14-day onboarding → Zero State	Max Dynamic Density, Zero LEZ violations	Pull-based production, zero-defect token lifecycle
Core Constraint	Regulatory/compliance logic	Physical city limits (LEZ, loading bays)	FAA/EASA certification & Tier-N dependencies
Key Innovation	AVA as Poka-Yoke	Integrity Passport as sovereignty veto	SUP-MVIT Dualism: strategic kernel + sovereign token
Angenic Result	Triumphant Disappearance of manual onboarding	Zero autonomous regulatory violations	Pull-by-Request fractal lifecycle governance

Table 2. Three-shot triangulation strategy — structural invariance across divergent industrial substrates.

A. 4.1 Case A: Microsoft Commercial Operations — The Zero Onboarding Ecosystem

Prior to transformation, Microsoft's commercial partner onboarding involved more than 230 steps — a manual-heavy, multi-team workflow requiring close coordination across sales, marketing, product, and finance functions. The Lean phase did not begin with AI: it began with process visibility. Under COO Carolina Dybeck Happe, Microsoft applied Kaizen principles to map what the process actually looked like — not what stakeholders believed it looked like — and reduced it from 230 to fewer than 40 steps (Bousquette & Lin, 2025). This is the Lean State in its most precise operational form: eliminating semantic waste before any autonomous delegation is attempted. The output of this phase was a unified process substrate — what this framework terms the Semantic Knowledge Core (SKC) — integrating CRM, compliance APIs, and sanctions databases into a single Golden Record. Without this consolidation, any agentic layer built on top would hallucinate.

The Agentic phase automated approximately 75% of the remaining steps using AI agents governed by the AI Agent Integrity Canvas (Bousquette & Lin, 2025). The Canvas ensure that the agent operate within hard constraint gates — financial limits, data sovereignty rules, and brand safety parameters — functioning as the system's Poka-Yoke: no autonomous action could proceed that violated the governance boundaries encoded in the Blueprint. The combined result of the Lean and Agentic phases reduced the onboarding cycle from days to hours. "In the end, what came out was a phenomenally better experience for the customers as well as for our teams," Dybeck Happe reported.

The projected Angenic state does not merely accelerate the existing cycle — it eliminates the cycle entirely through a Predictive Flywheel architecture that achieves near-zero onboarding latency. This architecture deploys four specialized agents: the ISA (Intent Sensing Agent) for continuous upstream signal ingestion — tender activity, consumption patterns, partnership ecosystem dynamics —; the VAA (Value Adoption Agent) for proactive contract drafting initiated before the partner formally requests onboarding; the AVA (Adversarial Verification Agent) as the system's Poka-Yoke, a hard constraint gate that halts any VAA proposal violating financial limits, data sovereignty rules, or brand safety parameters; and the ROA (Resource Orchestration Agent) for execution. The human Relationship Architect no longer manages a process; they govern a sovereign boundary, intervening only when the AVA flags a "Gray Zone" ambiguity that falls outside pre-encoded parameters. This is the precise operational definition of Human-on-the-Loop governance: the human is not removed from the system — they are elevated to its constitutional layer.

Practitioner note about Figure 4: Blocks 0–5 represent the critical governance design decisions requiring deep operational knowledge and strategic abstraction — particularly Block 3 (The Angenic Value Stream: Flywheel), which demands a thorough understanding of the agent ecosystem to be governed. Blocks 6 (Collaboration Contracts) and 7 (Core Enablers) are highly amenable to AI-assisted development, as demonstrated in this paper. The Angenic Flywheel remains the single most cognitively demanding artifact, requiring the practitioner to reason simultaneously about agent interactions, sovereignty boundaries, and systemic circularity.

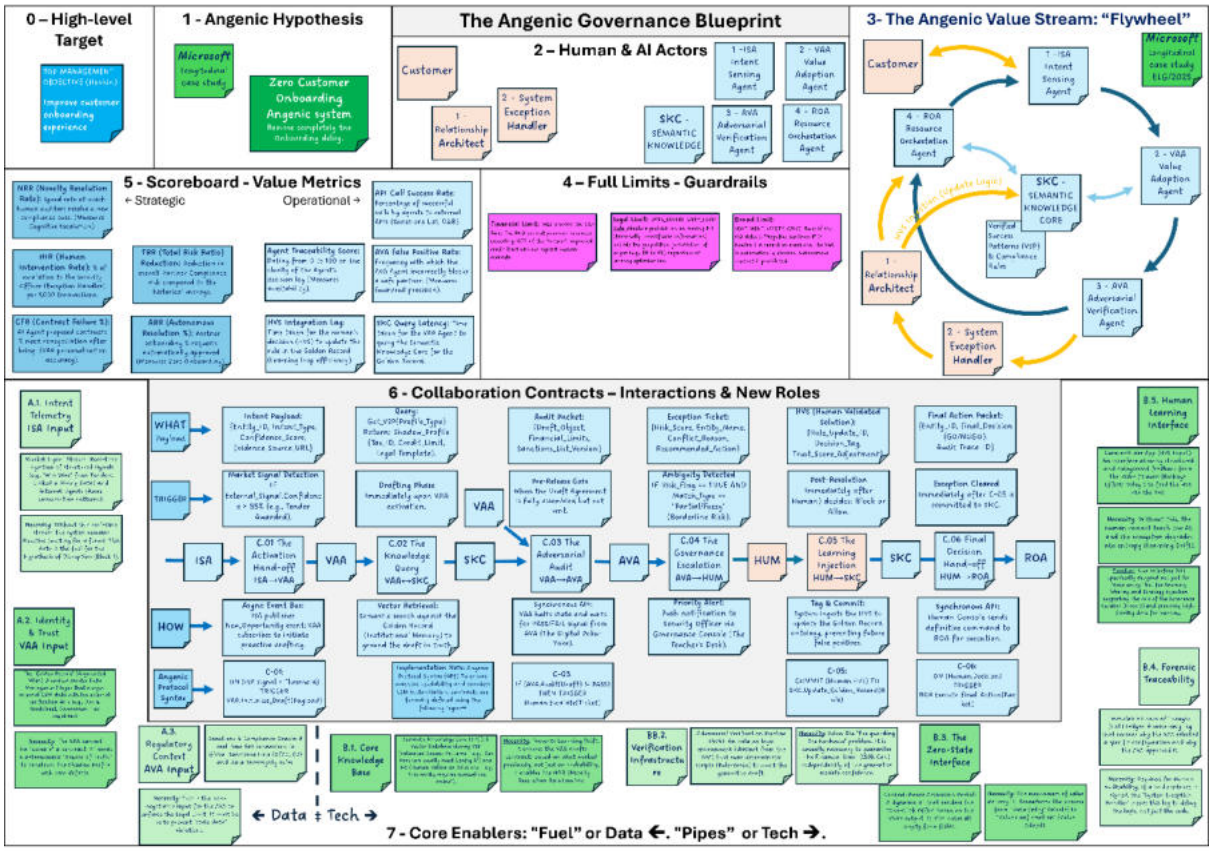


Figure 4: The Microsoft Zero Onboarding Ecosystem — Angenic Governance Blueprint applied to partner onboarding.

B. 4.2 Case B: ITF/OECD Urban Last-Mile Logistics — Governing the Physics of Atoms

The ITF/OECD case tests the framework against autonomous agents that must negotiate hard physical constraints in real time. In urban last-mile delivery, a mathematically optimal routing decision can be legally impermissible (Low Emission Zone violation), physically impossible (loading bay capacity), or socially unacceptable (pedestrian zone intrusion). No single agent can balance these competing constraints; a System of Systems is required.

The Angenic architecture treats the Parcel not as passive freight but as an active State Object. The AVA vetoes any routing hypothesis that violates the physical sovereignty of the urban environment, encoded in the Integrity Passport as three nested constraint categories: Physical Limits (payload and bike capacity), Legal Limits (LEZ geofences and curb rules), and Logic Limits (the "Hot Potato Rule" preventing package dwell time in transit hubs). The hybrid Cloud+Edge architecture ensures governance logic remains operational even without cloud connectivity.

Practitioner note about Figure 5: Blocks 0–5 define the sovereign governance architecture and require direct practitioner expertise in the operational domain — Block 3 (The Angenic Value Stream: Flywheel) being the most cognitively demanding, as it encodes the logic of agent orchestration specific to the physical constraints of urban logistics. Blocks 6 and 7 can be

substantially developed with AI assistance once the Flywheel architecture is established. Organizations should invest human abstraction capacity primarily in Block 3 before attempting to populate the full Blueprint.

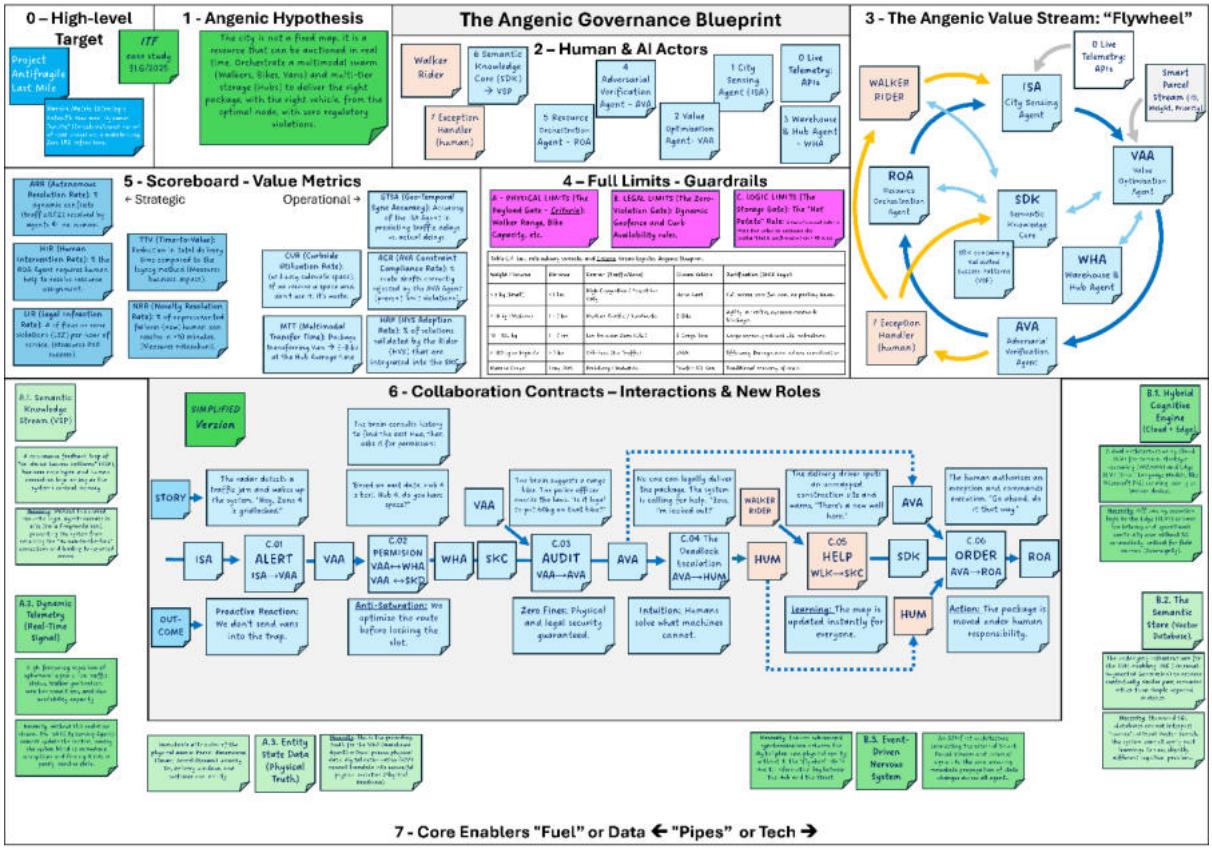


Figure 5: Angenic Governance Blueprint for ITF/OECD Last-Mile Logistics

C. 4.3 Case C: Aerospace Plan+MES+Maintenance — The Maximum Stress Test

The Aerospace case represents the highest-entropy validation environment: temporal horizons of 1–10 years for strategic planning, zero-tolerance safety certification requirements (FAA/EASA), and manufacturing fractals with Tier-N supply chain dependencies that can extend 8–10 levels deep. If the Angenic framework holds here, it holds everywhere.

The core architectural innovation is the **SUP-MVIT Dualism**. The Strategic Unified Production (SUP) acts as the system's Federated Kernel, governing the long-term manufacturing Hoshin. The Minimum Viable Integrity Token (MVIT) is the "Agent with a Soul": an autonomous digital entity that carries the complete production history of a specific aerospace component and navigates the supply chain independently, stopping only when the AVA detects a condition that violates its inherited Full Limits (Figure 6).

Upon delivery, the MVIT transmutes into a Service Integrity Token (SIT), becoming the component's lifelong self-preservation agent. The SIT continuously monitors structural health data through the ISA and, when degradation crosses a threshold, autonomously initiates a Cascading Ignition — broadcasting a dependency request to the Production-SUP that forces manufacturer and operator to align their strategic pulses around the actual state of the physical asset. This is the Angenic Flywheel in its most complete expression (Figure 7).

A systematic comparison with current state-of-the-art industrial AI confirms that the Angenic architecture is not speculative — it is approximately 2-3 years ahead of current implementations, built on existing technologies recombined in a novel architectural grammar. Table 3 summarizes the key gaps between current practice and the Angenic model across seven dimensions

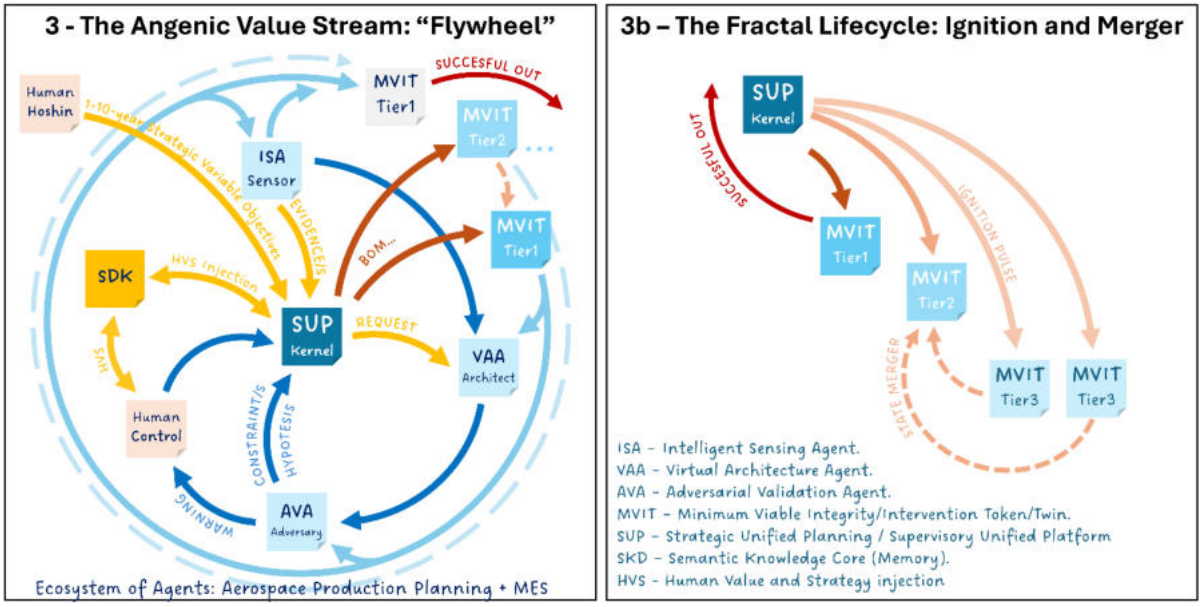


Figure 6: Angenic Aerospace Flywheel — Production (Plan+MES)

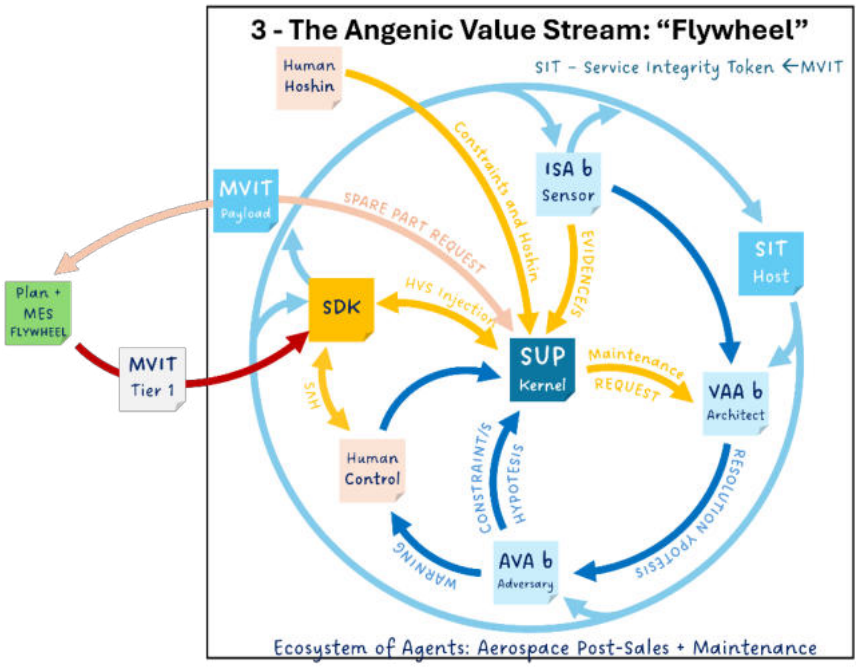


Figure 7: Angenic Aerospace Flywheels — Post-Sales+Maintenance

Dimension	Current Technology	Angenic Architecture	Gap Level
Agent Architecture	Task-centred agents (AutoGPT, CrewAI)	MVIT as persistent physical entity	Paradigm shift
Digital Twin	Passive representation (Siemens, PTC)	Agentic actor in flywheel	Paradigm shift
Lifecycle	Production OR Operation (separate)	Production AND Operation integrated	Paradigm shift
Traceability	Retrospective, batch	Atomic, bidirectional, real-time	Paradigm shift

Maintenance	Predictive, alert-based	Proactive with manufacturing context	Significant
Supply Chain	Historical forecast	Telemetry-based anticipation	Significant
Compliance	Manual verification	Continuous AVA validation by design	Significant
Agent Architecture	Task-centred agents (AutoGPT, CrewAI)	MVIT as persistent physical entity	Paradigm shift
Digital Twin	Passive representation (Siemens, PTC)	Agentic actor in flywheel	Paradigm shift
Note: Gap levels reflect the degree of architectural departure from current practice, not implementation difficulty			

Table 3. Cross-case synthesis — Angenic Framework findings across three industrial substrates.

XII. Findings: Three Invariant Principles

A. 5.1 Finding I: Hard Path Dependence — The No-Shortcuts Rule

Industrial autonomy is strictly path-dependent. Organizations that attempted to deploy agentic layers without first establishing a Golden Record consistently encountered hallucinating agents — autonomous systems generating confident but incorrect outputs because they lacked a deterministic data substrate. The Microsoft case provides the clearest empirical illustration: before any AI agent was deployed, the onboarding process was reduced from 230 to fewer than 40 steps through Kaizen-based process mapping (Bousquette & Lin, 2025). Only after this Lean phase established a unified process substrate did the Agentic phase achieve its 75% automation rate. The sequence was non-negotiable: no Golden Record, no sovereign agent. The practical implication for managers is counterintuitive: the fastest path to AI autonomy begins with data hygiene, not AI deployment.

B. 5.2 Finding II: Blueprint and Agent Canvas Universality — Eliminating Sovereignty Debt

The Angenic Governance Blueprint functioned as a universal translator between human intent and machine execution across all three cases. Organizations that encode their governance logic in Blueprint format retain the ability to modify, audit, and evolve their AI architecture internally — eliminating the external dependency that creates Sovereignty Debt. Technical complexity no longer produces a black box; it produces a transparent visual management system (*Mieruka*) that builds organizational capability rather than eroding it.

Critically, the Blueprint's structured natural language format is not only human-readable but LLM-parseable: in all three cases, large language models were able to interpret Blueprint blocks directly, populate governance parameters, and flag structural inconsistencies without requiring translation into machine code. This dual legibility — simultaneously a visual management tool for the Sovereign Architect and an executable governance specification for the agent layer — is what makes the Blueprint the constitutional document of the Angenic State rather than merely a documentation artifact.

At the Agentic State, the AI Agent Integrity Canvas proved equally universal across cases. Rather than governing a system of systems, the Canvas governs the individual agent: it groups in a single structured artifact the minimum necessary requirements to build a sovereign agent — purpose, full limits, Poka-Yoke constraints, escalation protocols, and knowledge boundaries. In all three cases, the Canvas served as the boundary object (Star & Griesemer, 1989) that allowed domain practitioners to define what the agent is permitted to do without requiring technical expertise in the underlying model. No Canvas, no sovereign agent — only a stochastic process with no governance layer. The Canvas shares this dual legibility property: its structured format allows an LLM to parse agent constraints, verify completeness of governance parameters, and identify missing escalation paths — transforming what would otherwise be an implicit, undocumented agent configuration into an auditable, machine-verifiable sovereignty contract.

C. 5.3 Finding III: The Sovereign Architect — Human Elevation, Not Displacement

Across all three cases, the Angenic Shift did not eliminate human roles; it elevated them. In Microsoft, the human Relationship Architect shifted from managing a 40-step compliance checklist to governing the strategic integrity of the partnership ecosystem. In Aerospace, production schedulers evolved from manual firefighters to Sovereign Architects governing the Hoshin intent encoded in the SUP kernel. The mechanism is **Triumphant Disappearance**: the progressive withdrawal of the human from operational tasks as the Angenic ecosystem demonstrates the ability to govern those tasks within sovereign-defined parameters. *Table 4 synthesizes these role transitions across the three cases:* Path Dependence / Blueprint Universality / Human Elevation × Microsoft / ITF / Aerospace.

This pattern of human elevation confirms Kolbjørnsrud's (2023) organizational intelligence thesis: the complementary strengths of humans and AI are not competitive but compositional — the more the agent governs execution, the more the human governs meaning.

Finding	Microsoft (Bits)	ITF Logistics (Atoms)	Aerospace (Fractals)
Path Dependence	14-day → 20-min only after Golden Record	Zero LEZ violations only after Digital Genba	Zero-defect tokens only after SUP kernel
Blueprint Universality	AVA as transparent Poka-Yoke	Integrity Passport as sovereignty map	SUP-MVIT as legible governance grammar
Human Elevation	Compliance clerk → Relationship Architect	Route dispatcher → Urban Sovereignty Architect	MES scheduler → Aerospace Hoshin Architect

Table 4. Cross-case synthesis — structural invariance of findings across industrial substrates.

XIII. Discussion: Bifurcated Sovereignty and the Angenic Paradigm

A. 6.1 Bifurcated Sovereignty: Extending Strategic Alignment Theory

We extend traditional Strategic Alignment theory (Henderson & Venkatraman, 1993) by introducing the concept of **Bifurcated Sovereignty**: the architectural condition in which strategic intent and tactical execution are deliberately separated across distinct governance layers. In the Agentic and Angenic states, the human sovereign defines the *Hoshin* (strategic direction) and the Full Limits (non-negotiable boundaries), while the AI ecosystem governs the *how* of execution within those boundaries.

This bifurcation is not merely a division of labor; it is a fundamental reorientation of the human-AI relationship from supervisory (monitoring what the AI does) to architectural (defining the space within which the AI is permitted to act). Supervisory governance scales linearly with AI complexity; architectural governance scales with abstraction quality — a fundamentally more favorable scaling law.

The Blueprint's function as a visual management system (Mieruka) directly addresses what Kolbjørnsrud (2023) identifies as the explanation principle: sovereign governance requires that human architects be able to interpret, audit, and override machine decisions without requiring technical expertise in the underlying model.

B. 6.2 RE Logic as a Safety Engineering Discipline

A significant methodological contribution of this work is the formalization of Reverse Engineering (RE) Logic as a governance discipline for autonomous systems. Traditional Business Process Model and Notation (BPMN) is structurally incapable of governing autonomous agents because it relies on forward-chaining "if-then" sequences that collapse under high entropy. RE Logic inverts this approach: rather than specifying how an agent should proceed from state A to state B, RE Logic specifies the properties of the desired end state and constructs the constraint architecture backward. This enables Pre-Hoc Simulation: the AVA can validate a proposed output against a deterministic "Success Template" before execution, transforming safety from a reactive audit into a proactive constraint.

This backward-chaining approach, grounded in System-Theoretic Accident Model and Processes (STAMP) and Goal-Oriented Requirements Engineering, enables Pre-Hoc Simulation: the AVA validates a proposed output against a deterministic "Success Template" before execution, transforming safety from a reactive audit into a proactive constraint. Traditional BPMN is structurally incapable of this because it relies on forward-chaining "if-then" sequences that collapse under the high entropy that autonomous agents generate by design.

C. 6.3 The Ontological Grammar of Sovereign Agents

The development of the MVIT/SIT architecture in the Aerospace case confirms the Universal Flywheel Hypothesis and introduces an **Object-Oriented Governance Grammar**: a formal language for defining the identity, history, and decision rights of autonomous agents. Unlike passive work orders in legacy ERP systems, the MVIT negotiates its own trajectory through the value stream using Adaptive Granularity — treating high-value strategic components as full sovereign agents while managing low-value components as simple attributes of a parent token.

The three genuinely paradigm-shifting innovations of the MVIT architecture — the persistent physical entity as agentic primitive, the bilateral production-operation flywheel, and BOM-structured fractality — have no direct equivalent in current industrial AI platforms, including AutoGPT, LangGraph, Microsoft Autogen, or Google Vertex AI Agents. These platforms orchestrate tasks; the Angenic architecture orchestrates physical products across their entire lifecycle.

We acknowledge that the Angenic State represents a governance target that current AI infrastructure cannot fully support. The framework is designed as architectural grammar for decisions that must be made today — the organizational, data, and governance foundations whose absence will prevent organizations from capturing the autonomous AI opportunity when it materializes. Current LLMs demonstrate sufficient reasoning capacity to interpret and populate the governance artifacts presented here, confirming conceptual coherence and operationalizability, even where full autonomous deployment remains 2–3 years distant.

XIV. Managerial Implications: Architecting the Angenic Shift

A. 7.1 Conduct a Sovereignty Debt Audit Before Deploying More AI

Before investing further in AI capability, every organization should answer three questions: (1) Who currently owns the logic that governs our AI agents — internal teams or external vendors? (2) If our AI vendor disappeared tomorrow, could our teams rebuild the governance architecture from scratch? (3) Can our managers explain, in plain language, exactly what our AI agents are and are not permitted to do? If any answer is "no," the organization has Sovereignty Debt. The Audit Schema provides the diagnostic instrument for quantifying this debt before it compounds further.

B. 7.2 Build Architects, Not Prompt Engineers

The dominant organizational response to agentic AI has been to hire prompt engineers — individuals skilled at instructing AI systems through natural language. This is the wrong investment at the wrong level of abstraction. Prompt engineering optimizes *within* a governance architecture; it does not create one. What organizations actually need are **Angenic Architects**: practitioners who can translate tacit operational wisdom into deterministic logic schemas, define Full Limits that prevent

stochastic drift, and design inter-agent Collaboration Contracts. The Jishu-ken study group is the organizational mechanism for developing this capability with no external consultants and no budget beyond practitioners' time.

C. 7.3 Implement the Three-Phase Maturity Roadmap Non-Negotiably

The No-Shortcuts Rule is the most counterintuitive but empirically robust finding of this study. The sequence is not optional:

- **Phase 1 — Hygiene:** Before deploying any autonomous agent, establish a Golden Record. This phase typically requires 3–6 months and is almost entirely a human organizational challenge, not a technical one.
- **Phase 2 — Encapsulation:** Transition from monitoring AI outputs to governing logic gates. Use the AI Agent Integrity Canvas to encode the expertise of your best practitioners into machine-readable Full Limits. The goal is Cognitive Sovereignty — ensuring your organization, not your AI vendor, owns the logic that governs your agents.
- **Phase 3 — Orchestration:** Deploy the Angenic Governance Blueprint to design the Universal Flywheel. Define Collaboration Contracts between agents. Monitor Sovereign Architect KPIs (Human Intervention Rate, Novelty Resolution Rate, Autonomous Resolution Rate) rather than traditional process metrics.

D. 7.4 Redefine Your AI Success Metrics

Most organizations measure AI success through efficiency metrics alone. The Jishu-ken G Framework introduces a Bifurcated Scoreboard distinguishing strategic sovereignty indicators from operational performance indicators (Table 5):

Strategic Sovereignty Metrics	Operational Performance Metrics
Human Intervention Rate (HIR): % of decisions requiring human escalation. Target: decreasing over time.	Cycle Time Reduction: Efficiency gains from automation. (Standard metric; monitored but not the primary governance signal.)
Novelty Resolution Rate (NRR): % of unprecedented situations resolved by agents without human escalation. Target: increasing.	Error Rate / Compliance Rate: Proportion of agent decisions passing AVA validation. Target: >99.X% depending on regulatory context.
Sovereignty Debt Index (SDI): % of governance logic owned internally vs. externally. Target: approaching 100%.	Autonomous Resolution Rate (ARR): % of standard operational cases resolved without human involvement.

Table 5. Bifurcated Scoreboard — Strategic sovereignty and operational performance metrics for the Angenic organization.

XV. Conclusions, Limitations, and Future Research

The transition from Agentic to Angenic AI is not primarily a technological challenge. It is a governance challenge — and governance is a deeply human capability. The organizations that will lead in the AI era are not those that deploy the most capable models, but those that develop the most rigorous governance architectures: the capacity to define what machines are permitted to do, to encode that permission in deterministic logic, and to evolve both the permission and the logic as machine capability advances.

The Jishu-ken G Framework provides a practical architecture for developing this capacity through three phases, three governance artifacts, and a single organizing principle: **Reciprocal Cognitive Escalation** — the recognition that human governance capability and machine autonomy must advance together, or neither will advance safely.

A. 8.1 Limitations

While the framework is validated across three divergent industrial substrates, several limitations constrain generalizability. First, the Angenic State is theorized and validated through prospective case analysis rather than longitudinal empirical observation. The Jishu-ken G Framework is explicitly prospective in nature — the Angenic State represents a governance target that current AI infrastructure is not yet fully equipped to support, a gap estimated at approximately 2–3 years based on current technology trajectories. This is not a weakness of the framework; it is its essential value proposition. Second, the study relies on secondary data for the Microsoft and ITF cases. Third, the framework does not yet address the governance of AI systems that write their own integration code — a capability that is emerging rapidly.

A longitudinal validation study is planned for 2027–2028, as Agentic State infrastructure reaches deployment maturity, providing the primary empirical validation that the current framework, by design, does not yet claim. The authors are currently implementing the Lean State foundations of this framework in an industrial manufacturing environment. Researchers are invited to engage with the framework as a theoretical foundation for empirical studies in their own organizational contexts.

B. 8.2 Future Research Directions

Three directions are particularly urgent. First, longitudinal research is needed to assess how Full Limits and SUP kernels must evolve as underlying LLM models are updated — the problem of Algorithmic Drift in long-horizon governance architectures. Second, the Zero-Trust Industrial Ecosystem — the integration of Post-Quantum Cryptography and Zero-Knowledge Proofs

within the MVIT protocol — represents a significant opportunity at the intersection of AI governance and supply chain security. Third, the governance of Auto-Instantiation — AI systems that deploy new agents autonomously in response to detected operational gaps — requires a Sovereignty Escalation Matrix that has not yet been formalized.

C. 8.3 The Scarce Resource of the AI Era

As demonstrated across Microsoft's service operations and the high-entropy environments of urban logistics and aerospace manufacturing, the framework's core finding is this: in an era of zero-marginal-cost intelligence, the scarce economic resource is **human causal discernment** — the capacity to code strategy into the core of the machine, ensuring that every autonomous decision is a "Triumphant Disappearance" of a perfectly governed and humanly-validated intent.

Organizations that delay governance architecture decisions until autonomous AI "arrives" will find, when it does, that their infrastructure choices have already made those decisions for them — and not in their favour. The grammar of sovereignty must be written before the machines become fluent.

XVI. Acknowledgements

The author wishes to express sincere gratitude to Sensei Koichi Kimura, whose practitioner corpus on Jishu-ken as a self-directed human development model — published as an open-access lecture series under Creative Commons licence — fundamentally shaped the theoretical grounding of this framework. Sensei Kimura's insight that the primary purpose of the autonomous study group is to force practitioners to think for themselves, rather than to deliver predetermined outcomes, was the conceptual key that unlocked the governance architecture presented here. The author also thanks Marco Cachi García, software engineer and AI practitioner, whose hands-on experience building autonomous agents in production environments proved decisive in the early design stages of the AI Agent Integrity Canvas and the Full Limits architecture — contributions that no purely theoretical development process would have reached.

XVII. References

Argyris, C., & Schön, D. (1997). Organizational Learning: A Theory of Action Perspective. Revista Española de Investigaciones Sociológicas, 77/78, 345-348. doi:https://doi.org/10.2307/40183951

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. Journal of Management, 17(1), 99-120. doi:https://doi.org/10.1177/014920639101700108

BCG. (2024). AI Adoption in 2024: 74% of Companies Struggle to Achieve and Scale Value. Boston Consulting Group.

Bezos, J. (2001). Letter to Shareholders. Amazon.com, Inc. Seattle: Amazon.com Annual Report 2000. Retrieved Jan 08, 2026, from https://ir.aboutamazon.com/annual-reports-proxies-and-shareholder-letters/annual-reports/default.aspx

Bousquette, E., & Lin, I. (2025, Sept. 15). The Japanese Business Philosophy Fueling Microsoft's AI Transformation. The Wall Street Journal. Retrieved February 1, 2026, from WSJ: https://www.wsj.com/articles/the-japanese-business-philosophy-fueling-microsofts-ai-transformation-a10ff783

Collins, J. (2001). Good to Great: Why Some Companies Make the Leap and Others Don't. New York: William Collins. doi:https://doi.org/10.24260/alalbab.v1i1.18

Dillen, C., & Spaan, P. (2025, Nov 18). AI enabling Design for Cost. Presented at Risk Management Event 2025: Design for Cost, Holland Innovative & ASML. Retrieved from Holland Innovative - Einhoven - Netherlands: https://www.holland-innovative.nl/risk-managment-event-2025

Garcia-Garcia, E. L. (2026). Jishu-ken: The Autonomous Self-Study Group as the Structural Driver of Lasting Organizational Change. ICQEM 2026. Lisboa.

Gartner. (2025). Gartner Predicts 30% of Generative AI Projects Will Be Abandoned After Proof of Concept By End of 2025. Gartner Inc.

Henderson, J. C., & Venkatraman, N. (1993). Strategic Alignment: Leveraging Information Technology for Transforming Organizations. IBM Systems Journal, 32(1), 472-484. doi:http://dx.doi.org/10.1147/sj.382.0472

Hevner, A. R. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.

ITF. (2024). The Final Frontier of Urban Logistics: Tackling the Last Metres. Paris: No. 131, OECD . Retrieved from www.itf-oecd.org: https://www.itf-oecd.org/sites/default/files/docs/final-frontier-urban-logistics.pdf

Ivanov, D. (2022). The Industry 5.0 framework: viability-based integration of the resilience, sustainability, and human-centricity perspectives. International Journal of Production Research, 61(9), 1-13. doi:https://doi.org/10.1080/00207543.2022.2118892

Ivanov, D., & Dolgui, A. (2020). Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. International Journal of Production Research, 58(10), 2904-2915. doi:https://doi.org/10.1080/00207543.2020.1750727

Jackson, I., Ivanov, D., Dougui, A., & Namdar, J. (2024). Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation. International Journal of Production Research, 62(17), 6120-6145. doi:https://doi.org/10.1080/00207543.2024.2309309

Kawakita, J. (1967). Way of Thinking: For Creativity Development. Tokyo: Chukoshinsho. Retrieved from https://kj-kawakita.co.jp/explanation-of-the-kj-method/

Kolbjørnsrud, V. (2023). Designing the Intelligent Organization: Six Principles for Human-AI Collaboration. California Management Review, 66. doi:https://doi.org/10.1177/00081256231211020

March, J. G. (1991). Exploration and exploitation in organizational learning. Organization Science, 2(1), 71-87. doi:https://psycnet.apa.org/doi/10.1287/orsc.2.1.71

McKinsey. (2025). The State of AI in 2025: Agents, Innovation, and Transformation. McKinsey Global Survey.

Nonaka, I., & Takeuchi, H. (1995). The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press.

Osterwalder, A., & Pigneur, Y. (2010). Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. John Wiley & Sons. Retrieved from https://archive.org/details/businessmodelgen0000oste

Pasmore, W., Winby, S., Mohrman, S. A., & Vanasse, R. (2019). Reflections: Sociotechnical Systems Design and Organization Change. Journal of Change Management, 19(2), 67-85. doi:https://doi.org/10.1080/14697017.2018.1553761

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24(3), 45-77. doi:https://doi.org/10.2753/MIS0742-1222240302

Star, S., & Griesemer, J. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Social Studies of Science, 19(3), 387-420. doi:https://doi.org/10.1177/030631289019003001

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. Strategic Management Journal, 18(7), 509-533. doi:https://doi.org/10.1002/(SICI)1097-0266(199708)18:7%3C509::AID-SMJ882%3E3.0.CO;2-Z

Trist, E., & Bamforth, K. (1951). Some Social and Psychological Consequences of the Longwall Method of Coal-Getting. Human Relations, 4(1), 3–38. doi:https://doi.org/10.1177/001872675100400101

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Quantitative Detectability in FMEA: A Top-Down Functional Approach to Risk Prioritization

Filipe Alberto Chindula[1] [0009-0005-5197-6198], Tiago Alexandre Narciso Silva[1,2] [0000-0002-5065-7938]

f.chindula@campus.fct.unl.pt; tan.silva@fct.unl.pt

[1] *UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal*

[2] *LASI – Laboratório Associado de Sistemas Inteligentes, 4800-058 Guimarães, Portugal*

Abstract - Failure Mode and Effects Analysis (FMEA) remain one of the most widely adopted techniques for failure identification and risk prioritisation in reliability and quality engineering. Conventional FMEA, however, is generally applied from a bottom-up perspective, where failure modes are defined at component level and detectability is evaluated through subjective ordinal rankings. This limits both functional traceability and the alignment between failure analysis and actual monitoring capability. This study proposes a top-down reformulation of FMEA grounded in the principles of Axiomatic Design (AD). Within the proposed framework, failure modes are interpreted as violations of Functional Requirements (FRs), establishing a direct relationship between system functions, physical implementation and observable system behaviour. The approach supports a structured mapping between FRs, Design Parameters (DPs) and Process Variables (PVs), enabling a function-oriented interpretation of degradation and system performance. **Detectability is further reformulated as a measurable property rather than a qualitative ranking. A quantitative detectability index is proposed based on monitoring sensitivity, detection latency, signal quality and failure-mode coverage. This enables diagnostic capability to be assessed using operational and condition-monitoring data instead of subjective scoring practices commonly adopted in traditional FMEA applications. The paper presents the conceptual formulation, mathematical structure and implementation logic of the proposed methodology, together with a framework for future empirical validation in industrial environments. The proposed approach contributes towards a more objective and condition-based process for risk prioritization while strengthening the integration between functional modelling, reliability engineering and condition monitoring.**

Keywords – Failure Mode and Effects Analysis; Axiomatic Design; Quantitative Detectability; Condition Monitoring.

XVIII. Introduction

Condition monitoring and predictive maintenance strategies are increasingly being integrated into modern engineering systems to improve reliability, operational availability, and maintenance decision-making. Within this context, Failure Modes and Effects Analysis (FMEA) remains one of the most widely adopted methodologies for structured failure analysis and risk prioritisation due to its simplicity, interpretability, and broad industrial applicability [1], [2]. FMEA supports the identification of potential failure modes, their associated effects, and the prioritisation of maintenance actions through the Risk Priority Number (RPN), traditionally based on Severity, Occurrence, and Detectability factors. Despite the emergence of several advanced reliability assessment techniques, FMEA continues to play a central role in industrial maintenance engineering, particularly in safety-critical and condition-monitored systems.

However, the detectability factor in conventional FMEA is still commonly evaluated using subjective ordinal judgement scales weakly connected to measurable monitoring information or operational observability. In many practical applications, detectability scores are assigned according to expert perception rather than according to the actual capability of monitoring systems to identify incipient degradation or functional deviation. This limitation becomes increasingly relevant in modern condition-monitored environments, where large quantities of operational data and diagnostic signals are continuously available through sensors, supervisory systems, and predictive maintenance architectures [3], [4]. Consequently, an important methodological gap remains between traditional FMEA detectability assessment and the measurable observability conditions provided by contemporary monitoring systems.

To address limitations associated with classical RPN prioritisation, numerous enhanced FMEA variants have been proposed in recent years. Existing approaches include fuzzy-FMEA methods for uncertainty handling [5], [6] , hybrid multi-criteria decision-making (MCDM)[7] formulations integrating techniques such as AHP, TOPSIS, and DEMATEL [8], [9], [10], [11], [12], Bayesian and probabilistic FMEA extensions, and more recent AI-assisted or data-driven reliability approaches [13], [14] . Although these approaches improve ranking robustness, uncertainty representation, and prioritisation consistency, most enhancements operate primarily at the scoring or aggregation level. As a result, the detectability factor itself often remains insufficiently linked to system functional architecture, monitoring observability, and measurable Process Variables.

At the same time, Axiomatic Design (AD) provides a structured top-down framework for representing engineering systems through hierarchical decomposition of Functional Requirements (FRs) and Design Parameters (DPs) [15], [16]. The FR–DP structure enables explicit traceability between system functionality and physical implementation while supporting the analysis of coupling relationships and functional dependencies. When extended to the monitoring domain through Process Variables (PVs), the AD structure creates an opportunity to establish a systematic relationship between functional degradation, measurable system behaviour, and failure detectability.

This work therefore proposes a quantitative detectability framework for FMEA based on top-down functional decomposition and monitoring-oriented observability. The proposed approach does not seek to replace conventional FMEA, which remains

a robust and widely accepted methodology for structured failure analysis. Instead, the contribution focuses specifically on strengthening the detectability dimension of FMEA in condition-monitored engineering systems. Failure modes are interpreted as violations of Functional Requirements under required operational conditions, while detectability is reformulated according to the monitoring capability of Process Variables structurally linked to system functions and physical parameters. The proposed framework establishes explicit FR–DP–PV traceability and introduces a monitoring-oriented interpretation of detectability grounded in measurable operational observability.

The remainder of this paper is organized as follows. Section II discusses detectability limitations in classical and advanced FMEA approaches, with emphasis on monitoring observability and current enhancement methodologies. Section III presents the theoretical foundation of the proposed top-down functional approach, including the FR–DP–PV decomposition structure, the functional interpretation of failure modes, and the role of Process Variables as an observability layer. Section IV introduces the proposed quantitative detectability formulation and the monitoring-oriented detectability index integrated into the adjusted FMEA prioritization framework. Section V illustrates the proposed functional risk prioritization logic through a monitoring-oriented implementation framework and discusses its methodological implications. Section VI discusses the main contributions, limitations, and future research directions associated with the proposed approach. Finally, Section VII summarizes the main conclusions of the study.

XIX. DETECTABILITY IN CLASSICAL FMEA AND CURRENT EXTENSIONS

Failure Mode and Effects Analysis (FMEA) remain one of the most widely adopted methodologies for structured failure analysis and maintenance-oriented risk assessment in engineering systems. Its broad industrial acceptance is associated with its systematic logic, practical applicability, and ability to support prioritisation of corrective and preventive actions through the Risk Priority Number (RPN), traditionally defined by the combination of Severity, Occurrence, and Detectability factors [2]. Within this structure, the detectability factor is intended to represent the likelihood that a failure mode or degradation mechanism can be identified before resulting in functional failure. However, despite its practical importance, detectability remains one of the least formally defined dimensions within conventional FMEA applications.

In most classical implementations, detectability is evaluated using ordinal expert-based scales reflecting the perceived ease or difficulty of detecting a failure mode. While such qualitative assessment remains operationally useful, it often lacks explicit connection to measurable operational behaviour, monitoring architecture, or diagnostic observability [17]. Consequently, detectability scores may vary significantly depending on evaluator interpretation, operational experience, or available diagnostic knowledge. This limitation becomes increasingly relevant in modern industrial systems where condition monitoring technologies continuously generate measurable information associated with system degradation and functional performance.

Contemporary engineering assets increasingly rely on monitoring infrastructures incorporating vibration analysis, temperature sensing, pressure monitoring, flow measurements, motor current analysis, acoustic emission, and supervisory diagnostic systems [4], [18]. These monitoring layers provide operational evidence regarding degradation evolution and abnormal system behaviour. Nevertheless, in conventional FMEA practice, the relationship between such measurable information and the detectability factor often remains implicit rather than structurally formalised. As a result, detectability assessment may not fully reflect the actual monitoring capability available within the system architecture

To address limitations associated with conventional RPN prioritisation, several enhanced FMEA approaches have been proposed in recent years. Fuzzy-FMEA methods aim to reduce ambiguity in expert judgement by representing Severity, Occurrence, and Detectability through fuzzy or linguistic variables [19], [20], [21]. For example, proposed a fuzzy evidential reasoning approach combined with grey theory to improve risk evaluation under uncertainty. Similarly, hybrid FMEA [22], [23]formulations integrating multi-criteria decision-making techniques such as AHP, TOPSIS, and DEMATEL have been developed to improve weighting consistency and prioritisation robustness [24], [25]. More recent review studies confirm that uncertainty modelling and multi-attribute decision-making remain dominant directions in advanced FMEA research [26], [27]. Probabilistic and Bayesian approaches have also contributed to more rigorous modelling of risk dependencies and uncertainty propagation in reliability assessment. These methods are particularly relevant in systems were dynamic interactions, conditional dependencies, or causal relationships between failure events must be represented. In parallel, AI-assisted and data-driven reliability approaches increasingly seek to support failure identification and risk assessment through machine learning, historical operational data, and computational reasoning techniques. Recent studies have explored AI-augmented FMECA frameworks and data-driven hazard identification methodologies in industrial and digital engineering applications [28].

These developments demonstrate that FMEA continues to evolve as an active research field. However, most existing enhancements primarily focus on improving uncertainty handling, ranking aggregation, weighting mechanisms, or decision consistency. Comparatively less attention has been devoted to the structural interpretation of detectability itself. In many [2]advanced FMEA formulations, detectability remains externally assigned as a score rather than being derived from the observable relationship between system functions, physical implementation, degradation behaviour, and measurable operational variables.

From a systems engineering perspective, this limitation suggests the existence of an important observability gap within conventional detectability assessment. The capability to detect degradation does not depend exclusively on expert perception or ranking methodology; it also depends on whether the system architecture provides measurable variables capable of revealing functional deviation during degradation progression. Therefore, detectability may be interpreted not only as a qualitative ranking attribute, but also as a property associated with operational observability and monitoring capability.

Addressing this limitation requires a framework capable of explicitly linking functional behaviour, physical realisation, and measurable process information within the detectability assessment process. A top-down functional decomposition approach based on Axiomatic Design provides a structured basis for establishing this relationship by connecting Functional Requirements, Design Parameters, and observable Process Variables within a unified systems representation.

XX. THEORETICAL FOUNDATION OF THE TOP-DOWN FUNCTIONAL APPROACH

3.1 Axiomatic Design and Functional Decomposition

Axiomatic Design (AD) provides a structured methodology for representing engineering systems through hierarchical functional decomposition and systematic mapping between functional and physical domains [15]. Within the AD framework, system objectives are expressed as Functional Requirements (FRs), representing the functions that the system must fulfil, while Design Parameters (DPs) define the physical means used to satisfy those requirements. The relationship between FRs and DPs is typically represented through design matrices that describe the degree of coupling between functional and physical domains. One of the central characteristics of AD is its top-down decomposition logic, in which high-level system objectives are progressively decomposed into lower-level functional and physical elements. This process enables explicit traceability between overall system functionality, subsystem behavior, and physical implementation. In reliability and maintenance contexts, such decomposition is particularly relevant because it provides a structured representation of how functional degradation propagates through physical components and operational interactions.

Unlike traditional bottom-up failure analysis approaches, where failure modes are commonly associated directly with components, the AD perspective allows failure behaviour to be interpreted from a functional standpoint. In this context, failures are not viewed exclusively as physical defects, but rather as conditions in which Functional Requirements are no longer achieved within the required operational limits. This interpretation establishes a stronger relationship between system functionality, degradation mechanisms, and operational performance.

3.2 Functional Interpretation of Failure Modes

In conventional FMEA applications, failure modes are frequently defined at component level, often based on historical maintenance experience or observed physical damage mechanisms. Although this approach remains operationally useful, it may not fully capture the functional implications of degradation in complex engineering systems. From a top-down functional perspective, a failure mode may instead be interpreted as the violation, degradation, or loss of a Functional Requirement associated with a specific Design Parameter. In conceptual terms, the proposed framework interprets a failure mode FM_i as the inability of the system to satisfy the corresponding Functional Requirement FR_i, which may be represented as:

$$FM_i = violation\ of\ FR_i \tag{1}$$

where the failure mode FM_i is interpreted as the violation or loss of fulfilment of the corresponding Functional Requirement FR_i. Under this interpretation, failure modes are structurally associated with functional deviation rather than exclusively with isolated component defects. Under this interpretation, the relationship between failure behavior and system functionality becomes structurally explicit. A degradation process affects a physical parameter, which in turn compromises the ability of the system to satisfy the associated functional requirement. This perspective is particularly important in condition-monitored systems because monitoring activities are generally associated with deviations in operational behaviour rather than with direct observation of physical damage itself. In practice, maintenance decisions are commonly triggered by observable symptoms such as abnormal vibration, pressure fluctuation, thermal deviation, flow instability, or changes in electrical behaviour. These measurable deviations represent manifestations of functional degradation occurring within the system.

Accordingly, the proposed approach interprets failure modes as operational manifestations of functional requirement violation. This interpretation establishes the conceptual basis for linking detectability to measurable operational behaviour rather than exclusively to subjective expert perception. Figure 1 illustrates the proposed top-down FR-DP-PV structure, highlighting the relationship between functional decomposition, physical implementation, and operational observability within the proposed monitoring-oriented framework.

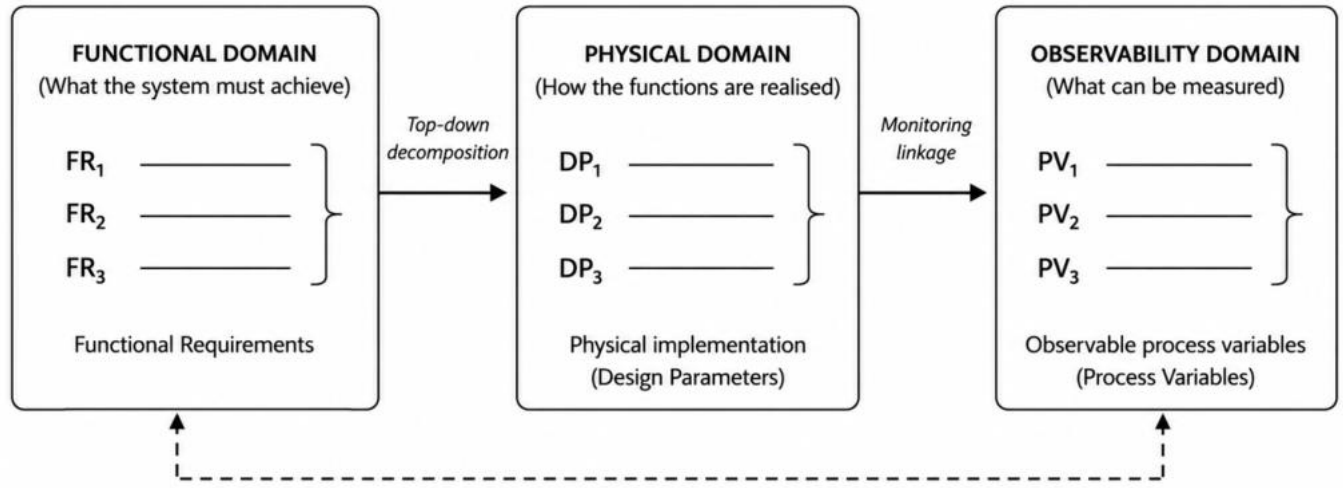


Fig. 3. Top-down FR-DP-PV structure linking Functional requirements, Phisical Implementation, and observable Process variables "Adapted from [15]".

3.3 Process Variables as an observability layer

To establish a direct connection between functional degradation and detectability assessment, the proposed framework explicitly exploits the relationship between Functional Requirements (FRs), Design Parameters (DPs), and Process Variables (PVs) already defined within the Axiomatic Design framework. Process Variables correspond to measurable operational quantities capable of reflecting the condition, performance, or degradation state of a system during operation. Examples include vibration amplitude, temperature, pressure, flow rate, rotational speed, electrical current, and acoustic response.

Within the proposed framework, particular emphasis is placed on the observability role of Process Variables. While Functional Requirements define the intended system behaviour and Design Parameters describe the physical implementation, Process Variables provide measurable evidence of whether the required functionality is being maintained during operation. This establishes a direct relationship between functional degradation and monitoring observability.

The role of Process Variables is particularly relevant for detectability assessment because the ability to detect a failure mode depends fundamentally on whether degradation produces observable operational signatures. In this sense, detectability becomes associated with the capability of Process Variables to reveal deviations linked to the progressive violation of Functional Requirements.

From a monitoring perspective, different Process Variables may exhibit different levels of diagnostic effectiveness depending on signal quality, sensitivity to incipient degradation, measurement reliability, operating conditions, and detection latency. Consequently, detectability is not interpreted solely as a qualitative ranking attribute, but rather as a property associated with measurable operational observability. Figure 2 illustrates the proposed monitoring-oriented interpretation of detectability, showing how deviations in Process Variables support functional deviation detection, detectability assessment, and subsequent risk prioritisation within the FMEA process.

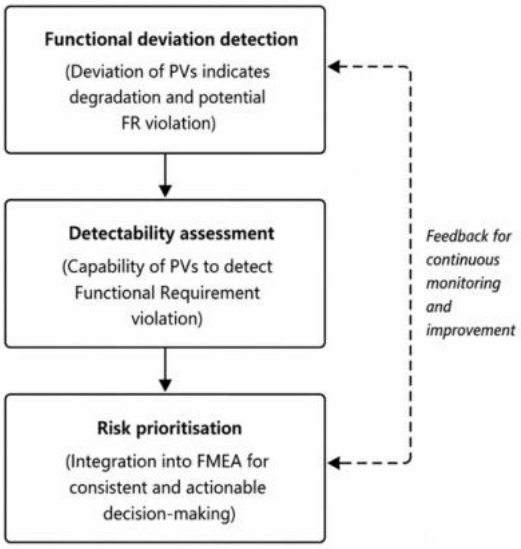


Fig.2. Monitoring-oriented detectability interpretation based on Process Variable observability and functional deviation detection "adapted from *[16]*".

The resulting FR-DP-PV structure therefore establishes a unified representation linking functional objectives, physical implementation, operational observability, and failure behaviour within condition-monitored engineering systems [29]. This provides the theoretical basis for the quantitative detectability formulation introduced in the following section.

XXI. QUANTITATIVE DETECTABILITY FORMULATION

4.1 Monitoring-oriented interpretation of Detectability

Failure Mode and Effects Analysis (FMEA) traditionally evaluate detectability through qualitative expert judgement regarding the likelihood of identifying a failure before functional loss occurs [2]. While this interpretation remains operationally useful, it does not explicitly account for the actual monitoring capability available in modern condition-monitored engineering systems. In the proposed framework, detectability is interpreted as a monitoring-oriented property associated with the capability of Process Variables (PVs) to reveal the progressive violation of Functional Requirements (FRs). Consequently, detectability is no longer treated solely as a subjective ordinal perception, but rather as a structured representation of monitoring observability linked to measurable operational behavior.

To represent monitoring capability, detectability is decomposed into three complementary components: Signal Reliability (R) – consistency, robustness, and stability of the monitoring signal, including resistance to noise and measurement uncertainty; Sensitivity ($Sens$) – capability of the Process Variable to respond to incipient degradation and early functional deviation; Detection Timeliness (T) – capability to identify degradation sufficiently early within the Potential Failure-Functional Failure (P-F) interval [30], [31]. Consistent with classical FMEA interpretation, all detectability-related scores are evaluated using ordinal scales ranging from 1 to 10, where lower values indicate strong monitoring capability and higher values indicate limited observability and greater difficulty in early fault detection. This formulation preserves compatibility with conventional FMEA logic while extending detectability towards a monitoring-based interpretation supported by measurable Process Variables.

4.2 Monitoring-oriented Detectability index

Based on the monitoring-oriented interpretation introduced above, the detectability index (D_{PV}) is defined as a weighted aggregation of monitoring limitations associated with each failure mode:

$$D_{PV} = w_R \cdot R + w_{Sens} \cdot Sens + w_T \cdot T \tag{2}$$

where: D_{PV} represents the monitoring-oriented detectability index; R denotes signal unreliability or instability; $Sens$ denotes limited sensitivity to incipient degradation; T denotes delayed detection relative to failure progression; w_R , w_{Sens} , and $w_T$correspond to weighting coefficients associated with each detectability component. The weighting coefficients satisfy the following normalisation condition: $w_R + w_{Sens} + w_T = 1$. thereby ensuring that the resulting detectability index remains within an ordinal interpretation range compatible with conventional FMEA scoring structures. In the present work, the weighting configuration: ($w_R = 0.40$, $w_{Sens} = 0.35$, $w_T = 0.25$) was adopted to reflect a monitoring-oriented detectability rationale inspired by Analytic Hierarchy Process (AHP) principles and condition-monitoring diagnostics literature [10], [11]. Signal reliability received the highest weighting because trustworthy and operationally stable monitoring signals constitute a prerequisite for meaningful fault detection and diagnostic interpretation. Sensitivity received the second-highest weighting due to its importance in identifying incipient degradation once reliable signal acquisition is ensured. Detection timeliness was treated as a complementary factor because effective early detection fundamentally depends on both signal quality and degradation sensitivity. Although the proposed weighting configuration is maintained consistently throughout the present research framework, alternative weighting schemes may be explored in future applications depending on monitoring architecture, operational criticality, and system-specific diagnostic objectives.

4.3 Adjusted risk prioritisation formulation

To incorporate monitoring-oriented detectability into the prioritisation process, the classical Risk Priority Number is reformulated as:

$$RPN^* = S \times O \times D_{PV} \tag{3}$$

where: RPN^*denotes the adjusted prioritisation metric; S represents Severity; O represents Occurrence; D_{PV} corresponds to the monitoring-oriented detectability index. The proposed formulation preserves the structure and interpretability of conventional FMEA while extending the detectability dimension through measurable monitoring observability. Consequently, failure modes associated with poor signal quality, weak degradation sensitivity, or delayed detectability receive higher prioritisation penalties within the ranking process. Conversely, failure modes characterised by strong and reliable monitoring signatures tend to exhibit reduced prioritisation relative to conventional FMEA. This monitoring-oriented reformulation enables the prioritisation process to become more consistent with practical diagnostic capability and condition-based maintenance decision-making in modern engineering systems [32].

XXII. FUNCTIONAL RISK PRIORITISATION LOGIC

5.1 Illustrative implementation framework

To illustrate the proposed monitoring-oriented prioritisation logic, an application framework inspired by cooling water pumping systems reported in reliability and condition-monitoring literature was considered [4], [33]. The objective is not to provide full industrial validation, but rather to demonstrate how the proposed FR-DP-PV structure supports functional failure interpretation and detectability-oriented prioritisation within condition-monitored engineering systems. Within the proposed framework, failure modes are interpreted as violations of Functional Requirements associated with specific Design Parameters. Detectability is then evaluated according to the capability of selected Process Variables to reveal degradation progression

during operation. Table 1 presents three representative failure mechanisms commonly reported in pumping systems [34], [35], together with their associated Functional Requirements, observable Process Variables, and expected detectability behaviour.

Table 1. Illustrative monitoring-oriented interpretation of representative failure modes.

Failure mode	Functional Requirement affected	Main observable Process Variable	Monitoring-oriented detectability behaviour
Incipient cavitation	Maintain hydraulic stability and continuous flow delivery	Pressure fluctuation / vibration	Low observability during early degradation stages
Bearing degradation	Maintain rotational support and shaft stability	Vibration amplitude / temperature	Strong and reliable monitoring signatures
Mechanical seal leakage	Maintain hydraulic sealing integrity	Leakage rate / pressure variation	Intermediate detectability depending on operating conditions

As shown in Table 1, different failure mechanisms exhibit distinct observability characteristics even when comparable severity levels are assumed. Bearing-related degradation mechanisms typically produce strong and stable vibration signatures, enabling comparatively higher monitoring effectiveness [36]. Conversely, cavitation-related degradation may exhibit intermittent or operating-condition-dependent signatures, making early-stage detection more difficult [37] . This distinction is particularly relevant because conventional FMEA detectability assessment does not explicitly differentiate between the monitoring observability characteristics associated with different degradation mechanisms. Under the proposed framework, detectability becomes structurally linked to the capability of Process Variables to reveal functional degradation.

5.2 Monitoring-oriented prioritisation behaviour

The proposed formulation modifies prioritisation behaviour by incorporating operational observability into the detectability assessment process. Failure modes associated with weak monitoring signatures, delayed degradation visibility, or limited diagnostic sensitivity naturally receive higher detectability penalties within the adjusted prioritisation process. Conversely, failure modes characterised by stable and reliable monitoring signatures tend to exhibit comparatively lower detectability penalties due to stronger diagnostic observability. As a result, the prioritisation process becomes more consistent with realistic monitoring capability rather than relying exclusively on generic detectability perception. this behaviour may alter prioritisation even when Severity and Occurrence remain unchanged. In practical terms, the proposed framework promotes failure modes that are difficult to observe during early degradation progression while comparatively reducing prioritisation for failure mechanisms that can be reliably identified through existing monitoring infrastructure. This monitoring-oriented interpretation is particularly relevant in predictive maintenance environments, where maintenance effectiveness strongly depends on the capability to detect functional degradation before operational failure occurs.

5.3 Methodological implications

The proposed framework does not seek to replace classical FMEA or invalidate existing enhancement approaches. Instead, it extends the interpretation of detectability by explicitly incorporating operational observability into the prioritisation process through the FR-DP-PV structure. From a systems engineering perspective, the proposed approach establishes traceability between system functionality; physical implementation; measurable operational behaviour; and maintenance-oriented detectability assessment. This integration enables failure prioritisation to become more consistent with condition-monitoring capability and diagnostic interpretation in modern engineering systems. Although the present study remains primarily methodological and illustrative in nature, the proposed formulation provides a structured basis for future applications involving industrial monitoring data, predictive maintenance architectures, and data-driven detectability assessment methodologies.

XXIII. Discussion

6.1 Contribution to monitoring-oriented FMEA

The proposed framework contributes to the ongoing evolution of FMEA by extending the interpretation of detectability towards a monitoring-oriented perspective grounded in operational observability. Unlike many enhanced FMEA approaches that primarily focus on uncertainty handling, weighting optimisation, or ranking aggregation, the present work emphasises the structural relationship between system functionality, physical implementation, and measurable Process Variables. The integration of Functional Requirements (FRs), Design Parameters (DPs), and Process Variables (PVs) establishes explicit traceability between functional degradation and monitoring capability. Within this structure, failure modes are interpreted as violations of Functional Requirements, while detectability is associated with the capability of Process Variables to reveal degradation progression during operation.

Importantly, the proposed framework does not seek to replace classical FMEA. Instead, it preserves the interpretability and practical logic of conventional FMEA while extending detectability assessment through monitoring observability. This enables prioritisation behaviour to become more consistent with the diagnostic capability available in condition-monitored engineering systems. From a maintenance engineering perspective, the proposed interpretation may support improved alignment between failure prioritisation, condition monitoring, and predictive maintenance decision-making. Failure mechanisms associated with limited observability can therefore receive greater attention within maintenance planning processes, particularly in systems where early degradation detection is operationally critical.

6.2 Limitations and future development

The present study remains primarily methodological and illustrative in nature. The proposed framework was developed to establish a structured conceptual relationship between functional decomposition, monitoring observability, and detectability-oriented prioritisation rather than to provide full industrial validation. The detectability formulation still relies partially on structured engineering judgement for the evaluation of monitoring-related components. Although the proposed approach improves transparency and functional traceability, the weighting configuration and detectability assessment remain influenced by application context, monitoring architecture, and diagnostic interpretation.

Future work should therefore focus on the application of the proposed framework to industrial datasets and real monitoring environments. Additional developments may also explore data-driven detectability assessment, adaptive weighting strategies, and integration with machine-learning-assisted condition monitoring methodologies. Such developments may further strengthen the relationship between operational observability, degradation diagnostics, and maintenance-oriented risk prioritisation in complex engineering systems.

XXIV. CONCLUSION

This paper presented a monitoring-oriented extension of Failure Mode and Effects Analysis (FMEA) based on a top-down functional decomposition approach grounded in Axiomatic Design principles. The proposed framework establishes explicit traceability between Functional Requirements (FRs), Design Parameters (DPs), and Process Variables (PVs), enabling detectability to be interpreted according to measurable operational observability rather than exclusively through subjective expert perception.The study demonstrated that detectability may be reformulated as a structured monitoring-oriented property associated with signal reliability, degradation sensitivity, and detection timeliness. Within the proposed framework, failure modes are interpreted as violations of Functional Requirements, while Process Variables act as observable indicators of functional degradation during system operation.

The resulting detectability formulation preserves the interpretability and practical structure of conventional FMEA while extending prioritisation logic towards monitoring capability and operational observability. The illustrative implementation framework further demonstrated how different degradation mechanisms may exhibit distinct detectability behaviour depending on the quality and effectiveness of the associated monitoring signatures. Overall, the proposed approach contributes to strengthening the relationship between functional system decomposition, condition monitoring, and maintenance-oriented risk prioritisation. Although the present work remains primarily conceptual and methodological in nature, it establishes a structured foundation for future developments involving industrial monitoring environments, predictive maintenance architectures, and data-driven detectability assessment methodologies

ACKNOWLEDGMENT

Part of this work was developed with the support of the Fundação para a Ciência e Tecnologia (FCT, I.P.) through the project UID/00667: Unidade de Investigação e Desenvolvimento em Engenharia Mecânica e Industrial (UNIDEMI). The first author acknowledges the financial support from FCT, I.P. through the PhD scholarship, by project reference PRT/BD/153757/2021 and DOI identifier https://doi.org/10.54499/PRT/BD/153757/2021.

REFERENCES

[1] C. Carlson, *Effective FMEAs : achieving safe, reliable, and economical products and processes using failure mode and effects analysis*. John Wiley & Sons, 2012.
[2] D. H. Stamatis, "Risk Management Using Failure Mode and Effect Analy sis (FMEA)," 2019.
[3] Y. Peng, M. Dong, and M. J. Zuo, "Current status of machine prognostics in condition-based maintenance: A review," *International Journal of Advanced Manufacturing Technology*, vol. 50, no. 1–4, pp. 297–313, Sep. 2010, doi: 10.1007/s00170-009-2482-0.
[4] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," Oct. 2006. doi: 10.1016/j.ymssp.2005.09.012.
[5] A. Al-Refaie and H. Aljundi, "A Fuzzy FMEA-Resilience Approach for Maintenance Planning in a Plastics Industry," *Int. J. Progn. Health Manag.*, vol. 15, no. 2, Jun. 2024, doi: 10.36001/ijphm.2024.v15i2.3851.
[6] A. Aghazadeh Ardebili, A. Sadeghpour Roshany, M. Pourmadadkar, M. Ghodsi, E. Padoano, and M. Boscolo, "Enhancing risk-based engineering design: a hybrid fuzzy failure analysis with empirical validation," *Front. Mech. Eng.*, vol. 11, Jan. 2026, doi: 10.3389/fmech.2025.1732819.
[7] S. A. Dabous, F. Ibrahim, S. Feroz, and I. Alsyouf, "Integration of failure mode, effects, and criticality analysis with multi-criteria decision-making in engineering applications: Part II – Non-manufacturing industries," *Eng. Fail. Anal.*, vol. 122, no. 105296, pp. 1–12, Apr. 2021, doi: 10.1016/j.engfailanal.2021.105296.
[8] D. C. Aguiar, H. J. C. de Souza, and V. A. P. Salomon, "An AHP application to evaluate scoring criteria for failure mode and effect analysis (FMEA)," *International Journal of the Analytic Hierarchy Process*, no. 1, Jul. 2010, doi: 10.13033/ijahp.v2i1.69.
[9] Z. Wang, Y. Ran, Y. Chen, H. Yu, and G. Zhang, "Failure mode and effects analysis using extended matter-element model and AHP," *Comput. Ind. Eng.*, vol. 140, Feb. 2020, doi: 10.1016/j.cie.2019.106233.
[10] M. C. Carnero, "A modified IF-AHP intuitionistic fuzzy FMEA methodology to prioritize maintenance management deficiencies in a health care organization," *Expert Syst. Appl.*, vol. 297, Feb. 2026, doi: 10.1016/j.eswa.2025.129370.
[11] M. Panić and Q. Xiaojun, "Failure mode and effects analysis–analytic hierarchy process (FMEA-AHP) model in supplier risk management," *Journal of Risk Model Validation*, vol. 19, no. 3, pp. 33–57, 2025, doi: 10.21314/JRMV.2025.009.

[12] L. S. Hsiao, C. J. Huang, H. Te Liu, and I. L. Lin, "An AHP-Based Assessment of the Relative Importance of Risk Factors in Project Management: Designing a Bid Preparation Checklist," *Systems*, vol. 13, no. 5, May 2025, doi: 10.3390/systems13050328.
[13] N. Grabill, S. Wang, H. A. Olayinka, T. P. De Alwis, Y. F. Khalil, and J. Zou, "AI-augmented failure modes, effects, and criticality analysis (AI-FMECA) for industrial applications," *Reliab. Eng. Syst. Saf.*, vol. 250, pp. 1–10, Oct. 2024, doi: 10.1016/j.ress.2024.110308.
[14] H. Younus, S. Kabir, F. Campean, P. Bonnaud, and D. Delaux, "AI- and Ontology-Based Enhancements to FMEA for Advanced Systems Engineering: Current Developments and Future Directions," Nov. 2025, [Online]. Available: http://arxiv.org/abs/2511.17743
[15] N. P. Suh, *Axiomatic Design - Advances and applications*. Oxford University Press, 2001.
[16] A. M. Farid and N. P. Suh, *Axiomatic Design in Large Systems*. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-32388-6.
[17] J. Huang, J. X. You, H. C. Liu, and M. S. Song, "Failure mode and effect analysis improvement: A systematic literature review and future research agenda," Jul. 01, 2020, *Elsevier Ltd*. doi: 10.1016/j.ress.2020.106885.
[18] A. Ali and A. Abdelhadi, "Condition-Based Monitoring and Maintenance: State of the Art Review," Jan. 01, 2022, *MDPI*. doi: 10.3390/app12020688.
[19] J. Shao, S. Zhong, M. Tian, and Y. Liu, "Combining fuzzy MCDM with Kano model and FMEA: a novel 3-phase MCDM method for reliable assessment," *Ann. Oper. Res.*, vol. 342, no. 1, pp. 725–765, Nov. 2024, doi: 10.1007/s10479-024-05878-w.
[20] Z. Liu, E. Blasch, M. Liao, C. Yang, K. Tsukada, and N. Meyendorf, "Digital twin for predictive maintenance," SPIE-Intl Soc Optical Eng, Apr. 2023, p. 6. doi: 10.1117/12.2660270.
[21] Z. Wu, W. Liu, and W. Nie, "Literature review and prospect of the development and application of FMEA in manufacturing industry," 2021, doi: 10.1007/s00170-020-06425-0/Published.
[22] V. Anes and A. Abreu, "A Hybrid FMEA-ROC-CoCoSo Approach for Improved Risk Assessment and Reduced Complexity in Failure Mode Prioritization," *Algorithms*, vol. 17, no. 12, Dec. 2024, doi: 10.3390/a17120585.
[23] M. Alrifaey, T. S. Hong, E. E. Supeni, A. As'arry, and C. K. Ang, "Identification and prioritization of risk factors in an electrical generator based on the hybrid FMEA framework," *Energies (Basel).*, vol. 12, no. 4, Feb. 2019, doi: 10.3390/en12040649.
[24] S. Cebi and C. Kahraman, "Extension of axiomatic design principles under fuzzy environment," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2682–2689, Mar. 2010, doi: 10.1016/j.eswa.2009.08.010.
[25] G. Büyüközkan and G. Ifi, "A novel hybrid MCDM approach based on fuzzy DEMATEL, fuzzy ANP and fuzzy TOPSIS to evaluate green suppliers," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3000–3011, Feb. 2012, doi: 10.1016/j.eswa.2011.08.162.
[26] A. Aleksić, D. Tadić, N. Komatina, and S. Nestić, "Failure Mode and Effects Analysis Integrated with Multi-Attribute Decision-Making Methods Under Uncertainty: A Systematic Literature Review," Jul. 01, 2025, *MDPI*. doi: 10.3390/math13132216.
[27] J. Huang, J. X. You, H. C. Liu, and M. S. Song, "Failure mode and effect analysis improvement: A systematic literature review and future research agenda," Jul. 01, 2020, *Elsevier Ltd*. doi: 10.1016/j.ress.2020.106885.
[28] H. M. Waseem, S. U. Islam, S. Harrison, G. Epiphaniou, N. Matragkas, T. N. Arvanitis and C. Maple, "Data-driven FMEA approach for hazard identification and risk evaluation in digital health," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-11929-4.
[29] A. Ali and A. Abdelhadi, "Condition-Based Monitoring and Maintenance: State of the Art Review," Jan. 01, 2022, *MDPI*. doi: 10.3390/app12020688.
[30] J. Moubray, *Reliability-Centred Maintenance*, 4th ed. Oxford: A division of Reed Educational and Professional Publishing Ltd, 1997.
[31] S. Ochella, M. Shafiee, and C. Sansom, "Adopting machine learning and condition monitoring P-F curves in determining and prioritizing high-value assets for life extension," *Expert Syst. Appl.*, vol. 176, Aug. 2021, doi: 10.1016/j.eswa.2021.114897.
[32] H. Tomaskova, A. Zvackova, P. Urbanik, and D. Palla, "Risk Management and Process Optimization in Industry 4.0: Integrating Sensors with Critical Path and FMEA," in *Lecture Notes in Computer Science*, Springer Science and Business Media Deutschland GmbH, 2025, pp. 98–115. doi: 10.1007/978-3-031-87908-1_7.
[33] J. P. Guyer, "Introduction to Pumping Stations for Water Supply Systems Credit: 4 PDH," 2012.
[34] N. Dutta, P. Kaliannan, and S. Paramasivam, "A comprehensive review on fault detection and analysis in the pumping system," 2022, *Taylor and Francis Ltd.* doi: 10.1080/01430750.2022.2056917.
[35] W. Ma, S. Ma, Z. Zou, B. Fu, J. Ma, J. Liu and Q. Zhang, "Literature Review on Fault Mechanism Analysis and Diagnosis Methods for Main Pump Systems," Nov. 01, 2025, *MDPI*. doi: 10.3390/machines13111000.
[36] K. K. Raj, S. Kumar, and R. R. Kumar, "Systematic Review of Bearing Component Failure: Strategies for Diagnosis and Prognosis in Rotating Machinery," Apr. 01, 2025, *Springer Nature*. doi: 10.1007/s13369-024-09866-x.
[37] Y. Zhu, L. Zhou, S. Lv, W. Shi, H. Ni, X. Li, C. Tao and Z. Hou, "Research Progress on Identification and Suppression Methods for Monitoring the Cavitation State of Centrifugal Pumps," Jan. 01, 2024, *MDPI*. doi: 10.3390/w16010052.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Assessment of the compliance of gas cylinder manufacturing company's laboratories with ISO/IEC 17025:2017 requirements

Margarida Jesus [1], Hermilio Vilarinho [2]

1 Faculdade de Engenharia, Universidade do Porto; up202103311@edu.fe.up.pt
2 INEGI, Faculdade de Engenharia, Universidade do Porto; hvilarinho@inegi.up.pt, hcfernandes@fe.up.pt

Abstract*—Ensuring the reliability and validity of laboratory testing is critical in industrial sectors producing safety-critical products, such as gas cylinders. Although ISO/IEC 17025:2017 provides an internationally recognized framework for laboratory competence, many industrial laboratories do not pursue formal accreditation due to cost, complexity, and organizational constraints. This study proposes a structured framework to assess and improve compliance with ISO/IEC 17025:2017 without formal accreditation. The framework combines the Analytic Hierarchy Process (AHP) with a gap analysis–based audit model to evaluate both the relative importance of standard requirements and their level of implementation. A two-level hierarchical structure, comprising criteria and sub-criteria aligned with the standard, was developed. Expert judgments were used to derive requirement weights, which were then integrated with audit results to calculate weighted compliance scores. The framework was applied to two testing laboratories within a gas cylinder manufacturing company. Initial compliance levels of 56.1% and 54.9% indicated moderate adherence to the standard. Based on the weighted assessment, thirteen improvement actions were identified and prioritized using an effort–benefit approach. The simulation of five priority actions increased projected compliance to 73.8% and 72.5%, respectively. The results demonstrate that substantial improvements in laboratory compliance can be achieved through targeted and strategically prioritized actions. The proposed framework provides a practical decision-support tool for quality engineering and management in industrial laboratory environments.

Keywords – ISO/IEC 17025:2017, Analytic Hierarchy Process, laboratory compliance, compliance assessment

XXV. Introduction

In an industrial context increasingly driven by quality, safety, and regulatory compliance, the reliability of laboratory testing plays a decisive role in ensuring product conformity and market competitiveness. For manufacturing sectors involving safety-critical products—such as gas cylinders—the consequences of testing failures can be severe, including risks to human safety, environmental damage, and significant economic losses. As a result, testing and calibration laboratories embedded within industrial organizations are required to operate under robust quality and competence frameworks.

The international standard ISO/IEC 17025:2017 establishes general requirements for the competence, impartiality, and consistent operation of testing and calibration laboratories. Compliance with this standard enhances confidence in laboratory results and supports decision-making across production, inspection, and regulatory processes. While full accreditation according to ISO/IEC 17025 is often pursued by independent laboratories, many industrial organizations face substantial financial, organizational, and operational barriers to accreditation, particularly when laboratories primarily serve internal production needs.

In parallel, globalization and increasing competitive pressure have intensified the need for industrial firms to adopt differentiation strategies based on quality, reliability, and safety rather than cost alone. For gas cylinder manufacturers, compliance with construction and testing standards requires a high level of confidence in laboratory results, even in the absence of formal accreditation. Consequently, organizations increasingly seek structured approaches that allow them to assess and improve laboratory performance without committing to the full accreditation process.

Despite the relevance of ISO/IEC 17025, existing studies often focus on accreditation outcomes, compliance audits conducted by accreditation bodies, or binary assessments of conformity. Less attention has been given to structured, quantitative self-assessment frameworks that account for the relative importance of different standard requirements in an industrial context. Moreover, decision-support tools that help prioritize improvement actions based on both impact and effort remain underexplored.

To address this gap, this study proposes a structured framework to evaluate and compare the compliance of testing laboratories with ISO/IEC 17025:2017 requirements using a weighted, quantitative approach. The framework integrates the Analytic Hierarchy Process (AHP) to determine the relative importance of standard requirements, combined with a gap analysis–based audit model to assess current performance. The approach is applied to two laboratories within a gas cylinder manufacturing company, enabling comparison, identification of critical gaps, and prioritization of improvement actions.

The main research question guiding this study is: How can the adherence of industrial testing laboratories to ISO/IEC 17025:2017 requirements be systematically evaluated, compared, and improved without formal accreditation?

The paper contributes to the quality engineering and management literature by (i) proposing a two-level AHP-based weighting structure aligned with ISO/IEC 17025 requirements, (ii) operationalizing a quantitative compliance assessment

suitable for industrial laboratories, and (iii) demonstrating its applicability through a real industrial case study. Unlike traditional checklist-based compliance assessments, the proposed framework incorporates the relative importance of ISO/IEC 17025 requirements through AHP-derived weights and supports the prioritization of improvement actions under resource constraints. The framework therefore combines compliance assessment, requirement criticality, and improvement planning within a single structured decision-support approach tailored to industrial laboratories.

XXVI. Theoretical Background

Quality Management Systems (QMS) are widely recognized as essential instruments for improving organizational performance, ensuring process consistency, and enhancing customer confidence (Oakland, 2014; Evans and Lindsay, 2016). Empirical studies have shown that effective QMS implementation contributes to improved operational control, reduced variability, and enhanced organizational learning (Zu, Fredendall and Douglas, 2008; Psomas, 2013). In laboratory environments, QMS implementation is particularly critical due to the direct impact of testing and calibration results on product conformity, safety, and regulatory compliance (Santos, Mendes and Barbosa, 2011).

ISO 9001:2015 provides a generic framework for quality management applicable to all types of organizations, emphasizing principles such as customer focus, leadership, the process approach, continuous improvement, and evidence-based decision-making (ISO, 2015). Several studies highlight the role of ISO 9001 in promoting structured management practices and process standardization (Fonseca, 2015; Psomas, 2013). However, ISO 9001 does not address the technical competence required to generate valid laboratory results. Consequently, certification to ISO 9001 alone is insufficient to demonstrate laboratory reliability in testing and calibration activities (Hoyle, 2017; Santos, Mendes and Barbosa, 2011).

In laboratory contexts, competence extends beyond documented procedures to include qualified personnel, validated methods, suitable equipment, controlled environments, and metrological traceability (JCGM, 2012; Magnusson and Ellison, 2008). Previous research emphasizes that deficiencies in any of these elements can compromise the results' validity and undermine confidence in laboratory outputs (Ellison and Williams, 2012). Ensuring comprehensive laboratory competence, therefore, requires a standard specifically tailored to laboratory operations, which has led to the widespread adoption of ISO/IEC 17025.

ISO/IEC 17025:2017 defines the general requirements for the competence, impartiality, and consistent operation of testing and calibration laboratories. Applicable to laboratories of any size and scope, the standard is globally recognized as the primary reference for laboratory competence (ISO/IEC, 2017; Westgard and Westgard, 2014). The 2017 revision introduced a structure aligned with modern management system standards and incorporated risk-based thinking, reflecting broader trends in quality and risk management (Fonseca and Domingues, 2018). Its requirements are organized into five main areas—general, structural, resource, process, and management system requirements—covering organizational impartiality, personnel competence, equipment and facilities, method validity, result reporting, and system effectiveness (Belezia and Almeida, 2021; Karthiyayini and Rajendran, 2021, Krismastuti and Habibie, 2022).

A key feature of ISO/IEC 17025:2017 is its flexibility. Laboratories are not required to adopt prescriptive documentation structures, such as mandatory quality manuals, but must demonstrate effective implementation and control of processes (ISO/IEC, 2017). This flexibility allows laboratories embedded within industrial organizations to adapt the standard to their operational context while maintaining technical rigor. Studies indicate that progressive or partial implementation of ISO/IEC 17025 requirements can deliver substantial benefits, including improved reliability of results, reduced non-conformities, and stronger internal quality assurance, even in the absence of formal accreditation (ILAC, 2025; Santos, Mendes and Barbosa, 2011; Westgard and Westgard, 2014).

ISO 9001 and ISO/IEC 17025 are complementary but serve distinct purposes. ISO 9001 focuses on the effectiveness of organizational processes and customer satisfaction, whereas ISO/IEC 17025 emphasizes technical competence and the validity of laboratory results (Oakland, 2014; Evans and Lindsay, 2016). While ISO/IEC 17025 incorporates management principles consistent with ISO 9001, it extends them with laboratory-specific technical requirements. Compliance with ISO/IEC 17025 does not automatically ensure conformity with ISO 9001, nor does ISO 9001 certification guarantee laboratory competence (Hoyle, 2017). Nevertheless, organizations with an established ISO 9001-based QMS often benefit from reduced implementation effort when adopting ISO/IEC 17025, as foundational management practices are already in place (Fonseca, 2015; ISO/IEC, 2017). For industrial laboratories, understanding the relationship between these standards is essential for designing efficient quality systems that avoid redundancy while ensuring both managerial effectiveness and technical reliability.

XXVII. Methodology

A. Overall methodological framework

This study adopts a quantitative, multi-criteria decision-support approach to assess and improve compliance with ISO/IEC 17025:2017 requirements in industrial testing laboratories. The proposed framework integrates a gap analysis–based audit model with the Analytic Hierarchy Process (AHP), enabling the simultaneous evaluation of the level of implementation of standard requirements and their relative importance.

The methodological framework follows four main stages: (i) structuring ISO/IEC 17025 requirements into a hierarchical model, (ii) assessing laboratory compliance through a structured audit, (iii) weighting requirements using expert judgments and AHP, and (iv) prioritizing improvement actions based on weighted compliance results and an effort–benefit analysis. This integrated approach supports transparent assessment, comparability between laboratories, and informed decision-making under resource constraints.

B. Compliance structure and gap analysis

ISO/IEC 17025:2017 requirements were organized into a two-level hierarchical structure composed of criteria and sub-criteria, aligned with the standard's five main requirement groups. This structure, shown in Fig. 4 allows for systematic assessment while preserving traceability to the original clauses of the standard. The main criteria corresponded to the five requirement groups defined in ISO/IEC 17025:2017, namely general, structural, resource, process, and management system requirements, each comprising multiple sub-criteria aligned with the standard.

A gap analysis–based audit was conducted to evaluate the degree of implementation of each sub-criterion. During the audit process, each sub-criterion was assessed using an ordinal scoring scale ranging from 1 (very low implementation) to 5 (full implementation), reflecting the extent to which the corresponding requirement was implemented in practice. This approach enables the identification of gaps between current practices and standard requirements and has been widely applied in quality management and compliance assessment contexts (Santos, Mendes and Barbosa, 2011; Fonseca and Domingues, 2018).

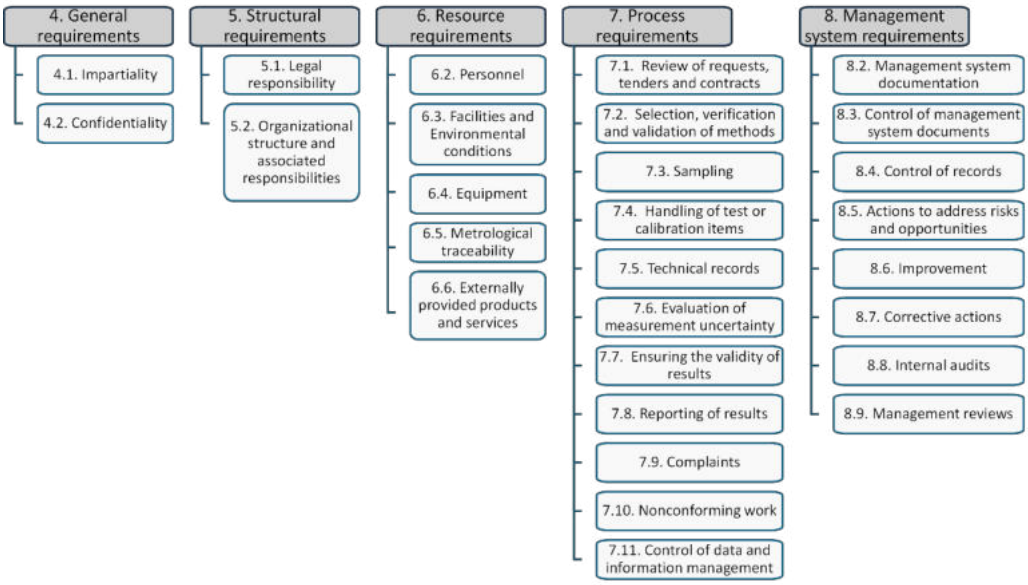


Fig. 4. Hierarchical structure of the evaluation framework based on ISO/IEC 17025:2017.

C. Weighting of requirements using the Analytic Hierarchy Process

To account for the relative importance of ISO/IEC 17025:2017 requirements, the Analytic Hierarchy Process (AHP) was applied. AHP is a well-established multi-criteria decision-making method that decomposes complex problems into hierarchical structures and derives priority weights based on pairwise comparisons between decision elements (Saaty, 1980). Its use is particularly suitable in contexts where both qualitative judgments and quantitative analysis are required, such as the assessment of compliance with technical standards. The method has been widely applied in operations management, quality engineering, and decision-support contexts due to its ability to structure complex evaluation problems and incorporate expert judgment (Subramanian and Ramanathan, 2012).

In this study, ISO/IEC 17025 requirements were structured into a two-level hierarchy comprising criteria and sub-criteria. Individual expert judgments were aggregated using the mean of the pairwise comparison values before calculating the final priority vectors. The pairwise comparisons were performed by experts with experience in laboratory quality management and ISO/IEC 17025 implementation, including professionals directly involved in laboratory operations and compliance activities. Pairwise comparisons were performed using Saaty's fundamental 1–9 scale, which allows experts to express both the direction and intensity of preference between two requirements (Saaty, 1994).

For each level of the hierarchy, pairwise comparison matrices were constructed based on the collected expert judgments. Pairwise comparisons were performed both among the main criteria and among the sub-criteria within each criterion. Priority vectors were then derived using the eigenvalue method, which computes normalized weights representing the relative importance of each criterion and sub-criterion within the hierarchy (Saaty, 1980). Consistency ratios were calculated for all comparison matrices to ensure the logical coherence of expert judgments, and all matrices satisfied the acceptable consistency thresholds recommended in the AHP literature (Saaty, 2008).

The resulting weights represent the relative importance of each criterion and sub-criterion within the hierarchy. For each criterion, the weights of its associated sub-criteria correspond to its internal decomposition within the AHP hierarchy. Expert judgments were aggregated by averaging the individual pairwise comparison values before calculating the final weights.

TABLE VI. presents an illustrative example of a pairwise comparison matrix between the main requirement groups of ISO/IEC 17025. This table is provided for methodological clarification only; the values shown do not correspond to actual expert judgments used in the analysis. The final AHP-derived weights reflect the relative criticality of the standard's requirements and serve as inputs for the weighted compliance assessment.

D. Weighted compliance assessment and prioritization of improvement actions

Weighted compliance scores were calculated by combining the results of the gap analysis–based audit with the AHP-derived weights. This integration allows the assessment to reflect not only the degree of implementation of ISO/IEC 17025 requirements, but also their relative importance for laboratory competence and result reliability. As a result, the weighted compliance score provides an aggregated and more informative measure of overall laboratory performance than unweighted approaches.

TABLE VI. ILLUSTRATIVE EXAMPLE OF PAIRWISE COMPARISON BETWEEN ISO/IEC 17025 REQUIREMENT GROUPS (AHP)

Requirement group	General req.	Structural req.	Resource req.	Process req.	MS req.
General requirements	1	3	5	5	3
Structural requirements	1/3	1	1	7	1
Resource requirements	1/5	1	1	7	1
Process requirements	1/5	1/7	1/7	1	7
Management system requirements	1/3	1	1	1/7	1

For each laboratory, weighted compliance was obtained by multiplying the compliance score assigned to each sub-criterion by its corresponding AHP weight and aggregating the results across all criteria. This procedure enables meaningful comparison between laboratories and facilitates the identification of critical weaknesses, particularly in highly weighted requirements where low implementation levels may have a disproportionate impact on overall compliance.

Based on the weighted compliance assessment, improvement actions were identified for sub-criteria exhibiting low compliance scores and high relative importance. These actions were subsequently prioritized using an effort–benefit approach, which considers both the expected impact of each action on compliance improvement and the estimated effort required for implementation. Actions associated with high expected benefit and relatively low implementation effort were classified as priority interventions, supporting efficient allocation of limited organizational resources.

To further support decision-making, the potential effect of selected improvement actions was simulated. The simulation was performed by updating the compliance scores associated with the selected improvement actions and recalculating the corresponding weighted compliance levels. This procedure allowed the estimation of projected compliance improvements resulting from the implementation of prioritized actions. This simulation allows managers to assess the expected gains from specific improvement strategies before implementation and to compare alternative action sets. Similar effort–benefit prioritization and simulation-based approaches have been widely recommended in quality improvement and decision-support literature as effective tools for planning and performance enhancement (Oakland, 2014; Evans and Lindsay, 2016).

XXVIII. RESULTS AND DISCUSSION

A. Case study description

The proposed framework was applied to two testing laboratories belonging to a gas cylinder manufacturing company, hereafter referred to as Laboratory A and Laboratory B. Both laboratories operate in an industrial production environment and perform testing activities relevant to product conformity and safety. Although neither laboratory holds formal ISO/IEC 17025 accreditation, both are required to ensure reliable testing results and compliance with applicable technical standards.

B. Initial compliance assessment

An initial compliance assessment was conducted using the gap analysis–based audit model described in Section XXVII.C, incorporating AHP-derived weights to account for the relative importance of different ISO/IEC 17025 requirement groups and sub-criteria. The final weights assigned to the main requirement groups are presented in TABLE VII. Among these groups, resource and process requirements received the highest relative weights, reflecting their critical role in ensuring technical competence, measurement reliability, and the consistency of testing activities.

The assessment indicated moderate adherence to ISO/IEC 17025 requirements in both laboratories, with overall compliance levels of 56.1% for Laboratory A and 54.9% for Laboratory B. Although basic quality and technical practices were already in place, the results revealed significant gaps across several requirement groups. By incorporating requirement importance into the assessment, the proposed approach supported the identification and prioritization of critical weaknesses according to their potential impact on laboratory competence and testing reliability.

TABLE VII. FINAL AHP-DERIVED WEIGHTS FOR ISO/IEC 17025 REQUIREMENT GROUPS

Requirement group	Final Weight
General requirements	0.091
Structural requirements	0.156
Resource requirements	0.250
Process requirements	0.272
Management system requirements	0.231

C. Compliance assessment results and interpretation

The incorporation of AHP-derived weights supported a more informed interpretation of the compliance assessment results by accounting for the relative importance of different ISO/IEC 17025 requirement groups. In particular, deficiencies associated with resource and process requirements were identified as more critical due to their higher influence on laboratory competence, equipment reliability, method validity, and the consistency of testing activities.

TABLE VIII. presents the compliance assessment results for the main requirement groups in both laboratories. The results demonstrate how the proposed framework supports the identification and prioritization of critical compliance gaps, providing a structured basis for improvement planning and managerial decision-making in industrial laboratory environments.

TABLE VIII. WEIGHTED COMPLIANCE BY ISO/IEC 17025 REQUIREMENT GROUP

Requirement group	Laboratory A (%)	Laboratory B (%)
General requirements	61.2	68.4
Structural requirements	54.0	57.6
Resource requirements	49.8	52.1
Process requirements	58.6	55.3
Management system requirements	57.1	56.2
Overall weighted compliance	**56.1**	**54.9**

D. Prioritization and Impact of Improvement Actions

Based on the weighted compliance assessment, improvement actions were identified for sub-criteria exhibiting low compliance scores combined with high relative importance. Thirteen improvement actions were defined and evaluated using the effort–benefit approach described in Section XXVII.D.

Fig. 5 presents the effort–benefit matrix derived from the analysis.

Fig. 5. Effort-benefit matrix for improvement actions.

TABLE VIII. summarizes the prioritized improvement actions, highlighting those classified as high-priority interventions due to their high or medium expected impact and relatively low implementation effort (actions 1, 3, 4, 10 and 13). This prioritization supports targeted allocation of resources toward actions with the greatest potential to enhance overall compliance.

To evaluate the potential impact of these actions, a simulation was performed by updating compliance scores for the affected sub-criteria and recalculating weighted compliance levels. The simulation of five priority actions increased projected compliance levels from 56.1% to 73.8% for Laboratory A and from 54.9% to 72.5% for Laboratory B. These results demonstrate that substantial improvements in laboratory compliance can be achieved through a limited number of strategically selected actions.

E. Discussion

The results demonstrate the added value of integrating a weighted, multi-criteria approach into the assessment of laboratory compliance with ISO/IEC 17025:2017. The weighted analysis revealed distinct performance profiles across requirement groups, highlighting critical weaknesses that would not be apparent under an equal-weight assessment. This finding reinforces the importance of considering the relative importance of standard requirements when evaluating laboratory competence.

In both laboratories, higher weights—and thus greater influence on overall compliance—were attributed to resource- and process-related requirements. These groups encompass key elements such as personnel competence, equipment control, metrological traceability, method validation, and handling of non-conforming work. Deficiencies in these areas have a disproportionate impact on the reliability of laboratory results, which explains their higher influence on overall weighted compliance. This outcome is consistent with prior studies emphasizing that technical and operational factors are central to laboratory performance and risk mitigation in testing environments.

TABLE IX. PRIORITIZED IMPROVEMENT ACTIONS BASED ON WEIGHTED COMPLIANCE AND EFFORT–BENEFIT ANALYSIS

No.	Description of Improvement action	Main requirement group	Expected benefit	Implementation effort	Priority
1	Include an impartiality policy in the management manual	General requirements	Medium	Low	High
2	Implement a formal confidentiality declaration	General requirements	Low	Low	Low
3	Develop a code of conduct for suppliers and clients	General requirements	Medium	Low	High
4	Implement an integrated human resource management system	Structural requirements	Medium	Low	High
5	Restrict access to the mechanical testing area	Resource requirements	Low	Medium	Low
6	Improve safety and environmental conditions in Laboratory B	Resource requirements	Low	High	Low
7	Develop software for integration of fatigue testing results	Resource requirements	Low	High	Low
8	Implement measurement uncertainty evaluation	Process requirements	Medium	High	Low
9	Define sample collection locations for Laboratory A	Process requirements	Low	Low	Low
10	Develop a laboratory risk matrix	Management system requirements	Medium	Low	High
11	Establish a method for collecting customer feedback	Management system requirements	Medium	Medium	Low
12	Implement a procedure for internal audits	Management system requirements	Medium	Medium	Low
13	Include laboratory performance in management review	Management system requirements	High	Low	High

The comparison between laboratories further illustrates how weighted compliance supports more meaningful benchmarking. Although Laboratory B exhibited higher compliance in some general and management-related requirements, its lower performance in highly weighted technical areas resulted in a lower overall weighted compliance score. This confirms that similar aggregate compliance levels may conceal substantially different risk profiles, reinforcing the limitations of binary or unweighted conformity assessments frequently used in internal audits.

The prioritization of improvement actions based on an effort–benefit approach proved effective in translating assessment results into actionable managerial guidance. For example, one of the prioritized actions involved the development of a laboratory risk matrix to systematically identify and monitor operational and technical risks associated with testing activities. Another high-priority action concerned the inclusion of laboratory performance indicators in management review processes, allowing the organization to monitor compliance evolution and support continuous improvement initiatives more effectively. These examples illustrate how the proposed framework can support practical and targeted improvements in industrial laboratory environments. The identification of thirteen improvement actions and the subsequent classification of priority interventions enabled a focused improvement strategy aligned with resource constraints. The simulation results show that implementing a limited number of high-priority actions can lead to substantial improvements in overall compliance, with projected increases exceeding 16 percentage points in both laboratories. This highlights the practical relevance of the proposed framework as a decision-support tool rather than a purely diagnostic instrument.

From a methodological perspective, the integration of AHP with gap analysis addresses a gap identified in the literature regarding the lack of structured, quantitative self-assessment tools for non-accredited industrial laboratories. While ISO/IEC 17025 accreditation remains the formal mechanism for demonstrating competence, the results suggest that progressive and targeted compliance strategies can deliver significant benefits in contexts where full accreditation is not immediately feasible. This supports the view that compliance with international standards can be approached as a continuous improvement process rather than a binary outcome.

Despite these contributions, the study has limitations that should be acknowledged. The application of the framework was limited to two laboratories within a single industrial organization, which may constrain the generalizability of the findings. In addition, AHP-based weighting relies on expert judgment, which, although systematically validated through consistency checks, may still be influenced by contextual factors.

Overall, the discussion confirms that the proposed framework enhances the assessment and management of laboratory compliance by linking technical rigor with practical decision-making. By revealing critical weaknesses, supporting prioritization, and enabling impact simulation, the framework contributes to both the academic literature on quality management and the practical needs of industrial laboratory environments.

XXIX. Conclusions And Future Work

This paper presented a structured framework for assessing and improving compliance with ISO/IEC 17025:2017 in industrial testing laboratories operating without formal accreditation. By integrating a gap analysis–based audit with the Analytic Hierarchy Process, the framework enables a weighted and transparent evaluation of laboratory compliance that goes beyond traditional checklist-based approaches.

The application of the framework to two industrial laboratories demonstrated its ability to support informed decision-making by identifying critical compliance gaps and guiding the prioritization of improvement actions under resource constraints. Rather than treating conformity as a binary outcome, the proposed approach supports a progressive and targeted path toward enhanced laboratory competence.

From a practical perspective, the framework offers industrial organizations a viable tool to strengthen internal quality assurance, improve the reliability of testing results, and align laboratory practices with international standards without the immediate burden of formal accreditation. For quality engineers and managers, the combination of weighted compliance assessment and impact simulation provides actionable insights to support planning and continuous improvement initiatives.

From an academic standpoint, this study contributes to the quality management and laboratory competence literature by operationalizing a quantitative, multi-criteria self-assessment approach tailored to industrial laboratory contexts. The proposed framework complements existing accreditation-focused research by addressing the needs of laboratories seeking structured compliance improvement strategies.

Future research could extend the application of the framework to a larger and more diverse set of laboratories, including ISO/IEC 17025-accredited and non-accredited ones, explore the sensitivity of results to alternative weighting schemes, or compare AHP with other multi-criteria decision-making methods.

REFERENCES

Belezia, L. C and Almeida, M. F. L. D. (2021), "Self-assessment model for testing and calibration laboratories based on ISO/IEC 17025: 2017 requirements", Journal of Physics: Conference Series, Vol. 1826 No. 1, pp. 012026.

JCGM (2012), International vocabulary of metrology – Basic and general concepts and associated terms (VIM), 3rd ed. JCGM 200:2012. Sèvres: Joint Committee for Guides in Metrology.

Ellison, S.L.R. and Williams, A. (2012), Quantifying uncertainty in analytical measurement, 3rd ed, Eurachem/CITAC, Teddington, UK.

Evans, J.R. and Lindsay, W.M. (2016), Managing for Quality and Performance Excellence. 10th, Cengage Learning, Boston, US.

Fonseca, L. M. and Domingues, J. P. (2018). "Empirical research of the ISO 9001: 2015 transition process in Portugal: Motivations, benefits, and success factors", Quality Innovation Prosperity, Vol. 22, No. 2, pp.16-46.

Fonseca, L.M. (2015), "ISO 9001 quality management systems through the lens of organizational culture", Quality – Access to Success, Vol. 16, No.148, pp.54–59.

Hoyle, D. (2017), ISO 9000 Quality Systems Handbook, 7th ed., Routledge, London, UK.

ILAC (2025), "The Value of Accreditation", International Laboratory Accreditation Cooperation, available at: https://a2la.org/the-value-of-accreditation/ (accessed 10 march 2026).

ISO (2015), ISO 9001:2015 – Quality management systems — Requirements, International Organization for Standardization.

ISO/IEC (2017), ISO/IEC 17025:2017 – General requirements for the competence of testing and calibration laboratories, International Organization for Standardization.

Karthiyayini, N., and Rajendran, C. (2021), "An approach for benchmarking service excellence in accredited services of Indian calibration and testing laboratories", Materials Today: Proceedings, Vol. 46, Part 17, pp. 8218-8225.

Krismastuti, F. S. H., and Habibie, M. H. (2022), "Complying with the resource requirements of ISO/IEC 17025: 2017 in Indonesian calibration and testing laboratories: current challenges and future directions", Accreditation and Quality Assurance, Vol 27, No. 6, pp. 359-367.

Magnusson, B. and Ellison, S.L.R. (2008), "Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements", Analytical and Bioanalytical Chemistry, Vol. 390, No. 1, pp.201–213.

Oakland, J.S. (2014), Total Quality Management and Operational Excellence, 4th ed, Routledge, London, UK.

Psomas, E. (2013), "The effectiveness of the ISO 9001 quality management system in service companies", Total Quality Management & Business Excellence, Vol. 24, No. 7-8, pp. 769–781.

Saaty, T. L. (1994), Fundamentals of decision making and priority theory with the analytic hierarchy process, 1st ed., McGraw Hill Publishers, New York, US.

Saaty, T.L. (1980), The Analytic Hierarchy Process, 2nd ed., McGraw Hill Publishers, New York, US.

Saaty, T.L. (2008), "Decision making with the analytic hierarchy process", International Journal of Services Sciences, Vol. 1, No. 1, pp. 83–98.

Santos, G., Mendes, F., and Barbosa, J. (2011), "Certification and integration of management systems: the experience of Portuguese small and medium enterprises", Journal of cleaner production, Vol. 19, No. 17-18, pp. 1965-1974.

Subramanian, N. and Ramanathan, R. (2012), "A review of applications of Analytic Hierarchy Process in operations management", International Journal of Production Economics, Vol.138, No. 2, pp.215–241.

Westgard, J.O. and Westgard, S.A. (2014), Basic Quality Management Systems, 1st ed., Westgard QC, Madison, US.

Zu, X., Fredendall, L.D. and Douglas, T.J. (2008), "The evolving theory of quality management: the role of Six Sigma. Journal of Operations Management", Vol. 26, No. 5, pp.630–650.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Evaluating Tourism Service Quality: A Comparative Analysis of Supply Standards and Consumer Expectations

Raffaele Tardio [1], Chiara Cardona [2], Sofia Menestò [3], Silvia Verdura [4], Kasra Hasani [5] and Francesco Picin [6]

[1] PhD student in Economics at University of Tuscia; raffaele.tardio@unitus.it
[2] PhD student in Economics at University of Tuscia; chiara.cardona@unitus.it
[3] PhD student in Economics at University of Tuscia; sofia.menesto@unitus.it
[4] PhD student in Economics at University of Tuscia; silvia.verdura@unitus.it
[5] Student in Economics at University of Venice; kasra6hasani@gmail.com
[6] Student in International Business at HZ University of Applied Sciences; pici0001@hz.nl

Abstract – *The assessment of service quality in tourism increasingly relies on such criteria as consumer expectations, digital channels, consumer reviews, trust symbols, emotions, and consumer experiences as a whole. Despite the fact that the legal requirements, international standards, and voluntary certification systems remain the most reliable means of conveying information about safety, transparency, sustainable operation, etc., they are not always based on the parameters considered by the tourist when choosing a service. In this paper, the problem is analyzed using a qualitative comparative approach to legal regulation, standard and non-standard certifications, consumer selection parameters, and consumer biases in online reviews. Comparing the objective supply-side mechanisms with the demand-side expectations allows highlighting the elements of congruence, partial congruence, and incongruence between the two parties' positions. Based on the findings, the existing standards and regulations are best applicable in objective and compliant issues such as safety, pricing transparency, and service consistency. At the same time, they only partially correspond to the subjective aspects of tourism quality, such as comfort, value, trust, employee behavior, and emotional satisfaction. The theoretical contribution of the study is the identification of a framework for evaluating tourism service quality in terms of a supply-demand relationship and consumer perceptions of quality. The paper aims to contribute to the quality management literature focusing on the goal for steady improvement and service excellence achievement and it also targets building a certification of updated standards to evaluate the actual quality level of tourism services through the development of future research.*

Keywords – *Quality Standards; Service Quality; Standard and non-Standard Certifications; Tourism Service Management.*

I. Introduction

The tourism industry is strongly influenced by customers' desire to explore new experiences therefore there are mechanisms that help tourists to make better decisions. These mechanisms, including standardized international certifications and voluntary eco-labels, prevent information asymmetry between consumers and service providers (Font, 2002). According to Signaling Theory, such instruments function as verifiable institutional signals, including ISO 9001, ministerial star classifications and sector quality marks, that aim to build the information gap and reduce consumer uncertainty in the evaluation of highly experiential services (Spence, 1973; Parasuraman, Zeithaml and Berry, 1988).

However, in recent years the process of decision making has shifted dramatically. Travelers now heavily rely on social media, online platforms and peer reviews to assess service quality which boosts their importance higher than institutional standards (Filieri, 2016). With the advent of digital platforms, consumers are engaged in continuous cycles of information exchange (Gretzel and Yoo, 2008). In this context, Electronic Word of Mouth (eWOM), defined as any positive or negative statement made by current or potential consumers and made available via the Internet (Hennig-Thurau et al., 2004), has taken a central role in the perceived quality evaluation process, progressively replacing traditional institutional certification channels. These information-exchange cycles generate greater perceived value for tourists than any institutional regulation (Gretzel and Yoo, 2008), thereby exposing serious shortcomings in traditional quality assessment systems. As Buckley (2012) notes, "certification schemes often measure what is easy to verify, rather than what tourists actually care about". This dynamic is further amplified by consumers' bounded rationality: faced with information overload, individuals resort to simplified heuristics that privilege the aggregated peer judgment conveyed through eWOM over the analysis of complex and often opaque technical standards (Simon, 1972).

This highlights a potential structural mismatch between institutional requirements and consumer expectations. While certification schemes aim to guide consumer choices, studies show that many tourists are either unaware of these labels or do not rely on them when evaluating hotel quality (Karlsson and Dolnicar, 2016). Corroborating this, research shows that more than half of European consumers are influenced by online reviews when choosing accommodation, while for one in three consumers the official star classification plays no role (Statista, 2022). With regard to sustainability certifications, green eco-labels do influence booking decisions, but to a lesser extent than price, rating and geographical location (Dimara et al., 2023); similarly, comparative empirical research on customer satisfaction between ISO 9001-certified and non-certified properties remains largely absent from the literature (Heras-Saizarbitoria et al., 2015).

Modern tourists seek authenticity, personalized services, and emotional value rather than purely standardized features of accommodation or infrastructure (Poon, 1993). With the rise of digital platforms, consumers have been engaging in

continuous information exchange cycles (Gretzel and Yoo, 2008). These cycles generate value to tourists more than any other institutional regulation which implies critical flaws in traditional rating systems. This shift has resulted in a widening gap between what formal standards measure often operational or technical criteria and what consumers report as important, including friendliness of staff, emotional comfort, uniqueness of the setting, and value for money (Filieri, 2016). The literature has consequently identified a structural gap which this work aims to fill: research has treated online reviews and eWOM as an almost exclusive proxy for perceived quality, systematically neglecting the relative weight of formal and informal certifications and their interaction with user-generated evaluations (Dedeoğlu and Demirer, 2015; Vatankhah et al., 2023). Understanding consumer perspective is therefore crucial as it reveals the misalignment between institutional quality frameworks and actual traveler expectations. So, this paper wants to understand which quality certification standards might be considered in tourism to reflect both operational requirements and experiential dimensions.

II. Methodology

This research primarily relies on desk analysis of different sources in the tourism sector, with particular attention to existing certifications and regulations regarding the tourism industry. Adopting a qualitative approach allowed us to explore how consumers evaluate tourism services and identify where their preferences align or conflict with current certification schemes adopted and proposed by their suppliers. This paper carried out a comparative analysis between consumer needs on the demand side, and the regulations and certifications used by the suppliers to demonstrate the quality of their services. This comparison has been made in order to identify eventual mismatches and partial alignments between supply and demand in tourism service quality evaluation.

III. Legal Certification

In the tourism industry there are some legal constraints and regulations which serve as both regulatory compliance tools and signals of quality and safety to consumers (Pike and Page, 2014). In the tourism and hospitality sector, several legal regulations establish the minimum mandatory requirements for safety, hygiene, consumer protection, and environmental responsibility. These frameworks play a crucial role in ensuring that service providers operate under uniform and enforceable standards. One of the most important is that food safety within hotels and restaurants is legally regulated through HACCP (Hazard Analysis and Critical Control Points) systems. The FAO/WHO (2003) guideline notes that HACCP provides a systematic approach in controlling food safety hazards, ensuring compliance with EU food safety law and protecting consumers. Together, these certifications and regulations integrate legal compliance with operational and sustainability standards, offering establishments a structured framework to meet both statutory obligations and market expectations. As Pike and Page (2014) emphasize, certifications and legal frameworks function not only as regulatory instruments but also as tools for market differentiation and consumer trust. Although these regulations provide a solid framework for consumers to guide through their decision-making process, there is still room to discuss whether consumers actually consider these factors or they see it as evident and need more motivation in their selection (Crespo and Gómez-López, 2021).

Legal protections extend beyond food safety. EU consumer protection law, including Directive 2011/83/EU, ensures transparent information, fair commercial practices, and standardized contract terms in tourism-related transactions (European Commission, 2018). The EU Omnibus Directive (2019/2161) further reinforces these principles by regulating the presentation of online reviews, requiring platforms and businesses to disclose whether reviews are verified, and prohibiting misleading practices such as fake ratings, which enhances consumer trust in digital booking environments (European Commission, 2020). Additionally, the principle of professional diligence—established under Directive 2006/123/EC on Services in the Internal Market and reinforced by the Unfair Commercial Practices Directive—requires tourism providers to act with the level of care and competence expected from a professional operator, thereby indirectly regulating aspects of service delivery and customer treatment (European Commission, 2016). Hotels and restaurants must also comply with occupational health and safety directives—most notably Directive 89/391/EEC—which require employers to ensure safe working environments (EU-OSHA, 2019). Similarly, fire safety and building regulations, such as the EU Construction Products Regulation (305/2011), mandate structural compliance and the implementation of emergency procedures. In many European countries, accommodation providers must additionally meet national tourism licensing laws and state-mandated hotel classification systems that regulate minimum service standards, equipment, hygiene requirements, and operational procedures (Pike and Page, 2014; OECD, 2020). Accessibility requirements—governed by Directive 2019/882 on accessibility of products and services—further ensure that tourism facilities accommodate visitors with disabilities (European Commission, 2021).

TABLE I. Legal Regulations

Legal Regulations	*Description*
HACCP (Hazard Analysis and Critical Control Points)	Systematic approach to identifying and controlling food safety hazards in hotels and restaurants, ensuring compliance with EU food safety law.
EU Consumer Protection Law (Directive 2011/83/EU)	Ensures transparent information, fair commercial practices, and standardized contract terms in tourism transactions.
Occupational Health and Safety (Directive 89/391/EEC)	Requires employers to provide safe working environments for employees in the hospitality sector.
Fire Safety and Building Regulations (EU Construction Products Regulation 305/2011)	Mandates structural compliance, emergency procedures, and safety standards in buildings used for tourism.
National Tourism Licensing Laws / Hotel Classification Systems	Set minimum operational, equipment, hygiene, and service standards for accommodation establishments.
EU Omnibus Directive (2019/2161)	EU Omnibus Directive (2019/2161)
Directive 2006/123/EC (Services Directive)	Establishes the principle of professional diligence, requiring tourism providers to act with competence, honesty, and care expected of a professional operator

IV. Standard And Non-Standard Certifications

In addition to legal regulations, tourism and hospitality establishments often adopt standard and non-standard certifications. These tools signal quality, sustainability, and operational excellence to consumers (Font, 2002). Unlike legal regulations, standard and non-standard certifications in tourism and hospitality are generally voluntary. While they provide guidance on quality management, environmental performance, and sustainability, establishments are not legally obligated to adopt them (ISO, 2015; ISO, 2018).

It is nonetheless worth noting that the landscape of voluntary standards is evolving: the forthcoming ISO 9001:2026, whose Draft International Standard was approved in December 2025 and whose publication is expected in autumn 2026 with a three-year transition period, signals a broader shift toward embedding quality culture, ethical conduct, and organizational resilience into management systems that go well beyond mere compliance (Draft ISO, 2025).

Among the most widely recognized international standards are the ISO series, which provide structured frameworks for management and environmental performance. ISO 9001 establishes a quality management system that helps organizations "consistently meet customer and regulatory requirements and enhance satisfaction" (ISO, 2015). The forthcoming ISO 9001:2026, will further strengthen this framework by introducing explicit requirements around quality culture, ethical behavior, and organizational resilience, positioning the standard as a more strategic tool for service excellence (DRAFT ISO, 2025). Hotels and tour operators often implement ISO 9001 to standardize service processes and improve customer experience, and the 2026 revision is expected to make this particularly relevant for hospitality providers, as it emphasizes leadership commitment to quality culture and the integration of sustainability considerations, including climate change, directly into the quality management system (ISO, 2025). Environmental management is addressed through ISO 14001, which guides organizations in reducing environmental impacts while complying with legal obligations. According to ISO (2015), ISO 14001 "enables organizations to achieve the intended outcomes of their environmental management system, including enhancing environmental performance". Tourism-specific standards, such as ISO 21401 and ISO 18513, further target the hospitality sector. ISO 21401 defines sustainability management for accommodation establishments, stating that it "provides requirements for establishing, implementing, maintaining and improving a sustainable management system" (ISO, 2018), while ISO 18513 standardizes service quality and terminology in tourist accommodation to ensure transparency and consistency for guests (ISO, 2003).

TABLE II. Standard and Non-Standard Certifications

Certification / Standard	*Type*	*Purpose*
ISO 9001 – Quality Management System	Standard	Provides a framework for quality management, ensuring organizations consistently meet customer expectations and regulatory requirements.
ISO 14001 – Environmental Management System	Standard	Guides organizations in reducing environmental impacts while complying with legal obligations.
ISO 21401 – Sustainability Management for Accommodation	Tourism-specific standard	Defines requirements for establishing and improving sustainability management systems in hotels and lodging facilities.
ISO 18513 – Tourism Services / Hospitality Terminology	Tourism-specific standard	Standardizes service quality and terminology in accommodations for consistency and transparency.
EMAS – Eco-Management and Audit Scheme	EU-backed certification	Enables organizations to monitor, report, and improve environmental performance based on legally recognized standards.
EU Ecolabel for Tourist Accommodation	EU-backed certification	Establishes environmental criteria for accommodation services, promoting sustainable practices.
Green Key	Voluntary certification	Signals sustainable and environmentally friendly practices in hotels and tourism establishments.
Blue Flag	Voluntary certification	Recognizes high environmental and quality standards for beaches and tourism facilities.
ISO 9001 – Quality Management System	Standard	Provides a framework for quality management, ensuring organizations consistently meet customer expectations and regulatory requirements.

There are also EU- backed certifications such as EMAS (Eco-Management and Audit Scheme), which "enables organizations to evaluate, report, and improve their environmental performance according to legally recognized standards" (European Commission, 2016), and the EU Ecolabel, which establishes environmental criteria for accommodation services (European Commission, 2023). Other widely recognized voluntary certifications are Green Key, Blue Flag, and GSTC-approved certifications, which signal sustainable practices, social responsibility, and quality to environmentally conscious travelers (Font, 2002).

V. Impact Of Customer Review

Despite their importance and usefulness, online reviews usually contain biases that prevent decision-making processes from being genuine and efficient. Reviews left online are generally positive. According to the data gathered from online platforms such as TripAdvisor in 2020, more than 82 percent of reviews in this platform are 4 or 5-stars, with nearly two-third 5-stars (TripAdvisor, 2025). This pattern is also repeated in other agencies like Airbnb, with average ratings around 4.7/5 and 90 percent of reviews at 4 or 5-stars (Skotis, Morfaki and Livas, 2023). In addition to positivity bias, negative reviews, although less frequent, have a greater impact (Skotis, Morfaki and Livas, 2023). Aggregate ratings heavily influence new reviewers.

Experiments have shown that previous high scores lead to higher subsequent ratings. Early customer reviews, if positive, can create a "rating bubble" that perpetuates high scores in future reviews (Cicognani, Figini and Magnani, 2022). Platforms like TripAdvisor and Booking.com for example, which prominently display review averages, may unintentionally encourage even more gregarious behavior. This bias compromises the independence of future reviews.

Reviewers self-select the option that most have to approve a review for it to be published: so, they generally publish comments from those who have had positive experiences, while those with moderate and average experiences are underrepresented. The "purchase bias" is also important: people expect to enjoy what they have booked and paid for. There is inconsistency between written feedback and numerical ratings because of personal review and judgment styles (Skotis, Morfaki and Livas, 2023). Some users always assign very high or very low ratings, thus making poor ratings absent. Without adjustment, aggregate ratings reflect these user-specific biases (Skotis, Morfaki and Livas, 2023).

Platform structures create bias among customers by modeling review bias. TripAdvisor, for instance, does not have a clear "neutral" score; 3 stars is often seen as negative, which pushes scores up. Trustpilot is vulnerable to biased input, such as companies pushing positive reviews or users "bombarding reviews," both positive and negative. The weakness of moderation in this area allows companies to game the system and thus influence the entire review ecosystem within the site. Design choices (e.g., scoring scales or default prompts) change the way users rate (Cicognani, Figini and Magnani, 2022).

False or misleading reviews remain common in the industry. TripAdvisor removed nearly 1 million fake reviews in 2020 (3.6 percent of the total) using automation for most detections (TripAdvisor, 2025). However, malicious positive or negative reviews continue to slip through and thus be present on the site. Even real reviews can be misleading: a user may write praise but assign 3/5 because of unclear personal standards or standards that differ from those intended by the platform itself or other users. Identifying such inconsistencies remains a key challenge.

TABLE III. CONSUMER BIAS

Type of Bias	***Description***	***Example Platform Issue***
Positivity Bias	Reviews are overwhelmingly positive, making it hard to differentiate truly exceptional experiences from average ones.	TripAdvisor had over 82% 4- or 5-star reviews in 2020; Airbnb shows a similar trend with 90% of reviews at 4 or 5 stars.
Negativity Bias	Though less frequent, negative reviews have a stronger impact on consumer decisions than positive ones.	A single 1-star review can deter more than several 5-star reviews attract.
Social Influence Bias	Earlier high ratings influence subsequent reviews, creating a "rating bubble."	TripAdvisor displays review averages that can encourage this "herding effect."
Selection Bias	Reviewers often self-select, meaning positive experiences are overrepresented, while moderate ones are underrepresented. Personal review styles can also lead to inconsistencies.	"Purchase bias" (expecting to enjoy what you've paid for) and user-specific rating habits can skew aggregate ratings.
Platform Design Bias	How a platform is designed (e.g., scoring scales, default prompts, lack of a neutral option) can influence how users' rate.	TripAdvisor lacks a clear "neutral" score, pushing ratings up.
Fake/Unreliable Reviews	False or misleading reviews persist, and even real reviews can be inconsistent with their numerical ratings due to unclear user standards.	TripAdvisor removed nearly 1 million fake reviews (3.6% of total) in 2020, but malicious reviews still appear.

VI. CONSUMER DECISION PARAMETERS

Tourist preferences when selecting a service tend to be shaped by a multifaceted set of parameters. These factors include objective elements (such as price and location) as well as subjective, psychological, or experiential elements, such as comfort, emotional satisfaction, perceived trust, and social proof (Mohamad, Idris and Manshoor, 2024). Empirical studies consistently confirm that service quality, amenities, and trust remain strong predictors of booking decisions. In the analysis conducted by Milawati et al. (2022), service quality, hotel facilities, and perceived trustworthiness significantly affected guest choice, suggesting that consumers value both tangible amenities and perceived reliability. Beyond basic features, physical comfort, aesthetics, emotional attachment, and cleanliness/safety are considered crucial in consumers' decision-making processes (Sirirat et al., 2024). According to Lee and Kim (2020), psychological comfort—how guests feel about and perceive the environment—is just as important as technical aspects. In the era of online booking, information signals such as customer reviews, review volume, and overall rating score play an increasingly important role. Using a two-stage choice model, recent research shows that during the initial "consideration set" stage, consumers pay attention to price, discounts, overall rating, review volume, and popularity; in the final booking stage, these variables continue to influence decisions (Hu and Yang, 2020). Additionally, staff behavior acts as a key emotional and psychological factor that shapes the overall guest experience beyond regulatory compliance or facility quality. Hospitality providers that invest in cultivating staff empathy and professionalism gain a competitive advantage through improved guest perceptions and repeat bookings (Sawant and Sakpal, 2021). Another important factor in consumers' decision making is perceived quality, which refers to the trade-off between the price paid and the quality expected. This concept emphasizes that consumers frequently weigh costs relative to expected comfort and service performance rather than simply selecting the cheapest option (Nsue Bela, 2024).

Complementing these elements, recent literature highlights additional determinants that have become increasingly influential in contemporary consumer decision making. Risk perception and health-related safety have gained significant importance, with variables such as flexible cancellation policies, hygiene protocols, and the reliability of information provided by the service provider shaping consumer preferences, particularly in the post-pandemic context (Bae & Chang, 2021).

Digital reputation, transparency in pricing, and consistency between online promises and the actual experience also play a crucial role in building or weakening consumer trust (Hlee, Lee & Koo, 2018). Furthermore, the co-creation of the experience, in which tourists actively participate in tailoring and personalizing services, enhances emotional engagement and strengthens behavioral intentions (Campos et al., 2018). Finally, the tourism brand image and its ability to convey identity-driven values such as authenticity, sustainability, and a sense of belonging influence consumer preferences, as travelers increasingly seek providers aligned with their personal values (Kladou & Kehagias, 2014).

TABLE IV. GAP ANALYSIS OF CONSUMER NEEDS AND EXISTING STANDARDS AND REGULATIONS

Consumer Needs	*Relevant Rule / Certification*	*Coverage*
Service Quality and Comfort	ISO 9001 (quality management)	Partially covered
Price and Discounts Transaparency	EU Consumer Rights Directive (2011/83/EU)	Fully covered
Emotional Saristaction	No direct regulation or standard	Not covered
Perceived Quality	EU Unfair Commercial Practices Directive (2005/29/EC)	Partially covered
Trust and Social Proof	EU Omnibus Directive (2019/2161)	Partially covered
Staff Behaviour	EU regulations apply only indirectly (via "professional diligence" in 2006/123/EC + UCPD)	Partially covered

VII. RESULTS

Our goal in the results section is to uncover whether there is any mismatch between regulation, standard and non-standard certification, and consumers' actual needs in order to fulfill the research gap highlighted by the literature about the linking in service perceived quality of these three elements (Dedeoğlu and Demirer, 2015; Vatankhah et al., 2023). This mismatch becomes particularly evident when consumer feedback and online reviews are analyzed alongside formal quality frameworks. Previous literature consistently shows that the dimensions tourists prioritize in their evaluations, such as staff behavior, room comfort, emotional atmosphere, and perceived value, differ substantially from the dimensions assessed by certification schemes, which tend to focus on operational consistency, procedural compliance, and verifiable physical features (Parasuraman, Zeithaml and Berry, 1988; Saleh and Ryan, 1991). Empirical studies drawing on user-generated content confirm this divergence: analyzing over 60,000 traveler ratings from online platforms, Stringam and Gerdes (2010) found that hotel service quality and room comfort were the dimensions most strongly correlated with overall satisfaction and willingness to recommend, neither of which is directly addressed by ISO 9001 or equivalent certification schemes. More recently, Vatankhah et al. (2023) proposed an extended SERVQUAL framework that incorporates smart service quality dimensions and identifies significant discrepancies between what formal standards measure and what consumer groups actually perceive as high quality. This structural divergence between standardized quality indicators and perceived service quality expressed through eWOM and customer reviews represents the analytical lens through which the following results should be read.

The relationship between consumer needs and existing tourism regulations reveals significant mismatches between what guests value and what regulatory frameworks actually cover. A core consumer priority is service quality and comfort, which is only partially covered by existing frameworks.

An additional perspective that helps interpret these mismatches comes from the literature on sustainability-oriented service innovation. The article "Managing sustainability-oriented innovation in services: proposal of a scale for the future of hotel companies and travel agencies" (Ruíz-Molina et al., 2024) shows how tourism firms are increasingly integrating innovative practices that combine business performance with environmental, social, and economic sustainability goals. The authors identify five key dimensions of Sustainability-Oriented Service Innovation (SOSI): technological innovation, non-technological innovation, and economic, social, and environmental sustainability. This contribution is particularly relevant to our study because it highlights that guests' perceived quality does not depend solely on regulated aspects or elements that can be measured in a standardized way, but also on experiential, organizational, and relational factors. The non-technological and social dimensions of the SOSI model underscore the importance of factors such as staff behavior, customer care, and the creation of a welcoming environment—elements that, while central to consumers, are not covered by current regulatory frameworks. The integration of SOSI practices within tourism firms therefore confirms the limitation of relying exclusively on regulatory compliance: to address customers' real needs, a broader approach is required, one that incorporates innovative processes oriented toward sustainability and experiential quality.

ISO 9001 provides a general quality management structure, but it is intentionally broad and does not specify service quality elements such as room comfort, ambiance, or guest experience (ISO, 2015). Most certification schemes focus on internal processes rather than consumer-facing experiential quality (Font, 2002). Consequently, while ISO 9001 can help tourism businesses standardize operations, it does not address the subjective comfort and service expectations that travelers commonly prioritize (Heras-Saizarbitoria, Arana and Boiral, 2015). This gap is confirmed by evidence from consumer-generated feedback: building on eWOM data from hotel review platforms, research shows that the attributes guests most frequently mention, room quality, staff responsiveness, and comfort, are precisely those that formal standards treat as peripheral or leave entirely unaddressed (Stringam and Gerdes, 2010).

By contrast, price and discount transparency is fully covered, reflecting strong EU commitment to consumer rights. The Consumer Rights Directive (2011/83/EU) ensures clear price information, cancellation rights, and transparent contract terms—factors that heavily influence booking decisions (European Commission, 2018). When hotels provide pricing information in a

transparent, sufficiently detailed way, consumers perceive the price as fairer, and their willingness-to-pay increases (Miao and Mattila, 2007).

Emotional satisfaction, an important psychological and experiential factor, refers to how the environment makes guests feel, a dimension shown to play a central role in hotel choice (Lee and Kim, 2020). By analyzing existing standards, we realized there was no standard or regulation referring to this aspect of decision making. For instance, Spence (2022) notes that multisensory appeal and emotional cues are critical for guest satisfaction and return intentions, none of which are measured by ISO or EU regulatory frameworks. This absence becomes particularly telling when compared with the content of consumer reviews: qualitative analysis of online hotel reviews reveals that emotional and relational dimensions, how guests feel about the atmosphere, the staff, and the overall experience, are among the most frequently discussed topics in user-generated content, yet they remain entirely absent from formal certification criteria (Vatankhah et al., 2023). Capturing emotional satisfaction requires quantifiable measures. Instruments such as the Destination Emotion Scale (DES) have been developed and validated to measure guests' feelings, emotional reactions, and experiential satisfaction (Hosany and Gilbert, 2010).

Another partially mentioned factor regards trust and social proof. The Omnibus Directive (EU 2019/2161) strengthened consumer-protection rules by requiring platforms to ensure the authenticity of online reviews and prohibiting manipulation or posting of fake feedback, thereby enhancing transparency in digital marketplaces (European Parliament, 2019). This is particularly relevant in tourism and hospitality, as social proof strongly influences hotel-booking decisions. Studies show travelers perceive Consumer Generated Media (CGM) to be more trustworthy than official destination websites or mass media. In other words, credible reviews supply powerful social proof that shapes consumers' perceptions of value and trust and drives booking intentions (Filieri, 2015; Li and Mattila, 2007). Importantly, the distinction between objective quality, as measured by star ratings and certification schemes, and perceived quality as expressed through eWOM is not merely terminological: research drawing on a large European dataset of over 5,000 hotels demonstrates that both objective quality, as measured by star ratings, and perceived quality, as expressed through consumer scores, significantly influence eWOM behavior, This suggests that what consumers reward through their online evaluations goes beyond what formal classification systems are designed to capture (Sánchez-González and González-Fernández, 2021). Regulation alone cannot fully resolve trust issues because online review systems remain susceptible to bias and manipulation. The OECD notes that fake or incentivized reviews and the suppression of negative feedback persist and that ratings often rely on small, unrepresentative samples (OECD, 2019).

Furthermore, the UCPD prohibits misleading commercial practices such as false advertising or omission of material information, its coverage is largely limited to verifiable characteristics such as price, features, guarantees, or terms of service (European Commission, 2018). Consequently, it primarily protects consumers from objective misrepresentations rather than subjective experiences. Research in service management shows that perceived quality is a multidimensional construct reflecting the trade-off between what consumers pay and what they receive in terms of comfort, amenities, and service performance (Zeithaml, 1988; Parasuraman, Zeithaml and Berry, 1988). Therefore, even when all advertised information is accurate, guests may perceive poor value-for-money if the actual experience does not meet their expectations, leaving this dimension largely unprotected under EU consumer law (WIPO, 2023). In addition, perceived service quality is influenced by intangible factors such as staff responsiveness, attentiveness, and reliability, which cannot be fully standardized or verified prior to consumption (Ladhari et al., 2017; Karatepe, 2013). This gap underscores the limitations of relying solely on transparency and truthful marketing to guarantee consumer satisfaction.

Staff behavior in hospitality including responsiveness, professionalism, and interpersonal interactions is a critical determinant of customer satisfaction and service quality (Ladhari et al., 2017; Karatepe, 2013). This aspect is partially covered in UCPD (Directive 2005/29/EC) as professional diligence, but its potential is constrained by its generality. Although the UCPD imposes a duty of "professional diligence," this requirement is deliberately broad and flexible. The Costs Lawyer Standards Board notes that professional diligence is an objective standard of special skill and care commensurate with honest market practice and the general principle of good faith. Even by considering the baselines, variations in attentiveness, responsiveness, or interpersonal skills can still lead to negative guest experiences (Ladhari et al., 2017).

In addition to certifications mentioned above, some basic legal requirements like HACCP are often perceived as basic legal requirements rather than value-adding features. A 2025 survey of 916 U.S. consumers on food-safety expectations found that 85% of respondents considered food safety a minimum quality standard, yet they rated it as secondary to other attributes such as freshness, product quality or local origin (Archila-Godínez et al., 2025). Consequently, although essential for safety, certifications like HACCP compliance typically play little role in shaping perceived quality or guiding consumer decisions.

Taken together, these findings confirm that the gap between formal quality standards and consumer expectations is not incidental but structural: as Saleh and Ryan (1991) demonstrated in an early application of the SERVQUAL model to hospitality, gaps between management perceptions and guest expectations are a persistent source of dissatisfaction, and the evidence from contemporary online reviews suggests that this gap has not narrowed, but has instead shifted toward experiential, emotional, and relational dimensions that current regulatory frameworks are ill-equipped to address.

VIII. Conclusion

The present study analyses the connection between formal tourism service quality signals and expectations that shape consumer decision making. On the one hand, there is sufficient evidence that rules, standards and certifications remain indispensable in setting minimum levels of safety, transparency, sustainability and reliability. On the other hand, these elements of service quality do not reflect the whole gamut of criteria used by customers when choosing a service. The most apparent evidence supporting this assumption comes from comparing six major needs that consumers have with existing regulatory and certification frameworks. Thus, only price and discount transparency seems to be fully covered by rules and certification, whereas service quality and comfort, perceived quality, trust and social proof and staff behavior have some overlap with existing frameworks. Emotional satisfaction of consumers is not mentioned at all, which reflects a greater focus

on measurable criteria. It appears that standards and regulations tend to cover areas where verification is possible and necessary, whereas those aspects that relate to subjective perceptions, interaction and emotional side of services are neglected. Another issue that is worth mentioning relates to how crucial online reviews are in terms of consumer decision making. It seems that the fact that almost half of the reviews on TripAdvisor are positive as well as the fact that 975 thousand fake reviews were removed from the platform in 2020 suggests that online reviews offer good social proof while being biased towards the positive experience of consumers. In a similar way, evidence related to food safety suggests that legal compliance of a product/service should not be regarded as something special or unique because the customer views it as a minimum requirement, rather than as something that sets a service apart. Therefore, the findings of the study can contribute to the literature in terms of introducing a new approach to assessing tourism service quality – namely, the supply-demand alignment approach. According to this concept, tourism service quality should not be defined based on what suppliers are able to prove or certify, but according to whether formal quality signals correlate with criteria that shape consumer value perception. Thus, even if this contribution is conceptual, it offers important insights into why some existing frameworks may seem adequate from a supply-side perspective while being insufficient from a consumer-side perspective.

IX. Research Limitations

While our analysis reveals several mismatches between consumer needs and present certifications/regulations used by suppliers, it is grounded only in secondary sources. This approach, consistent with documentary comparative reviews, allows for the construction of a service evaluation framework, but does not permit empirical verification of consumer preferences or measurement of the actual impact of certifications on purchasing decisions. Future studies could address this gap through structured surveys or discrete choice experiments targeting tourists, aimed at quantifying the relative weight of different quality attributes in the decision-making process. In addition, our review may not consider all emerging demands and regulations or standards which might develop in future years.

X. Further Research

Future research might confirm empirically the results achieved in the present study by looking at the priorities that customers use while deciding which indicators have greater weight in their preferences. The use of surveys, interviews, consumer behavior tests, and analysis of platform data might help investigate whether such indicators as emotional satisfaction, behavior of personnel, quality perception, trust, and credibility of reviews have more importance in the tourists' decisions than certificates or compliance. Future research might also work on creating a revised version of the tourism quality certificate model that includes such criteria as operational, legal, sustainability, emotional, interpersonal, and review-based factors. So, the results mentioned in this study open up an opportunity to modify and re-create existing quality certifications and evaluation parameters in the tourism industry. Developing new models that better capture the dynamics between providers' offerings and travelers' expectations could improve alignment between market demand and market supply. Furthermore, a new model to be developed through future research can enable business actors in this sector to tailor their offerings to better respond to the market demand on the quality of tourism services.

References

Archila-Godínez, J. C., Kotanko, C., Wiatt, R., Marshall, M. I. and Feng, Y. (2025) "Consumers' food safety expectations and risk perceptions of produce from small and medium-sized farms", Journal of Food Science, 90(9), pp. e70527. doi: 10.1111/1750-3841.70527.

Bae, S.Y., Chang, P.-J. (2021). "The effect of coronavirus disease-19 (COVID-19) risk perception on travel intention".

Buckley, R. (2012) "Sustainable tourism: Research and reality", Annals of Tourism Research, 39(2), pp. 528–546.

Campos, A.C., Mendes, J., Valle, P.O., and Scott, N. (2018). "Co-creation of tourist experiences: A literature review and research agenda".

Costs Lawyer Standards Board (2025) Professional diligence: Guidance note. Costs Lawyer Handbook. London: CLSB.

Crespo, A. and Gómez-López, R. (2021) "Quality-certified hotels and sustainability: Do certifications influence consumers?", Sustainability, 13(22), pp. 12599.

Dedeoğlu, B. B., and Demirer, H. (2015). Differences in service quality perceptions of stakeholders in the hotel industry. International Journal of Contemporary Hospitality Management, 27(1), 130–146.

Dimara, E., Baloglu, S., and Fyall, A. (2023). The importance of green certification labels/badges in online hotel booking choice: A conjoint investigation. International Journal of Hospitality Management, 113, 103528.

EU-OSHA (2019) Directive 89/391/EEC: Safety and health at work (EU Occupational Safety & Health Agency guidance). Luxembourg: Publications Office of the European Union.

European Commission (2016) The Services Directive (2006/123/EC) and professional diligence: Implementation report. Brussels: European Commission.

European Commission (2018) Consumer Rights Directive (2011/83/EU): Key facts for businesses and consumers. Brussels: European Commission.

European Commission (2020) Omnibus Directive (EU 2019/2161): New rules for consumer protection. Brussels: European Commission.

European Commission (2021) European Accessibility Act: Directive (EU) 2019/882 on accessibility requirements for products and services. Brussels: European Commission.

European Commission (2023) EU Ecolabel criteria for tourist accommodation: 2023 revision. Brussels: European Commission.

European Parliament (2019) "Directive (EU) 2019/2161 of 27 November 2019 on better enforcement and modernisation of Union consumer-protection rules (Omnibus Directive)", Official Journal of the European Union, L328, pp. 7–28.

Filieri, R. (2015) "What makes online reviews helpful? A diagnosticity–adoption framework to explain informational and normative influences in e-WOM", Journal of Business Research, 68(6), pp. 1261–1270.

Filieri, R. (2016) "What makes an online consumer review trustworthy?", Annals of Tourism Research, 58, pp. 46–64.

Filieri, R., Alguezaui, S. and McLeay, F. (2015) "Why do travellers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth", Tourism Management, 51, pp. 174–185.

Font, X. (2002) "Environmental certification in tourism and hospitality: Progress, process and prospects", Tourism Management, 23(3), pp. 197–205.

Gretzel, U. and Yoo, K.-H. (2008) "Use and impact of online travel reviews", in O'Connor, P., Höpken, W. and Gretzel, U. (eds.) Information and communication technologies in tourism 2008. Vienna: Springer, pp. 35–46.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., and Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? Journal of Interactive Marketing, 18(1), 38–52.

Heras-Saizarbitoria, I., Arana, G. and Boiral, O. (2015) "Do ISO 9001-certified hotels get a higher customer rating than non-certified ones?", International Journal of Hospitality Management, 49, pp. 98–104.

Hlee, S., Lee, H. and Koo, C. (2018) "Hospitality service quality and trust: the role of digital reputation and information consistency", Sustainability, 10(8), p. 2991.

Hosany, S. and Gilbert, D. (2010) "Measuring tourists' emotional experiences: Further validation of the Destination Emotion Scale", Journal of Travel Research, 49(4), pp. 513–526.

Hu, X. and Yang, Y. (2020) "Determinants of consumers' choices in hotel online searches: A comparison of consideration and booking stages", International Journal of Hospitality Management, 86, pp. 102370.

Ipsos (2024) Understanding consumer needs in relation to food hygiene ratings in an online food ordering environment. London: FSA Research and Evidence. doi: 10.46756/001c.123520.

ISO (2003) ISO 18513:2003 – Tourism services — Hotels and other types of tourism accommodation — Terminology. Geneva: International Organization for Standardization.

ISO (2015a) ISO 9001:2015 – Quality management systems – Requirements. Geneva: International Organization for Standardization.

ISO (2015b) ISO 14001:2015 – Environmental management systems – Requirements with guidance for use. Geneva: International Organization for Standardization.

ISO (2018) ISO 21401:2018 – Tourism and related services – Sustainability management system for accommodation establishments – Requirements. Geneva: International Organization for Standardization.

ISO (2025) ISO/DIS 9001 – Quality management systems – Requirements (Draft International Standard). Geneva: International Organization for Standardization. Available at: https://www.iso.org/standard/88464.html

Karatepe, O. M. (2013) "High-performance work practices and hotel employee performance: The mediating role of work engagement", International Journal of Hospitality Management, 32, pp. 132–140.

Karlsson, L. and Dolnicar, S. (2016) "Does eco certification sell tourism services? Evidence from a quasi-experimental observation study in Iceland", Journal of Sustainable Tourism, 24(5), pp. 694–714.

Kladou, S. and Kehagias, J. (2014) "Assessing destination brand equity: An integrated brand equity model", Journal of Brand Management, 21(1), pp. 25–40.

Ladhari, R., Souiden, N. and Dufour, B. (2017) "The role of emotions in utilitarian service settings: Effects of emotional satisfaction on product perception and behavioural intentions", Journal of Retailing and Consumer Services, 34, pp. 10–18.

Lee, M.-J. and Kim, H.-J. (2020) "Psychological comfort and its impact on hotel satisfaction and loyalty", International Journal of Hospitality Management, 91, pp. 102639.

Li, M. and Mattila, A. S. (2007) "How do price transparency and market norm information impact consumers' fairness perceptions in the restaurant industry?", Journal of Foodservice Business Research, 10(4), pp. 7–26.

Miao, L. and Mattila, A. S. (2007) "How and how much to reveal? The effects of price transparency on consumers' price perceptions", Journal of Hospitality & Tourism Research, 31(4), pp. 530–545.

Milawati, A., Budiyono, B. and Yuliani, Y. (2022) "Service quality, facilities and trust as predictors of hotel selection", Journal of Tourism and Hospitality Management, 10(3), pp. 120–132.

Mohamad, M., Idris, N. and Manshoor, N. (2024) "Decision-making factors in hotel selection: Price, location, trust and emotional satisfaction", Tourism Economics, 30(1), pp. 182–198.

Nsue Bela, J. (2024) "Perceived quality and the price–quality trade-off: Evidence from tourist accommodations", Journal of Travel Research, 63, pp. 1127–1139.

OECD (2019) Understanding online consumer ratings and reviews. OECD Digital Economy Papers, No. 264. Paris: OECD Publishing.

OECD (2020) Tourism policy and regulation: National licensing and classification frameworks. OECD Tourism Papers, 2020/2. Paris: OECD Publishing.

Parasuraman, A., Zeithaml, V. A. and Berry, L. L. (1988) "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality", Journal of Retailing, 64(1), pp. 12–40.

Pike, S. and Page, S. J. (2014) "Destination marketing organizations and destination marketing: A narrative analysis of roles and responsibilities", Tourism Management, 41, pp. 202–227.

Poon, A. (1993) Tourism, technology and competitive strategies. Wallingford: CAB International.

Ruíz-Molina, M. E., Gil-Saura, I., Berenguer-Contrí, G. and Belda-Miquel, S. (2024) "Managing sustainability-oriented innovation in services: Proposal of a scale for the future of hotel companies and travel agencies", Journal of Tourism Futures. doi: 10.1108/jtf-09-2023-0209.

Saleh, F. and Ryan, C. (1991) "Analysing service quality in the hospitality industry using the SERVQUAL model", Service Industries Journal, 11(3), pp. 324–345.

Sánchez-González, G. and González-Fernández, A. M. (2021) "The influence of quality on eWOM: a digital transformation in hotel management", Frontiers in Psychology, 11, p. 612324.

Sawant, P. and Sakpal, P. (2021) "Determinants of guest satisfaction: The role of empathy and professionalism in service staff", Journal of Hospitality and Tourism Insights, 4(2), pp. 167–180.

Simon, H. A. (1972) "Theories of bounded rationality", Decision and Organization, 1(1), pp. 161–176.

Sirirat, P., Thampanichwat, W., Pongsermpol, C. and Moorapun, S. (2024) "The hotel architectural design factors influencing consumer destinations: A case study of three-star hotels in Hua Hin, Thailand", Buildings, 14(8), p. 2428. doi: 10.3390/buildings14082428.

Skotis, P., Morfaki, M. and Livas, C. (2023) "Rating distributions in peer-to-peer accommodation: Exploring positivity bias on Airbnb", International Journal of Contemporary Hospitality Management, 35(5), pp. 1845–1862.

Spence, C. (2022) "Sensehacking the guest's multisensory hotel experience: Designing for all five senses", i-Perception, 13(5), pp. 1–22.

Spence, M. (1973) "Job market signaling", The Quarterly Journal of Economics, 87(3), pp. 355–374.

Statista / Booking.com (2022) Online hotel reviews are more important to consumers than star classification. Statista Research. Available at: https://www.statista.com/chart/28992/

Stringam, B. B. and Gerdes Jr, J. (2010) "An analysis of word-of-mouse ratings and guest comments of online hotel distribution sites", Journal of Hospitality Marketing & Management, 19(7), pp. 773–796.

TripAdvisor (2021) TripAdvisor transparency report 2020. Needham, MA: TripAdvisor LLC.

Vatankhah, S., Alizadeh, R. and Ariffin, A. A. M. (2023) "Smart service quality in hospitality: A quantitative assessment using MCDM and clustering methods", International Journal of Hospitality Management, 115, p. 103590.

WIPO (2024) Global intangible asset values in 2023: The intangible asset economy. Geneva: World Intellectual Property Organization.

Zeithaml, V. A. (1988) "Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence", Journal of Marketing, 52(3), pp. 2–22.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Quality 4.0 Unlocked: an implementation roadmap for its progressive adoption in Portuguese companies

Diana Oliveira[1], Maria João Rosa[2] and Helena Alvelos[3]

1 DEGEIT, University of Aveiro; dianao@ua.pt

1 CIPES & DEGEIT, University of Aveiro; m.joao@ua.pt

1 CIDMA & DEGEIT, University of Aveiro; helena.alvelos@ua.pt

Abstract*—The digital transformation driven by the Industry 4.0 paradigm is reshaping manufacturing environments and creating new opportunities for the evolution of quality management systems. In this context, Quality 4.0 has emerged as the integration of traditional quality management practices with advanced digital technologies, enabling more data-driven, predictive, and interconnected approaches. Despite growing interest, empirical guidance on how organizations can effectively implement Quality 4.0 remains limited. This paper aims to identify practical approaches that support the successful and sustainable implementation of Quality 4.0 in industrial organizations, based on the results of a mixed-method research study, which combined a systematic literature review, a survey of Portuguese manufacturing companies, and an in-depth case study of an industrial organization undergoing Industry 4.0 transformation. The study findings reveal that the adoption of Quality 4.0 is typically gradual and structured, beginning with the consolidation of traditional quality management practices as a foundation, followed by the digitalization of quality processes and data collection, the development of data-driven analytical capabilities, and the integration of advanced technologies such as artificial intelligence, machine learning, and computer vision. Successful implementation also depends on strong leadership commitment, organizational culture, and workforce competencies, highlighting the central role of human and managerial factors alongside technological adoption. Based on these insights, the paper proposes a practical roadmap that guides organizations through progressive stages of Quality 4.0 adoption, emphasizing the alignment of strategy, technology, processes, and people. This roadmap offers actionable guidance for industrial organizations seeking to advance towards fully integrated, data-driven quality management within the Industry 4.0 context, maximizing performance, efficiency, and competitive advantage.

Keywords: Quality 4.0; Industry 4.0; Implementation roadmap

I. Introduction

The rapid digital transformation of industrial systems, driven by the emergence of the Industry 4.0 paradigm, is fundamentally reshaping manufacturing environments and organizational processes (Lee et al., 2022). Technologies such as the Internet of Things (IoT), cyber-physical systems, advanced analytics, and artificial intelligence (AI) are enabling unprecedent levels of connectivity, automation, and real-time data exchange across production systems. These developments are transforming not only operational processes but also the way organizations manage and control quality throughout the value chain (Chiarini, 2020).

Traditionally, quality management has relied on well-established frameworks and methodologies such as the ISO 9000 standards, Total Quality Management (TQM), Lean, and Six Sigma, which emphasize continuous improvement, process control, and customer focus (Sony et al., 2021). While these approaches remain essential for ensuring process stability and operational excellence, the increasing availability of digital technologies and large volumes of operational data has created new opportunities to enhance quality management practices through real-time monitoring, predictive analytics, and automated decision support (Weckenmann et al., 2019; Oliveira et al., 2025). This transformation has led to the emergence of the concept of Quality 4.0, which integrates traditional quality management principles, methodologies and tools with advanced digital technologies to support more proactive, data-driven approaches to quality (Maganga & Taifa, 2022; Antony et al., 2024).

Despite the growing interest in Quality 4.0, much of the existing literature remains largely conceptual, focusing on frameworks, maturity models, and theoretical discussions. Empirical evidence on how organizations actually adopt Quality 4.0, particularly across different industrial contexts and levels of digital maturity, remains limited (Oliveira et al., 2025). Existing frameworks frequently identify key dimensions and technological enablers of Quality 4.0 adoption, but often provide limited empirical validation and little practical guidance regarding how organizations can progressively implement these practices in real industrial settings. In particular, many models present Quality 4.0 as a high-level maturity objective, without sufficiently addressing the sequential and operational challenges faced by companies during implementation (Mittal et al., 2022; Antony et. al, 2024; Sony et al, 2021; Zulqarnain et al., 2022).

This gap is especially relevant in countries where industrial structures are strongly characterized by small and medium-sized enterprises (SMEs), which often face specific challenges related to limited resources, technological capabilities, and organizational readiness for digital transformation (Jokovic et al., 2023; Wawak et al., 2023). Understanding how companies combine traditional quality practices, digital technologies, data management, and organizational capabilities is therefore essential for supporting the practical implementation of Quality 4.0.

Based on previous empirical findings, this paper proposes a practical roadmap aimed at supporting organizations in the progressive and sustainable adoption of Quality 4.0. By linking empirical evidence with practical recommendations, the research contributes to a better understanding of how companies can navigate the transition toward data-driven quality management in the context of Industry 4.0. Unlike many existent conceptual frameworks, the proposed roadmap is grounded in empirical evidence collected through a mixed-method approach combining literature analysis, survey data, and an industrial case study. The roadmap emphasizes the progressive integration of traditional quality management practices, digitalization, data-driven capabilities, and advanced technologies, while also considering organizational and human-related factors that influence implementation.

II. Literature Review

The emergence of the Industry 4.0 paradigm has significantly transformed industrial environments through the integration of cyber-physical systems, the Internet of Things (IoT), cloud computing, and advanced automation (Lee et al., 2022). In this context, traditional quality management approaches are being complemented by digital technologies capable of enabling real-time monitoring, advanced analytics, and more proactive forms of quality control (Chiarini, 2020).

This transformation has led to the concept of Quality 4.0, commonly defined as the integration of established quality management principles, methodologies and tools with emerging digital technologies such as artificial intelligence, machine learning, big data analytics, and computer vision (Maganga & Taifa, 2022; Antony et al., 2023). Rather than replacing traditional quality methodologies, Quality 4.0 extends and enhances them by enabling data-driven insights, automation, and improved decision-making across the production system.

A central element of Quality 4.0 is the growing importance of data-driven quality management. The increasing availability of operational data from sensors, connected equipment, and integrated information systems allows organizations to move from reactive quality control toward predictive and preventive approaches (Oliveira et al., 2025). Technologies such as Manufacturing Execution Systems (MES), Enterprise Resource Planning (ERP), and automated inspection systems enable real-time monitoring of processes and improved traceability across the value chain (Weckenmann et al., 2019).

However, the implementation of Quality 4.0 involves more than technological adoption. The literature highlights the importance of organizational factors such as strategic alignment, leadership commitment, workforce competencies, and change management (Sharma & Joshi, 2023). Digital transformation in quality management therefore requires organizations to simultaneously address technological, organizational, and cultural dimensions.

Although empirical evidence remains limited, several studies have already proposed conceptual frameworks, maturity models, and implementation roadmaps to support the adoption of Quality 4.0, highlighting key dimensions such as data integration, digital capabilities, leadership, and organizational readiness (Antony et al., 2023; Maganga & Taifa, 2022; Sony et al., 2021). However, these studies tend to focus either on the technological dimensions of Quality 4.0 (Jamkhaneh et al., 2022; Sureshchandar, 2022) or on its maturity level assessment (Nenadál et al., 2022; Schumacher et al., 2016), while offering limited guidance on how organizations can operationalize the transition from a traditional quality management system towards fully integrated Quality 4.0 approach.

In addition, prior research suggests that organizational size and digital maturity influence the adoption of Industry 4.0 and Quality 4.0 practices. Larger firms typically possess greater financial and technological resources, while small and medium-sized enterprises often face constraints related to infrastructure, skills, and investment capacity (Jokovic et al., 2023; Wawak et al., 2023; Mittal et al., 2018; Frank et al., 2019). As a result, the transition toward Quality 4.0 is frequently gradual and heterogeneous across organizations.

This research seeks to address these gaps by proposing an empirically grounded roadmap for small and medium sized enterprises based on evidence collected from Portuguese industrial companies. The proposed roadmap is intended to provide structured guidance for organizations seeking to align traditional quality practices with emerging digital technologies, even if their practical application is still evolving.

III. Methodology

This study uses the results obtained from a mixed-method research approach based on three complementary sources of information: a systematic literature review (Oliveira et al., 2025), an in-depth case study in a Portuguese industrial company (Oliveira et al., 2026), and a survey applied to Portuguese manufacturing firms (Oliveira et al., 2024; 2025). The integration of these sources allowed both a broad and contextualized understanding of the current state of Quality 4.0 implementation and supported the development of practical recommendations for organizations. The use of a mixed-method approach also enabled methodological triangulation, combining quantitative and qualitative evidence to strengthen the robustness and validity of the findings.

The literature review provided the conceptual foundation of the study, enabling the identification of key dimensions associated with Quality 4.0, including digital technologies, quality management practices, organizational readiness, and workforce competencies (Oliveira et al., 2025). These dimensions guided the design of the empirical research instruments and the interpretation of the results. In addition, the literature review supported the definition of the main analytical assumptions explored in the empirical phase, namely the relationship between Industry 4.0 maturity, organizational characteristics, and the adoption of Quality 4.0 practices.

To complement the literature review, an in-depth case study was conducted in a Portuguese manufacturing company engaged in Industry 4.0 transformation initiatives (Oliveira et al., 2026). Methodologically, the case study was developed in two phases. In the first phase, the company's progress toward Industry 4.0 was analysed using an adapted version of the readiness model proposed by Lichtblau et al. (2015). While maintaining the original structure of the model, the adapted instrument was

used to obtain a descriptive overview of the company's digital transformation initiatives and the implementation of Industry 4.0 technologies, rather than to determine a formal maturity level. The instrument was completed by the company's industrial engineer responsible for Industry 4.0 projects, and the collected information was analysed to identify patterns, best practices, and gaps related to digital transformation and quality management.

The second phase focused on the analysis of Quality 4.0 practices through qualitative research. Eight semi-structured interviews were conducted with all employees of the quality department, including the department director, a senior quality technician, a laboratory technician, and five quality controllers. Different interview guidelines were developed according to the role of each participant in order to ensure contextual relevance and obtain detailed insights regarding operational and managerial quality practices. The interviews addressed topics such as quality management methodologies and tools, data collection and analysis practices, workforce competencies, quality control mechanisms, and ongoing Quality 4.0 initiatives.

The interviews were fully transcribed and subjected to content analysis. To support the analysis, a coding framework was developed based on the main themes covered in the interviews' guidelines and aligned with the objectives of the study. The coding scheme included the following thematic categories: (i) quality management methodologies and tools; (ii) data collection and analysis practices; (iii) quality control mechanisms; (iv) workforce competencies and skills; and (v) Quality 4.0 projects and digital initiatives. An inductive approach was subsequently adopted within each category to identify recurring themes, patterns, and relationships emerging from the interviews data.

To improve the reliability and consistency of the analysis, the coding structure was iteratively reviewed throughout the analytical process, and the interpretation of the results was continuously compared with evidence collected from on-site observations and company documentation. This triangulation process contributed to reducing interpretation bias and strengthening the credibility of the findings.

In addition to the interviews, one of the authors conducted on-site observations of the company's operations, routines, and quality management practices. The observations enabled a better understanding of the production environment, the technologies used to support digital transformation, and the real-time quality control mechanisms implemented in practice. Informal conversations with employees also provided complementary insights that supported the interpretation of the interviews' data and the results obtained during the first phase of the case study.

In addition, a survey-based study was conducted to assess the implementation of Quality 4.0 practices across Portuguese industrial companies (Oliveira et al., 2025). Based on the results of the literature review and the insights from the case study, a questionnaire was developed, with a set of questions organized according to the flowing sections: i) the company's profile; ii) Industry 4.0 maturity level; iii) quality management practices in place; and iv) Quality 4.0 implementation level. The survey was distributed online to manufacturing companies operating in Portugal. In total, 1,957 questionnaires were distributed, resulting in 173 valid responses.

The survey results analysis combined descriptive statistics and hypothesis testing to explore relationships between companies' characteristics, Industry 4.0 maturity, and Quality 4.0 adoption. Descriptive statistics was used to characterize the sample and identify general implementation patterns of both Industry 4.0 and Quality 4.0 approaches. Hypothesis tests were subsequently applied to identify the existence of statistically significant differences in Quality 4.0 adoption between different types of companies (micro, small and medium-sized enterprises (SMEs) vs. large companies; companies that are still in their initial/preparatory stages of developing an I4.0 strategy vs. companies actively implementing/consolidating an I4.0 strategy). Given the ordinal nature of the variables under analysis and the absence of guaranteed normality assumptions, the Mann–Whitney U and the Kruskal–Wallis non-parametric tests were employed. Additionally, chi-square tests of independence were applied to assess associations between categorical variables.

The integration of the survey results with the qualitative insights obtained from the case study enabled a comprehensive analysis of Quality 4.0 implementation and supported the development of the practical roadmap presented in the results section. This integration process allowed the comparison and triangulation of quantitative data with qualitative evidence from the case study, contributing to a more detailed understanding of how Quality 4.0 implementation evolves in practice across different organizational contexts.

IV. Results

The aforementioned three complementary sources of empirical evidence, contributed with distinct but interrelated insights into the nature, challenges, and maturity of Quality 4.0 implementation in Portuguese industrial companies, as presented in Fig. 1.

From a theoretical perspective, the literature review characterizes Quality 4.0 as an evolution of traditional quality management, where digital technologies enhance, rather than replace, established methodologies. It also highlights a strong emphasis on real-time data and the digitalization of quality management systems, while stressing that traditional quality principles remain essential. However, the literature reveals a significant lack of empirical evidence and practical implementation guidelines, reinforcing the need for real-world validation of Quality 4.0 implementation roadmaps.

From a practical perspective, the case study demonstrated that the transition toward Quality 4.0 is gradual and hybrid in nature, marked by the coexistence of traditional and digital quality practices, with strong emphasis on system integration, real-time monitoring, and workforce development. The findings emphasize strengths such as technological integration, and the strategic role of quality, while also identifying key limitations, including low interoperability, partial digitalization, and limited use of advanced analytics. These results show that real-world implementation is progressive, incomplete, and multidimensional.

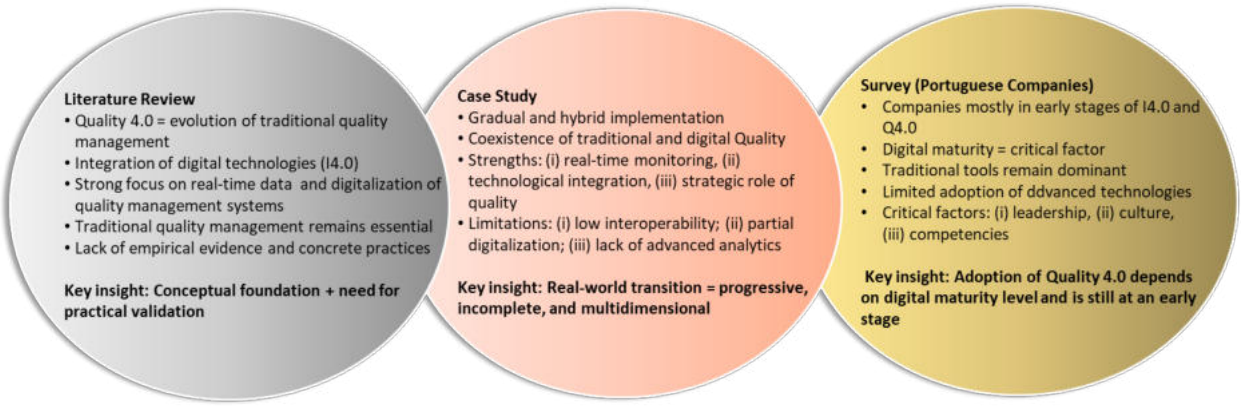


Fig. 8. Synthesis of the key findings from the mixed-method research approach (literatura review, case study, and survey.)

The survey results show that most Portuguese industrial companies are still at early stages of Industry 4.0 and Quality 4.0 adoption, with digital maturity emerging as a key determinant of implementation, while traditional quality tools remain widely used across all organizations and advanced technologies are still limited. Additionally, across all data sources, organizational and human factors, such as leadership, culture, skills, and change management, consistently emerge as critical enablers of successful Quality 4.0 adoption.

Overall, these findings indicate that the transition toward Quality 4.0 is not a purely technological shift but a progressive and multidimensional transformation, involving the alignment of quality management principles, digital technologies, data capabilities, and organizational readiness.

Based on these empirical findings, a practical implementation roadmap has been developed to support organizations in the progressive and sustainable adoption of Quality 4.0. The roadmap is structured into six sequential and complementary phases, presented in the following sections. Each phase identifies key practices and highlights their implications for the successful implementation of Quality 4.0.

A. *Phase 1 – Strengthening Quality Management Foundations*

The implementation of Quality 4.0 begins with the consolidation of traditional quality management practices and methodologies, which provide a stable foundation for digital transformation. **Key practices include**:

- Standard quality approaches: inspection, quality control, quality assurance, and quality improvement;
- Structured methodologies: PDCA cycle, 5S, FMEA, 8D, and A3 Reports;
- Quality management tools: flowcharts, cause and effect diagrams, Pareto diagrams, 5 Whys, control charts.

Implications: Before introducing digital or advanced Quality 4.0 technologies, companies must ensure their quality processes are standardized, stable, and free of inefficiencies, providing a reliable base for further transformation.

The survey results reveal that traditional quality approaches, namely inspection and quality control, are transversal and widely adopted across all companies' sizes and level of digital maturity. In parallel, larger organizations tend to adopt more structured methodologies and analytical tools, suggesting higher levels of formalization and process maturity.

The case study findings further reinforce the importance of process standardization and continuous improvement practices as prerequisites for digital transformation, particularly through the integration of Lean initiatives and structured quality management routines.

These findings suggest that Quality 4.0 implementation should not be interpreted as a purely technological transition. Instead, organizations must first ensure that quality processes are stable, standardized, and capable of generating reliable and consistent operational information before progressing towards more advanced digitalization initiatives.

B. *Phase 2 – Digitalization of Quality Data*

Step 2 in the journey to Quality 4.0 involves digitalizing quality-related data, moving from manual or paper-based collection to automated and system-supported processes. **Key practices include**:

- Automating data collection across production and quality functions;
- Integrating data with ERP or MES systems;
- Utilizing data visualization tools, such as Power BI or specialized statistical software (e.g., R, Minitab, SPSS);

Implications: Digital data collection is essential for accurate, timely, and actionable quality information, acting as the backbone of Quality 4.0 adoption.

The empirical evidence obtained from the survey demonstrates that the level of digitalization of quality data collection is strongly associated with both organizational size and digital maturity. Larger firms tend to present more advanced digitalized data collection systems, whereas micro and smaller firms still rely more heavily on manual or hybrid approaches. Additionally, the results indicate that Excel remains the most widely used tool for quality data processing across all organizational profiles, reflecting its consolidated role as a foundational quality management instrument. However, companies at more advanced stages of Industry 4.0 implementation report higher adoption of ERP systems, Power BI, and specialized statistical software, including real-time monitoring and automated statistical process control (SPC), demonstrating a progression towards more integrated and analytical quality management systems.

The case study findings corroborate these results by highlighting the implementation of paperless systems, automated quality data collection, and MES integration projects aimed at enabling real-time monitoring and operational visibility.

Overall, the findings suggest that digitalized quality data collection constitutes a critical enabler of Quality 4.0, as it provides the informational infrastructure necessary to support real-time analysis, process visibility, and data-driven quality management practices.

C. Phase 3 – System Integration and Real-Time Visibility

Once data is digitalized, organizations can advance to systems integration and the establishment of real-time monitoring of processes and quality indicators. **Key practices include:**

- Connecting sensors, devices, and production systems via IoT, cloud platforms, and mobile interfaces;
- Implementing dashboards and business intelligence tools to track KPIs in real time;
- Integrating quality data with production and operational systems (MES/ERP).

Implications: This phase transforms quality management from reactive control to proactive, data-driven oversight, preparing the organization for analytics-driven quality practices.

The survey findings show that organizations with higher levels of Industry 4.0 maturity demonstrate significantly greater adoption of system integration technologies, including IoT, mobile connectivity, cloud computing, and real-time process monitoring systems. These organizations also report higher levels of automated statistical process control and real-time visualization of quality indicators.

The case study evidences further illustrate the practical implementation of these concepts through the development of dashboards, integration of production and quality systems, and deployment of automated monitoring platforms capable of supporting real-time operational decision-making.

The empirical results therefore suggest that systems integration represents a critical transition point in the Quality 4.0 journey, enabling organizations to move from reactive quality control practices toward proactive and continuous quality monitoring, supported by real-time operational visibility.

D. Phase 4 – Data-Driven Quality Management

With integrated and real-time data, companies can move towards analytics-based decision-making. **Key practices include:**

- Using statistical and predictive analytics to support quality decisions;
- Applying data-driven approaches for process improvement and defect prevention;
- Developing competencies in Big Data analysis, structured problem solving, and data-driven decision making.

Implications: Developing analytical capabilities and fostering a culture of data-driven decision-making is critical for the successful adoption of a Quality 4.0 approach.

The empirical findings reveal that although Portuguese industrial companies generally demonstrate having human resources with strong human and relational competencies, such as teamwork, leadership, adaptability, and cross-functional cooperation, more advanced analytical competencies remain comparatively underdeveloped. In particular, Big Data analysis capabilities and knowledge of advanced Industry 4.0 technologies are still rather incipient, especially among less digitally mature organizations. Remarkably, companies at advanced stages of Industry 4.0 implementation report significantly stronger competencies in data-driven decision-making, structured problem-solving, Big Data analysis, and technological knowledge, indicating that digital maturity strongly influences the development of analytical capabilities within the quality function.

The case study findings reinforce this perspective by identifying increasing organizational emphasis on predictive analytics, real-time process analysis, and data-supported operational decision-making as part of ongoing Quality 4.0 initiatives.

These findings suggest that the transition toward Quality 4.0 depends not only on technological adoption, but also on the development of analytical competencies and a culture capable of supporting evidence-based and data-driven decision-making practices.

E. Phase 5 – Adoption of Advanced Quality 4.0 Technologies

After establishing analytics-driven practices, companies can adopt advanced digital technologies to further enhance quality management. **Key technologies include:**

- Artificial Intelligence (AI) and Machine Learning for predictive quality control;
- Computer vision and augmented reality (AR) for defect detection and process monitoring;
- Collaborative robots and automation of quality-related processes.

Implications: Advanced technologies' adoption represents the final stage of Quality 4.0 maturity, enabling higher automation, predictive capabilities, and quality integration across the value chain.

The empirical findings indicate that the adoption of advanced Quality 4.0 technologies remains limited among Portuguese industrial companies and is largely concentrated in organizations with higher levels of Industry 4.0 maturity. Survey results show that most companies are still focused on foundational digital transformation activities, such as process digitalization, system integration, and data management, while more advanced technologies are not yet widely implemented.

The case study findings reinforce this observation. Although the company has introduced technologies such as artificial vision systems for defect detection, automatic sorting systems, RFID-based traceability, and real-time process monitoring, these initiatives coexist with traditional quality management practices, including manual inspections and sample-based controls. Furthermore, predictive analytics and autonomous quality management capabilities have not yet been implemented, indicating that the transition towards more advanced Quality 4.0 technologies is still ongoing.

These findings suggest that many organizations are currently establishing the digital and analytical foundations required for more sophisticated Quality 4.0 applications. As digital maturity increases, technologies such as artificial intelligence, machine learning, augmented reality, computer vision, and collaborative robotics may progressively support quality management activities by enhancing automation, predictive capabilities, and decision-making effectiveness. Therefore, the adoption of advanced Quality 4.0 technologies represents a longer-term stage of maturity that builds upon the successful implementation of earlier digitalization and analytics-driven initiatives.

F. Phase 6 – Organizational Transformation and Capability Development

Throughout the Quality 4.0 journey, the study's empirical results highlight that organizational and human factors are critical to success. **Key practices include**:

- Strong top management involvement and leadership commitment;
- Employee engagement and cross-functional collaboration;
- Change management and the development of an organizational culture supportive of innovation;
- Employees training and development on Quality 4.0 skills, including data literacy and data-driven decision-making, statistical and analytical skills, knowledge of digital and automation technologies (e.g., IoT, AI, and information systems), as well as transversal competencies such as problem-solving, adaptability, and systems thinking.

Implications: Quality 4.0 requires a holistic transformation, combining people, processes, and technology, where organizational readiness is as important as the digital infrastructure.

The survey results identify top management involvement, employee engagement, organizational culture, and change management as some of the most critical success factors for Quality 4.0 implementation. It is important to note that these organizational and human dimensions are perceived as more critical than purely technological aspects. Additionally, companies at more advanced stages of Industry 4.0 implementation attribute significantly greater importance to organizational culture, people involvement, connectivity, scalability, and analytical infrastructures, reinforcing the strong relationship between organizational maturity and digital transformation success.

The case study findings further emphasize the importance of continuous employee training, adaptability, leadership commitment, and organizational readiness in supporting the integration of Quality 4.0 initiatives within operational practices.

Overall, the results suggest that Quality 4.0 should be understood as a holistic organizational transformation process involving the simultaneous alignment of people, processes, technologies, and organizational culture. Consequently, organizational readiness and capability development emerge as equally important as technological infrastructure for achieving successful Quality 4.0 implementation.

V. DISCUSSION

The findings of this study reinforce the idea that Quality 4.0 implementation should not be understood as a purely technological transition, but rather as a progressive organizational transformation involving the alignment of quality management practices, digital technologies, analytical capabilities, and organizational culture. The empirical evidence obtained through the literature review, survey, and case study consistently suggests that organizations tend to evolve gradually from traditional quality management approaches towards more integrated and data-driven practices.

One of the main contributions of this research is the proposal of an empirically grounded implementation roadmap that translates existing conceptual discussions into a more operational and sequential approach. While previous studies have frequently focused on maturity models (Nenadál et al., 2022; Schumacher et al., 2016) or technological dimensions

(Jamkhaneh et al., 2022; Sureshchandar, 2022), the proposed roadmap emphasizes the practical progression between different stages of Quality 4.0 adoption and highlights the interdependence between technological, organizational, and human-related factors.

However, the findings also reveal that the implementation of Quality 4.0 involves important trade-offs and challenges across the different phases of the roadmap. In the early stages, organizations often face difficulties related to process standardization and the consolidation of traditional quality management practices. Although digital technologies may accelerate transformation, insufficiently stabilized processes can compromise the effectiveness of digitalization initiatives and generate inconsistent or low-quality data.

As organizations progress toward digital integration and real-time monitoring, additional challenges emerge regarding system interoperability, infrastructure investment, cybersecurity, and data governance. The empirical findings suggest that many companies, particularly SME, still face important limitations in terms of technological resources, digital competencies, and financial capacity, which may slow down the implementation process.

The transition toward data-driven quality management and advanced technologies such as artificial intelligence and machine learning also introduces organizational and managerial risks. These include resistance to change, lack of analytical skills, difficulties interpreting complex data, excessive dependence on technology, and challenges associated with integrating digital tools into existing operational routines. In this context, the study reinforces that technological adoption alone is insufficient to ensure successful Quality 4.0 implementation without appropriate organizational readiness and workforce development.

The results additionally suggest that the proposed roadmap may not be equally applicable across all industrial contexts. Organizations operating in highly regulated industries, low-digital-maturity environments, or sectors characterized by limited technological infrastructure may require slower or adapted implementation approaches. Similarly, smaller firms may encounter greater barriers due to resource constraints and limited access to specialized knowledge and advanced technologies. Therefore, the roadmap should be interpreted as a flexible and adaptive framework rather than a rigid implementation model.

VI. Conclusions

The findings of this study demonstrate that the successful adoption of Quality 4.0 depends on the coordinated integration of processes, people, and technology. Traditional quality management practices, such as inspection and quality control, and methodologies like the PDCA and the 5S provide a stable foundation that supports digital transformation and enables organizations to structure their quality systems effectively. Building on this foundation, companies advance in their journey towards Quality 4.0 by digitalizing processes, adopting integrating management systems, such as ERP and MES, and implementing real-time monitoring, which together form the essential infrastructure for a Quality 4.0 adoption. The development of data-driven decision-making capabilities and predictive analytics further enhances operational performance and strategic decision-making, allowing organizations to leverage insights from increasingly complex data streams. Advanced technologies, including artificial intelligence, machine learning, computer vision, and collaborative robotics, emerge as powerful enablers but are most effective when deployed after the consolidation of foundational practices and digital integration. Equally important is the readiness of people and the organization, as strong leadership commitment, employee engagement, supportive culture, and effective change management are decisive factors in transforming quality practices. Overall, the study provides empirical evidence of the need for a staged approach to Quality 4.0 adoption in which organizations progress from traditional practices to digitalization, system integration, advanced analytics, and, finally, technology-driven and human-centered organizational transformation. Although much of the existent literature on the topic already claims for such an approach, the novelty of our work is to propose it based on empirical evidence, based on the analysis of the Portuguese industrial organisations' context. This implementation roadmap illustrates that Quality 4.0 is not merely a technological challenge but a holistic transformation that, when implemented effectively, drives sustainable quality performance, enhances competitiveness, and prepares companies for the demands of Industry 4.0.

Although this study provides relevant empirical insights, some limitations should be acknowledged. First, the survey focused exclusively on Portuguese industrial companies, which may limit the generalizability of the findings to other countries or industrial contexts with different levels of digital maturity, economic structures, or technological infrastructures. Second, the case study was conducted in a single organization, which limits the possibility of fully capturing the diversity of Quality 4.0 implementation practices across industries. Third, the cross-sectional nature of the survey restricts the analysis of how Quality 4.0 adoption evolves over time.

Future research could expand the geographical scope of the analysis, compare implementation patterns across different industrial sectors and countries, and develop longitudinal studies to better understand the long-term evolution of Quality 4.0 maturity. Additional research could also explore sector-specific implementation challenges and develop quantitative maturity assessment models capable of supporting benchmarking between organizations.

XXX. Acknowledgments

This work was supported by CIDMA (https://ror.org/05pm2mw36) under the Portuguese Foundation for Science and Technology (FCT, https://ror.org/00snfqn58), projects UID/04106/2025 (https://doi.org/10.54499/UID/04106/2025) and UID/PRR/04106/2025 (https://doi.org/10.54499/UID/PRR/04106/2025) and by CIPES (https://ror.org/020r0s364) under the Portuguese FCT (FCT, https://ror.org/00snfqn58), projects UIDB/00757/2025 (DOI: https://doi.org/10.54499/UID/00757/2025) and UID/PRR2/00757/2025 (DOI: https://doi.org/10.54499/UID/PRR2/00757/2025).

XXXI. REFERENCES

Antony, J., Kaul, A., Bhat, S., Sony, M., Kaul, V., Zulfiqar, M., and McDermott, O. (2024), "Critical failure factors for Quality 4.0: an exploratory qualitative study". International Journal of Quality and Reliability Management, Vol. 41 No. 4, pp. 1044–1062. https://doi.org/10.1108/IJQRM-07-2023-0240

Chiarini, A. (2020). Industry 4.0, quality management and TQM world. A systematic literature review and a pro posed agenda for further research. The TQM Journal 32 (4):603–16. doi: 10.1108/TQM-04-2020-0082.

Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. International Journal of Production Economics, 210, 15–26. https://doi.org/10.1016/j.ijpe.2019.01.004

Jamkhaneh, H. B., Jalali, R., Shahin, R., Lima, R. M., Rasouli, E., and Jamali, G. (2022). "Analysis and Prioritization of Quality 4.0 dimensions and companies' readiness to adapt to industry 4.0 evolutions through Bayesian Best-Worst method." Proceedings of the 5th ICQEM Conference, University of Minho, 1-17

Jokovic, Z., Jankovic, G., Jankovic, S., Supurovic, A., and Majstorović, V. (2023), "Quality 4.0 in Digital Manufacturing – Example of Good Practice". Quality Innovation Prosperity, Vol. 27 No. 2, pp.177–207. https://doi.org/10.12776/QIP.V27I2.1870

Lee, C. H., Li, L., Wang, D., Wee, H. J. (2022). Strategic servitization design method for Industry 4.0-based smart intralogistics and production. Expert Systems with Applications 204

Maganga, D. P., Taifa, I. W. R. (2022). Quality 4.0 transition framework for Tanzanian manufacturing industries. TQM Journal 35(6), 1417-1448

Mittal, S., Khan, M. A., Romero, D., & Wuest, T. (2018). A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs). Journal of Manufacturing Systems, 49, 194–214. https://doi.org/10.1016/j.jmsy.2018.10.005

Nenadál, J., Vykydal, D., Halfarová, P., and Tylečková, E. (2022). "Quality 4.0 Maturity Assessment in Light of the Current Situation in the Czech Republic". Sustainability (Switzerland), 14(12). https://doi.org/10.3390/su14127519

Oliveira, D., H. Alvelos, and M. J. Rosa. (2025). Quality 4.0: Results from a systematic literature review. The TQM Journal 37 (2):379–456. doi: 10.1108/TQM-01-2023-0018.

Oliveira, D., Alvelos, H., Rosa, M.J. (2026). Quality 4.0 in practice: The case of a Portuguese ceramics company. Quality Management Journal, 1–46 https://doi.org/10.1080/10686967.2025.2591612

Oliveira, D., Rosa, M.J., Alvelos, H. (2025). Implementation of Quality 4.0 in Portuguese Companies: Diagnosis and Challenges for Future Development. Paper presented at the 31st IJCIEOM – International Joint Conference on Industrial Engineering and Operations Management, Polytechnic of Bari, Bari, Italy, 23-25 June 2025.

Oliveira, D., Rosa, M.J., Alvelos, H.: Getting an insight into Quality 4.0 reality: Design of a questionnaire. (2024). In: Proceedings of the International Conference on Quality Engineering and Management. pp. 748–762, Girona, Spain, 13–14 June

Sharma, M., and Joshi, S. (2020), "Digital supplier selection reinforcing supply chain quality management systems to enhance firm's performance". TQM Journal. https://doi.org/10.1108/TQM-07-2020-01

Sony, M., Antony, J., Douglas, J. A., and McDermott, O. (2021). "Motivations, barriers and readiness factors for Quality 4.0 implementation: an exploratory study". TQM Journal, 33(6), 1502–1515. https://doi.org/10.1108/TQM-11-2020-0272

Schumacher, A., Erol, S., and Sihn, W. (2016), "A Maturity Model for Assessing Industry 4.0 Readiness and Maturity of Manufacturing Enterprises". Procedia CIRP, Vol. 52, pp. 161–166. https://doi.org/10.1016/j.procir.2016.07.040

Sureshchandar, G. S. (2022). "Quality 4.0 – understanding the criticality of the dimensions using the analytic hierarchy process (AHP) technique". International Journal of Quality and Reliability Management, 39(6), 1336–1367. https://doi.org/10.1108/IJQRM-06-2021-0159

Wawak, S., Sütőová, A., Vykydal, D., and Halfarová, P. (2023), "Factors affecting Quality 4.0 implementation in Czech, Slovak and Polish organizations: Preliminary research". Advances in Production Engineering And Management, Vol. 18 No. 3, pp. 345–356. https://doi.org/10.14743/apem2023.3.477

Weckenmann, A., Bodi, Ş., Popescu, S., Dragomir, M., Hurgoiu, D., and Comes, R. (2019). "Hit or miss? Evaluating the potential of a research niche: A case study in the field of virtual quality management". Sustainability (Switzerland), Vol. 11 No. 5, pp. 1–26. https://doi.org/10.3390/su11051450

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

A Discussion on Quality Definitions for Artificial Intelligence Systems

André M. Carvalho [1]

[1] UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal; andremc@fct.unl.pt

***Abstract*— Artificial intelligence (AI) is rapidly being adopted across industrial, service and societal applications, yet the understanding of quality in AI systems remains fragmented and lacks a clear quality management perspective. While classical notions of quality may address the characteristics of AI-enabled systems, several shortcomings are found when trying to provide a comprehensive approach to AI Quality.**

This opinion paper argues that AI cannot be understood solely in terms of model performance or compliance with predefined requirements. Instead, AI quality should be understood as an emergent property of data-driven sociotechnical systems, reflecting the extent to which they remain fit for their intended purpose over time and within a given operational context.

Beyond this contextual framing, the paper further distinguishes two closely related quality challenges that deserve dedicated attention: data quality, as a prerequisite for trustworthy AI, and the societal impacts of AI-enabled decisions, which require multidisciplinary governance extending beyond quality management alone.

This paper contributes to the ongoing discussion on AI quality and hopes to provide a foundation for future research into context-specific quality models, assessment frameworks and management practices for AI-enabled.

Keywords – Artificial Intelligence, Quality Definitions;

1. Introduction

Quality has long been central to human activity, but its meaning has never been stationary. Juran defined quality as "fitness for use", focusing on the adequacy of a product for a particular purpose (Juran, 1988). Crosby, by contrast, defined quality as "conformance to requirements", emphasizing specification, control and defect prevention (Crosby, 1979). Deming approached quality through the performance of the system itself, linking it to process stability and continuous improvement (Deming, 1986). ISO 9001 later defined quality as the "degree to which a set of inherent characteristics of an object fulfils requirements", while reinforcing the importance of a process approach (ISO, 2015).

These formulations have established a shared understanding of quality in regards to use, conformance and system performance. However, such understanding becomes less obvious in the face of Artificial Intelligence (AI) systems. Unlike products, services or even "conventional" software, AI systems are data-dependent, probabilistic and adaptive. Their behavior depends not only on design and code, but also on training data, deployment conditions and evolving contexts of use. As a result, perspectives of quality as the compliance with fixed specifications or as the match to stable technical attributes is no longer offers a fitting match.

The problem is not merely that AI adds new quality criteria. More fundamentally, it unsettles the assumptions on which established definitions rest. This is visible in contemporary discussions of AI, where evaluation increasingly invokes notions such as robustness, fairness, accountability and trustworthiness. While these terms broaden the evaluative scope, they also blur conceptual boundaries. AI Quality is used to refer at once to these multiple dimensions, but often fails in clarifying or establishing the relationship between them. As a result, "Ai Quality" risks becoming both overloaded and under-specified.

Following this logic, the application context becomes fundamental to the understanding of quality in AI systems. Although certain quality characteristics, such as reliability, robustness or safety, remain broadly applicable, their interpretation and relative importance depend on the intended use of the system and the environment in which it operates. Consequently, quality cannot be defined solely in abstract terms but must be understood in relation to the specific objectives, stakeholders and constraints of each AI application. This reinforces the need to move beyond a generic notion of AI quality towards context-specific conceptualizations and assessment approaches.

Against this background, this opinion paper discusses **how adequate are classical definitions of quality in the context of Artificial Intelligence systems, what limitations they reveal, and which challenges remain for understanding quality in adaptive, data-driven AI systems.** The departure point is that classical definitions remain valuable in guiding the understanding of quality in AI-driven systems, but that they assume predefined requirements, bounded rules and relatively stable system behavior in ways that do not translate directly to AI systems.

2. Reframing Quality in AI

Classical quality management offers a set of highly influential definitions that continue to shape contemporary thinking. These definitions remain powerful because they capture different but complementary aspects of quality: fit-for-use, process-thinking, compliance and system performance. At the same time, they rest on assumptions that become challenging in the context of AI. "Fitness for use" presumes that use can be identified with reasonable clarity, yet not always AI systems operate in settings where purposes are clear; or in other cases, where the purpose either evolve or unintended (but useful) uses may emerge. "Conformance to requirements" assumes that requirements can be specified in advance, but above all, that they remain stable over time. In AI contexts, requirements are frequently evolving (either via scope-enlargement or scope-creep) incomplete, shifting or contested. Deming's focus on stability remains relevant, but AI systems are characterized by data drift, retraining and probabilistic outputs, making variation an intrinsic characteristic.

What this perceived clash reveals is not that classical definitions are no longer useful, but that their adequacy is partial. They were developed for contexts in which the object of quality was relatively well-defined and where assessment criteria were stable. AI systems, by contrast, are adaptive, data-dependent and embedded in broader sociotechnical arrangements. Their quality cannot be understood solely at the level of the model or technical artefact, since performance depends on data, context, organizational processes and human interaction.

The significance of classical definitions in this context is therefore double. On the one hand, they provide the conceptual starting point for thinking about quality in AI systems. On the other hand, their limitations expose the need for a broader conception of quality, one capable of accounting for adaptation, uncertainty and the distributed character of system performance. If classical definitions remain important but insufficient, the question becomes how quality should be understood in the context of AI. In AI systems, behavior emerges from the interaction of models, data, deployment conditions and individual or organizational practices; contexts where multiple (traditional) definitions of quality co-exist.

3. Configurations of Quality and their fitness for AI Systems

A further implication of this discussion is that the understanding of quality in AI systems is unlikely to converge into an single, uniform and existing definition of Quality. As AI becomes embedded in different organizational and service arrangements, it increasingly becomes subject to distinct modes of use and interaction.

Nevertheless, these different definitions - or configurations - offer important insights, pieces of a puzzle that aims to frame quality in AI systems. Departing from the understanding of Products quality (and thus looking at **AI as a product),** quality is defined primarily towards the technical aspect of an artefact, assessed in terms of performance, reliability and conformance against specified expectations (Sebastianelli and Tamimi, 2002). This framing remains useful but assumes relatively stable system boundaries and evaluation criteria. AI can also in some contexts be understood **as a service** and evaluated accordingly through dimensions of service quality such as tangibility, reliability or empathy (Ramya, Kowsalya, and Dharanipriya, 2019). This perspective captures delivery and use, but not the full systemic quality of the AI. A service may be highly available, easy to use, and responsive, yet still produce biased outputs, deteriorate under drift, rely on poor training data, or behave in ways that are misaligned with the decision context. Those are not just service delivery issues.

Human–machine interaction introduces a different but equally important shift. In many AI applications, system outputs do not act directly on the world but are mediated through human interpretation, response and intervention. Quality therefore depends not only on system precision or Quality of Service (QoS)/Quality of Experience (QoE), and it should also include the system's ability to support human–machine interaction, handle cognitive workload, and respond to the physical and mental needs of the users (Redzepagic et al., 2020; Banuelos-Lozoya et al., 2021). An AI system that performs well in formal terms may still exhibit poor quality if its outputs are systematically misunderstood, over-trusted or under-utilized. Similarly, the use of AI in organizational contexts brings challenges in the assessment of AI-supported decision-making and process management systems. In **AI in decision-support systems,** quality depends not only on technical adequacy, but on the consequences of system outputs within contexts where errors carry significant human, legal or social implications (Matta and Bolli, 2023). As for **AI embedded in organizational processes,** quality is shaped by the fit between the AI system and the workflow, decision structure and institutional routines within which it operates (Chalmers, et al., 2026). The relevant issue is not only whether the system performs well technically, but whether it supports and sustains effective organizational action.

A further development concerns continuously learning or **adaptive systems**. Here, quality cannot be fully assessed at the point of deployment because the system continuously absorbs new data and evolves over time, rendering static testing methods and offline validation datasets inadequate for ensuring reliability, fairness, and robustness (Grover, 2025). Quality becomes inseparable from monitoring, maintenance and institutional capacity for ongoing adjustment. Table 1 summarizes different Contextual Configurations of Quality and their link to AI.

Table 1. Contextual Configurations of Quality and their link to Artificial Intelligence.

Configuration	Primary focus of Quality	Main Quality dimensions	Why the classical definition becomes insufficient on its own
AI as a product	Technical artefact or model.	Performance, reliability, specification fulfilment.	Presumes bounded system and stable requirements.
AI as a service	Service arrangement combining model, infrastructure and provider relationship.	Availability, interoperability, update reliability, continuity of performance.	Captures delivery and use, but not the full systemic quality of the AI.
Human–machine interaction	Interaction between system output and human interpretation or intervention.	Usability, intelligibility, trust calibration, decision support adequacy.	Quality is co-produced in use rather than residing in the artefact alone.
AI in decision-support systems	Sociotechnical arrangement linking model outputs to consequential decisions.	Robustness, interpretability, accountability, contextual appropriateness.	Technical performance alone is insufficient for judging adequacy.
AI embedded in organizational processes	Coupling between AI capability and workflow or managerial routine.	Process fit, role allocation, escalation mechanisms, institutional alignment.	Quality depends on organizational integration, not only technical correctness.
Adaptive or continuously learning AI	Temporal behavior of the system across updates and changing data.	Drift management, sustained adequacy, monitoring and adjustment.	Quality cannot be fixed at deployment or assessed once and for all.

These different perspectives should not be understood as competing definitions of AI quality, but rather as complementary views of the same phenomenon. Each highlights a particular aspect of how quality is created, experienced or maintained, depending on the role that AI plays within a given context. Consequently, any comprehensive definition of quality in AI systems should acknowledge these multiple dimensions, recognizing that their relative importance will vary across applications, but that together they provide a more complete understanding of AI quality.

4. Asking the Machine: How AI Defines Quality

Before advancing towards a discussion on the meaning and scope of Quality in AI, a simple exercise was conducted. Its goal is not to provide a standing definition, but rather to appreciate the current "self-"understanding of Quality in AI systems, as a results of the training of different AI models. While this paper is more focused on industry or service application of AI, this exercise provides an important insight into how quality is being perceived by the leading developers of AI solutions.

Several large language models (LLMs) were asked to define the concept of quality in AI systems. Although these responses do not constitute scientific evidence, they provide an interesting reflection of the concepts and perspectives that emerge from the body of knowledge on which these models have been trained, including research literature, standards, and policy documents. Examining these definitions makes it possible to identify recurring themes, points of convergence, and differences in emphasis, providing a useful complement to the preceding review before proposing a consolidated definition of quality for AI systems.

Four widely used large language models (LLMs) were asked to provide a definition of quality in AI systems: ChatGPT, Claude, Microsoft Copilot and DeepSeek. All models received exactly the same prompt, with no additional context, examples or constraints, allowing the responses to reflect the models' general understanding of the concept. The objective was not to compare the performance of the models, but rather to identify recurring concepts and points of convergence in the way AI systems themselves synthesize and communicate the notion of quality. To ensure comparability, the prompt explicitly requested a concise, one-sentence definition.

Table 2. Definition of Quality in AI Systems as provided by four widely used large language models (LLMs): ChatGPT, Claude, Microsoft Copilot and DeepSeek

Model	Definition of Quality in AI Systems
ChatGPT	*Quality for AI systems is the degree to which an AI-enabled system consistently delivers accurate, reliable, trustworthy, and context-appropriate outcomes while meeting technical, human, ethical, and regulatory requirements throughout its lifecycle.*
Claude	*Quality in an AI system is the degree to which it reliably produces correct, safe, and useful outputs for its intended users and context.*
Copilot	*Quality in AI systems is the degree to which an AI reliably delivers accurate, safe, fair, and consistent outcomes that meet user and organizational expectations.*
DeepSeek	*Quality in AI systems is the measurable degree to which an AI model consistently delivers safe, reliable, fair, and accurate outcomes that meet stakeholder needs across its entire lifecycle.*
Prompt: "Give me a short, one-sentence definition of Quality in AI Systems."	

The four definitions above show a high degree of convergence. All define quality as the degree to which an AI system delivers desirable outcomes, consistently highlighting attributes such as accuracy, reliability and safety. At the same time, they incorporate broader concepts including fairness, trustworthiness, stakeholder needs and lifecycle considerations, reflecting many of the principles found in recent AI standards and governance frameworks.

Despite this convergence, the definitions remain necessarily generic. They provide a broad understanding of what quality in AI systems entails but offer little guidance on how quality should be interpreted or assessed in different application domains. This reinforces the argument that a single, universal definition is unlikely to be sufficient. Instead, the concept of quality in AI systems requires further refinement according to the type of AI system, its intended purpose and the context in which it operates.

5. Re-defining the scope

AI systems shift the conditions under which quality is defined and maintained. Quality can no longer be understood only as compliance with predefined requirements or as the possession of stable attributes. It must also account for evolving uses, changing contexts and the distributed interaction of technical and organizational elements.

In this sense, the problem is not only how to define quality in AI systems in general, but perhaps more how quality is constituted in particular AI settings. While we remain in the "early ages" of mass AI usage, understanding Quality in AI systems remains highly tied to the application context. Any work on quality in AI systems should thus be explicit about the configuration under consideration. In this sense, it is essential to clearly define the scope which we wish to address.

For quality Engineers and managers, the scop e of intervention should not be to define quality for AI in its broadest sense, nor to address the specific characteristics of general-purpose AI models or conversational interfaces. Instead, the discussion should focus on AI systems developed to support industrial, service and societal applications, where AI constitutes one component of a broader sociotechnical system designed to fulfil a defined purpose. In these settings, the quality of the AI system cannot be dissociated from the quality of the process, organization and context in which it operates. This does not mean that a definition of quality for AI in its broader sense should not be pursued, but rather admitting that such pursuit in is multidisciplinary challenges that requires not inputs for other sciences, but also a great maturity in the definition of AI quality within our community.

There are certain characteristics, such as reliability or trustworthiness, that remain broadly applicable. However, the relative importance of individual quality dimensions depends on the intended application and operational context. Consequently, the discussion promoted by this paper adopts a context-oriented perspective, where quality is understood as property of specific AI-enabled sociotechnical systems, rather than as an intrinsic property of the AI model alone. This provides a more appropriate basis for discussing quality in the context of engineering, operations, public services and other organizational environments where AI is increasingly being deployed.

As the discussion moves from general-purpose AI towards domain-specific applications, more concrete interpretations of AI quality begin to emerge. This reality can already be observed in recent standards and industrial guidance. ISO/IEC 42001 (International Organization for

Standardization, 2023a) frames AI as a managed organizational capability requiring governance, monitoring and continual improvement, while ISO/IEC 25059 extends traditional software quality models with characteristics specific to AI systems International Organization for Standardization, 2023b). Similarly, the NIST AI Risk Management Framework emphasizes properties such as validity, reliability, robustness and transparency throughout the AI lifecycle (Tabassi, 2023). Although developed for different purposes, these initiatives converge towards a common understanding that AI quality cannot be reduced to model performance alone, but must instead be addressed as a property of the broader sociotechnical system in which AI operates.

This is particularly evident in sectors where quality management is already well established, and AI is being integrated into existing engineering and operational practices. A representative example is the guidance published by the German automotive industry's Quality Management Center (Verband der Automobilindustrie, 2026), which addresses the application of artificial intelligence in automotive quality management. Rather than treating AI quality as an abstract concept, the guidance frames it in terms of deployment within established quality processes, emphasizing aspects such as data quality, validation, traceability, governance and continuous monitoring. Such initiatives reinforce the view that the quality of AI is increasingly being understood in relation to the quality of the sociotechnical systems in which it is embedded, rather than as a property of the AI model alone.

In more general terms, we should understand that when mentioning Quality in AI systems, we ought to look for definitions that will depend on the extent to which a data-driven sociotechnical system remains fit for its intended purposes over time and under a certain context, even if that context is marked by conditions of uncertainty and change. This framing preserves the core insight of classical quality theory, particularly the emphasis on adequacy and system performance, while extending it to contexts in which requirements are evolving, variation is intrinsic and quality must be maintained rather than simply verified. Such a reframing does not remove classical quality management. Rather, it shows that its foundational concepts remain useful only if they are interpreted in a more dynamic and systemic way. AI systems do not render quality irrelevant; they make its conceptual clarification more necessary.

6. Adjacent Challenges to AI Quality

Although this paper considers AI systems from a contextual perspective, viewing quality as an emergent property of a data-driven sociotechnical system, two closely related challenges deserve particular attention, while being context-independent. These correspond to the two ends of the AI lifecycle: the quality of the data that feeds the system and the societal impacts of the outcomes that it produces. Besides being somewhat independent from the context in which they occur, they also constitute research domains that extend beyond the scope of AI-specific applications. As such, they deserve dedicated consideration. Quality professionals have an important role to play in both areas, although their degree of leadership and responsibility is expected to differ.

The first challenge concerns data quality. The quality of data has been the subject of extensive research for several decades (Wang and Strong ,1996) and is widely recognized as a prerequisite for the successful deployment of data-driven systems (Kenett and Shmueli, 2014). Numerous frameworks have been proposed to characterize data quality through dimensions such as accuracy, completeness, consistency, timeliness and relevance (Carvalho et al., 2025), most of which developed for contextual applications (Montenegro et al., 2026). Nevertheless, the discussion remains highly relevant, particularly as organizations become increasingly dependent on data-driven decision-making. Although data quality has traditionally been led by the information systems and data management communities, quality professionals should play a much more prominent role (Carvalho et al., 2021). As the custodians of organizational processes and, in many cases, the stewards of the data generated by those processes, they possess both the knowledge and the methodologies required to ensure that data is fit for its intended use. Rather than acting as supporting actors, there is a strong argument that quality professionals should co-lead, or even lead, this discussion within organizations.

The second challenge concerns the quality of AI outcomes and their broader implications. Beyond technical performance, AI systems increasingly influence decisions that affect individuals, organizations and society, raising questions of fairness, transparency, accountability, safety and human oversight (Akinrinola et al., 2024; Cheong, 2024.). These issues extend beyond the boundaries of quality management and require contributions from disciplines such as computer science, law, ethics, public policy and the social sciences. Nevertheless, quality professionals have an important contribution to make. Their experience in defining requirements, managing risk, monitoring performance and promoting continual improvement provides a valuable foundation for ensuring that AI systems remain aligned with their intended purpose and societal expectations. Rather than leading this discussion alone, quality professionals should become active participants in the multidisciplinary effort required to govern the impacts of AI responsibly.

7. Discussion and conclusions

This paper shows that there is still no universally accepted definition of quality for AI systems, and the challenges ahead in pursuing one. Existing perspectives differ according to disciplinary background, application domain and intended purpose, with computer science, software engineering, quality management and AI governance each emphasizing different aspects of the concept. Nevertheless, a clear convergence can be observed. Across the literature, standards and industrial guidance, quality is increasingly understood as a multidimensional concept that extends beyond predictive performance to include data, governance, robustness, transparency, human interaction and lifecycle management. This evolution suggests that quality in AI systems should not be viewed as a fundamentally new concept, but rather as an extension of established quality principles to a new class of complex sociotechnical systems.

In less general terms, and from a Quality Science point of view, we understand that when mentioning Quality in AI systems, we should look for a definition that will depend on the **extent to which a data-driven system remains fit for its intended purposes over time and under a certain context, even if that context is marked by conditions of uncertainty and change**. This framing preserves the core insight of classical quality theory, particularly the emphasis on adequacy and system performance, while extending it to contexts in which requirements are evolving, variation is intrinsic and quality must be maintained rather than simply verified. Such a reframing shows that Quality's foundational concepts remain useful only if they are interpreted in a more dynamic and systemic way. AI systems do not render quality irrelevant; they make its conceptual clarification more necessary.

This works aims to hopefully offer a conceptual foundation for the discussion of a Quality-oriented approach for AI Systems, rather than a final or exhaustive formulation of such. Future work should refine and operationalize this concept for different classes of AI systems and application domains, recognizing that the relative importance of quality characteristics will inevitably vary across contexts. In particular, further research is needed to translate this conceptual understanding into assessment models, quality indicators and management practices that support the design, deployment and continuous improvement of AI systems in industrial, service and societal applications.

References

Akinrinola, O., Okoye, C.C., Ofodile, O.C. and Ugochukwu, C.E., 2024. Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. GSC Advanced Research and Reviews, 18(3), pp.050-058.

Banuelos-Lozoya, E., Gonzalez-Serna, G., Gonzalez-Franco, N., Fragoso-Diaz, O. and Castro-Sanchez, N., 2021. A systematic review for cognitive state-based QoE/UX evaluation. Sensors, 21(10), p.3439.

Carvalho, A.M., Sampaio, P., Rebentisch, E. and Saraiva, P., 2021. 35 years of excellence, and perspectives ahead for excellence 4.0. Total Quality Management & Business Excellence, 32(11-12), pp.1215-1248.

Carvalho, A.M., Soares, S., Montenegro, J. and Conceição, L., 2025. Data Quality: revisiting dimensions towards new framework development. Procedia Computer Science, 253, pp.247-256.

Chalmers, D., Hunt, R.R., Pachidi, S., Potočnik, K. and Townsend, D., 2026. The acceleration of artificial intelligence: Rethinking organization and work in an era of rapid technological change. Journal of Management Studies, 63(2), pp.285-314.

Cheong, B.C., 2024. Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. Frontiers in Human Dynamics, 6, p.1421273.

Crosby, P.B., 1979. Quality Is Free: The Art of Making Quality Certain. New York: McGraw-Hill.

Deming, W.E., 1986. Out of the Crisis. Cambridge, MA: MIT Press.

Grover, S., 2025. Engineering robust AI products through continuous quality assurance: a framework for testing, monitoring, and validation of adaptive live learning AI/ml systems in dynamic production environments. International Journal of Applied Mathematics, 38(2s), pp.1092-1113.

International Organization for Standardization, 2015. ISO 9000:2015 Quality Management Systems: Fundamentals and Vocabulary. Geneva: International Organization for Standardization

International Organization for Standardization, 2023a. ISO/IEC 42001:2023 Information technology - Artificial intelligence - Management system. Geneva: International Organization for Standardization

International Organization for Standardization, 2023b. ISO/IEC 25059:2023 Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems. Geneva: International Organization for Standardization

Juran, J.M., 1988. Juran on Planning for Quality. New York: Free Press.

Kenett, R.S. and Shmueli, G., 2014. On information quality. Journal of the Royal Statistical Society Series A: Statistics in Society, 177(1), pp.3-38.

Matta, S.S. and Bolli, M., 2023. Trustworthy ai: Explainability & fairness in large-scale decision systems. Review of Applied Science and Technology, 2(04), pp.54-93.

Montenegro, J., Conceição, L., Soares, S., Beernaerts, J., Hovestad, M. and Carvalho, A.M., 2026. A Data Quality Framework on Urban Vehicle Access Regulations. Transportation Research Procedia, 95, pp.129-136.

Ramya, N., Kowsalya, A. and Dharanipriya, K., 2019. Service quality and its dimensions. EPRA International Journal of Research & Development, 4(2), pp.38-41.

Redzepagic, A., Loffler, C., Feigl, T. and Mutschler, C., 2020, November. A sense of quality for augmented reality assisted process guidance. In 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (pp. 129-134). IEEE Computer Society.

Sebastianelli, R. and Tamimi, N., 2002. How product quality dimensions relate to defining quality. International Journal of Quality & Reliability Management, 19(4), pp.442-453.

Tabassi, E., 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Verband der Automobilindustrie, 2026. Artificial Intelligence in Quality Management. Berlin: Verband der Automobilindustrie (VDA).

Wang, R.Y. and Strong, D.M., 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, *12*(4), pp.5-33.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

A Framework for Quality Assessment in Service SMEs Based on a Process Weighting Model

Juan Canela [1], Esperanza Marcos [2], Valeria de Castro [3], Sofía Canela [4] and Jaime Dionisio [5]

1 Universidad Rey Juan Carlos, Madrid, España; juan.canela@urjc.es
2 Universidad Rey Juan Carlos, Madrid, España; esperanza.marcos@urjc.es
3 Universidad Rey Juan Carlos, Madrid, España; valeria.decastro@urjc.es

4 Universidad Rey Juan Carlos, Madrid, España; sofia.canela@urjc.es

5 Universidad Rey Juan Carlos, Madrid, España; jaime.dionisio@urjc.es

***Abstract*— The service sector is characterized by high competitiveness and constant change, making quality management a key factor in the sustainability of organizations. However, existing quality frameworks, such as EFQM, ISO 9001, ITIL, and CMMI, are designed primarily for large companies, which makes their application in small and medium-sized enterprises (SMEs) difficult due to their complexity and implementation costs. To address this limitation, LightSM³ is proposed, an evolution of the LightSME framework, which broadens its scope through a multidimensional approach to quality assessment in service management. The model integrates three dimensions: organizational characteristics, maturity levels, and company size. This article focuses on the development of the "size" dimension of the LightSM³ model, which, unlike traditional models that impose homogeneous requirements, adapts the evaluation criteria according to the size of the company, avoiding unnecessary burdens for SMEs. To this end, an evaluation methodology is proposed based on an adaptation of the Analytic Hierarchy Process (AHP), which uses pairwise comparisons and priority vectors to assign different weights to the evaluated processes depending on the company's size.**

Keywords – quality management, LightSME, Analytic Hierarchy Process (AHP), pairwise comparisons,

XXXII. Introduction

The service sector is characterized by high competitiveness and constant evolution, where quality is a key factor in ensuring the sustainability of service-providing organizations. To help companies address the challenges associated with innovation and the improvement of their services, processes, and practices, there are various quality frameworks, models, and standards. However, existing quality frameworks, including EFQM (2021), ISO 9001:2015, ISO/IEC 20000-1:2018, ITIL (Agutter, 2020), and CMMI (CMMI Institute, 2018), as well as VeriSM (Agutter et al., 2017), FitSM (2016), and P-CMM (Curtis et al., 2009), are typically designed with the needs and structures of large organizations in mind. These approaches require significant resources (financial, human, and knowledge-related, among others) that are often scarce or inaccessible to SMEs, making their adoption difficult in smaller organizations with limited resources.

LightSME (Feversani, 2023) is a lightweight framework for assessing quality in service organizations, specifically designed to be adopted by SMEs. It is based on the ISO/IEC 33000 family of standards (ISO/IEC 33000:2015) and integrates key features from the most representative service management models and standards, including EFQM (2021), VeriSM (Agutter et al., 2017), ISO/IEC 20000-1:2018, ITIL (Agutter, 2020), CMMI (CMMI Institute, 2018), and ISO 9001:2015. The framework was validated through its application in three companies of different sizes (micro, small, and medium-sized enterprises) and across different sectors, including market research, training, automotive repair, and the hospitality industry. A key insight derived from its application is that quality should not be assessed uniformly across organizations of different sizes; rather, company size is a critical factor in determining the processes required to achieve a given level of quality.

Current models require a certain level of process compliance to achieve a given level of maturity without considering the size of the organization being evaluated. However, the fact that large organizations need to implement specific processes to reach a certain maturity level does not imply that these same requirements should be imposed on micro or small enterprises. On the contrary, such requirements may be counterproductive, potentially overburdening SMEs with the management of processes that are unnecessary for their context.

LightSM³, a multidimensional model (cube) for the lightweight evaluation of service management, is proposed as an evolution of the LightSME framework. This new model incorporates a third dimension—the size of the organization—resulting in a structure based on three dimensions: (1) service organization characteristics (people, organizational governance, customers and consumers, and services); (2) maturity levels (immature, basic, improved, and excellent); and (3) organization size (micro, small, medium, and large), as shown in Fig. 1. Unlike LightSME, which only considered organizational characteristics and maturity levels, LightSM3 extends its approach to better adapt to the structural diversity of the sector. Therefore, it is a model specifically designed for service organizations, considering their operational particularities and the need for flexibility, enabling its application across different types of companies according to their size.

This paper focuses on the size dimension and proposes a method for evaluating processes in a differentiated manner according to organization size. To this end, an adaptation of the Analytic Hierarchy Process (AHP), proposed by Saaty and Vargas (2022), is used as a reference, focusing exclusively on the assignment of weights to processes without the need to select a specific alternative. These weights are then incorporated into the calculation of a weighted average, enabling the determination of the degree of completeness.

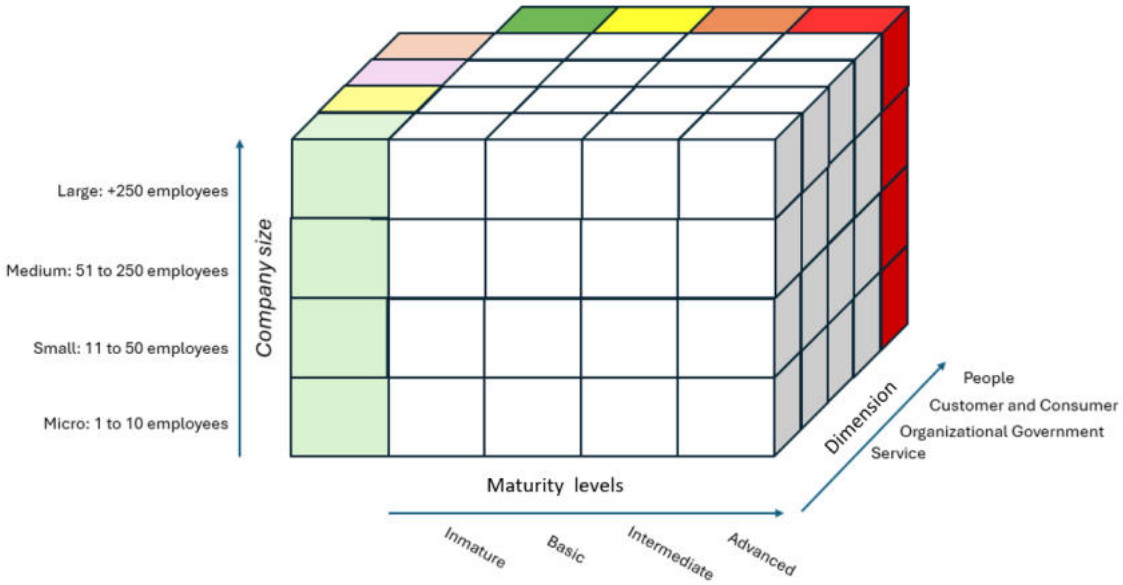


Fig. 6. LightSM^3 multidimensional model (cube) for the evaluation of service management, including organization size, maturity levels, and service characteristics.

The remainder of the paper is structured as follows: Section 2 reviews the main related work; Section 3 presents the LightSM^3 model, including the process reference model, the process assessment model, and the maturity model; finally, Section 4 summarizes the main conclusions and outlines future work.

XXXIII. RELATED WORK

In LightSME (Feversani, 2023), a comprehensive study of existing models for assessing quality in service organizations is presented. Fig. 2 provides a summary of this analysis.

The results of this study reveal several important limitations in current approaches when applied to small and medium-sized enterprises (SMEs). On the one hand, most existing models, such as EFQM (2021), ISO 9001:2015, ITIL (Agutter, 2020), and CMMI (CMMI Institute, 2018), have been primarily designed for large organizations, assuming the availability of significant financial, human, and technical resources. Consequently, their implementation in SMEs is often costly, complex, and difficult to sustain over time.

On the other hand, these approaches tend to adopt a "one-size-fits-all" perspective, defining homogeneous requirements regardless of the organizational context. This lack of adaptability may lead to inefficient implementations, as SMEs are forced to adopt processes and practices that are not aligned with their size, structure, or operational needs.

Furthermore, many of these models originate in domains such as software engineering or IT service management, which limits their applicability to general service sectors. As a result, organizations operating in areas such as education, consulting, or hospitality may encounter difficulties when transferring these approaches to their specific context.

These limitations highlight the absence of lightweight and adaptable frameworks specifically designed for SMEs, capable of adjusting to the characteristics and constraints of organizations according to their size and context. This gap motivates the need for new approaches that incorporate organizational size as a key factor in quality assessment.

The main contributions of this study can be summarized as follows: the framework is structured in accordance with the requirements established in the ISO/IEC 33000 standard (ISO/IEC 33000:2015), and it integrates key characteristics from international models and standards for quality and service management. This lightweight framework was developed with a focus on service-oriented SMEs, regardless of their sector, and is aimed at improving service quality through the assessment and enhancement of processes. In addition, the framework places particular emphasis on people within the organization and on customer service culture. From its application to several case studies, several lessons learned were identified, including the importance of considering company size, as well as the type of service provided, when measuring quality. Based on these findings, work has begun on an extended version, LightSM^3, which further adapts the framework to the specific needs of each organization by explicitly incorporating company size as a key factor.

Among the models that enable process evaluation, and with the objective of assigning weights to processes, several approaches have been identified, including direct assignment methods and comparison-based methods (Barba-Romero and Pomerol, 1997). Within direct assignment methods, simple ranking and simple rating can be distinguished. Simple ranking

requires the decision-maker to order the criteria from most to least important by assigning values from 1 to n (where n is the number of criteria), which are then normalized so that their sum equals one, yielding the corresponding weights. In contrast, simple rating involves assigning a value to each criterion according to a predefined scale (e.g., 0–5), after which the values are normalized to obtain the weights. Among comparison-based methods, the Revised Churchman–Ackoff Technique (RCAT) proposed by Knoll and Engelberg, and the Analytic Hierarchy Process (AHP) developed by Saaty and Vargas (2022), can be highlighted. The RCAT method is based on successive comparisons, where the decision-maker initially ranks the criteria, as in simple ranking, and then iteratively compares them in groups to adjust their weights. Finally, the AHP method consists of constructing a pairwise comparison matrix, in which the decision-maker compares criteria using a scale from 1 to 9. From this matrix, priority vectors are derived, and the consistency of the judgments is assessed.

	Characteristics											
Models and standards	**LCO**	**Per**	**Pro**	**CaS**	**CuS**	**Amb**	**Alc**	**MRP**	**MEP**	**MMO**	**Agi**	**Ori**
EFQM	✓	P	✓	X	P	ORG y SVC	✓	X	X	X	X	ORG
VeriSM	✓	✓	✓	✓	✓	SVC y SVC-TI	✓	X	X	X	✓	TDO
ISO 9001	P	✓	✓	✓	P	PRO y SVC	P	P	X	X	X	CPS
ISO/IEC 20000-1	P	✓	✓	✓	P	SVC-TI	✓	✓	X	X	X	TI
CMMI-SVC	X	X	✓	P	P	SVC-TI	✓	P	✓	✓	X	SW
ITIL	P	P	✓	P	X	SVC-TI	✓	✓	P	P	P	TI
FitSM	P	P	✓	P	X	SVC-TI	✓	✓	P	X	P	TI
ISO/IEC 33000	X	X	✓	P	X	EVPRO	✓	✓	✓	✓	X	SW
P-CMM	X	✓	✓	P	X	ORG	✓	P	✓	✓	X	SW
CENT/TS 16880	✓	✓	✓	✓	✓	SVC y SVC-TI	✓	X	X	X	X	EXS

Fig. 7. Comparative analysis of models and standards for quality assessment in service organizations, based on their main characteristics.

TABLE X. LEGEND

✓	Indicates full coverage of the characteristic	SVC	General Services
P	Indicates partial coverage of the characteristic	SVC-TI	IT Services
×	Does not consider the characteristic	ORG	Organizations
EVPRO	Process evaluation	PRO	Products
TDO	Digital transformation of organizations	SW	Software
CPS	Quality of products and services	TI	Information Technologies
EXS	Service excellence		

The AHP method is selected due to its robust structure and its ability to ensure consistency in the assignment of weights. As shown in Table 2, direct assignment methods are quick and easy to apply; however, their complete reliance on the subjective judgment of the decision-maker may lead to inconsistencies, especially when multiple criteria are evaluated. In contrast, although the AHP method is also based on subjective judgments, it incorporates mechanisms that provide greater rigor and consistency. It enables a more precise evaluation by relying on pairwise comparisons and includes a consistency verification mechanism, thereby reducing the likelihood of contradictions in the assignment of weights. Compared to the RCAT method, which requires successive adjustments and may become cumbersome when dealing with many criteria, AHP offers a more structured and mathematically grounded solution, ensuring that the assigned weights accurately reflect the relative importance of each criterion. Furthermore, AHP allows for the aggregation of judgments from multiple experts through techniques such as the geometric mean. In this context, it is common to consider a sample of approximately 20 participants as acceptable for this type of analysis. For these reasons, AHP is considered the most suitable method for this study, providing a balance between precision, reliability, and applicability in process evaluation.

TABLE XI. WEIGHT ASSIGNMENT METHODS

Method and Type	*Advantages*	*Disadvantages*
Simple Ranking (Direct Assignment)	Fast and easy to apply, without the need for complex calculations. Ideal for quick decision-making.	Subjective, depends on the decision-maker's judgment. May lead to inconsistencies when many criteria are involved.
Simple Rating (Direct Assignment)	Flexible and easy to apply with few criteria. Allows expressing differences in importance between criteria.	Does not ensure consistency between weights and may generate inconsistencies when many criteria are involved.
AHP (Analytic Hierarchy Process) (Pairwise Comparison)	Robust and structured, allows verification of the consistency of judgments.	Requires calculations and advanced tools. May become cumbersome when many criteria are involved.
RCAT (Revised Churchman-Ackoff Technique) (Successive Comparison)	Reduces inconsistencies in weight assignment and allows progressive adjustments through successive comparisons.	Tedious when many criteria are involved and does not guarantee mathematical validation.

XXXIV. MULTIDIMENSIONAL MODEL

A. Model Framework

LightSM^3 is focused on service management in general and promotes values such as lightness, flexibility, adaptability, and scalability when applied within organizations. It follows the same structure as the LightSME framework (Feversani, 2023), which is composed of three models: the Process Reference Model (**PRM**), the Process Assessment Model (**PAM**), and the Maturity Model (**MM**), while introducing an extension to the Process Assessment Model.

The Process Reference Model is structured into four characteristics, which are further divided into key factors that group related processes, as shown in Table 3. The Process Assessment Model incorporates evaluation indicators that enable the assessment of process performance and capability, that is, to understand and evaluate the extent to which processes meet the requirements defined in the Process Reference Model.

TABLE XII. KEY FACTORS OF THE LIGHTSM^3 PROCESS REFERENCE MODEL (PRM)

Organizational Governance	**People**
• Strategy • Portfolio • Service Culture • Compliance • Data and Information Management	• Administration and Human Resources • Training • Communication and Team Management
Customers and Consumers	**Services**
• Communication Channels • Needs and Expectations • Customer Management and Loyalty	• Design and Innovation • Service Management • Service Delivery • Incident and Request Management

The LightSME-PAM (Feversani, 2023) is taken as a reference. To determine the capability level of processes, it is necessary to evaluate the performance of the attributes or tasks that constitute each process. After evaluating these tasks, the degree of fulfillment of each factor is determined using the method presented in the following section, which is adjusted according to the size of the organization, allowing for greater adaptability. The Maturity Model defines four maturity levels (Immature, Basic, Improved, and Excellent). The maturity of an organization is determined by the maturity of its processes. The quality level of these processes establishes the different maturity levels that an organization can achieve. Therefore, to determine the maturity level of an organization, it is necessary to assess the capability level of each process.

B. Weight Assignment Method

This section presents a method for evaluating processes in a differentiated manner according to the size of each organization. To this end, different weights are assigned to the processes within each key factor, which will influence the weighted average used to calculate the degree of completeness of each key factor.

Multi-criteria decision-making (MCDM) aims to evaluate and select the best option among alternatives by considering multiple criteria. Some of these criteria have greater relevance depending on factors such as the decision-maker's preferences and are therefore assigned weights or importance coefficients that reflect their relative significance.

As previously mentioned, there are various methods for weight assignment (Barba-Romero and Pomerol, 1997). In this study, the Analytic Hierarchy Process (AHP), proposed by Saaty in the late 1970s (Saaty and Vargas, 2022), is selected. AHP is a weight assignment method based on the calculation of the dominant eigenvector of a pairwise comparison matrix of criteria. AHP is typically used to assign weights to different criteria to select among multiple alternatives (Saaty and Vargas, 2022). In this work, however, the core of this method is applied to assign different weights to processes depending on the size of the organization—micro (1 to 10 employees), small (11 to 50 employees), medium (51 to 250 employees), or large (more than 250 employees)—without the need to subsequently select a specific alternative.

A weight vector $w = [w_1, w_2, \ldots, w_n]$ must be assigned to the processes within each key factor, such that $\sum_j w_j = 1$. This vector varies according to the size of the organization. The same procedure is applied to each characteristic and each key factor.

As an example, as shown in Table 4, within the "People" characteristic, the processes PM1, PM2, and PM3 are associated with the "Administration and Human Resources" factor. The objective is to assign a weight vector w = [w_1, w_2, w_3], which varies depending on the size of the organization.

TABLE XIII. PROCESSES ASSOCIATED WITH THE "ADMINISTRATION AND HUMAN RESOURCES" FACTOR IN THE LIGHTSM³ MODEL

Administration and Human Resources	**PM1:** Manage employee onboarding, compensation, and offboarding
	PM2: Manage productivity and performance
	PM3: Manage talent acquisition

In order to assign the weight vector, the following steps must be followed (Barba-Romero and Pomerol, 1997; Mendoza et al., 2019).

- **Representation of judgments and scale**. In order to compare processes, it is first necessary to establish a value scale. Pairwise comparison judgments in the Analytic Hierarchy Process (AHP) are applied to pairs of homogeneous elements. The fundamental scale used to represent the intensity of judgments is shown in Table 5. This scale has been validated, among others, against 28 alternative scales that were tested, demonstrating its effectiveness not only in numerous applications by different users, but also through a theoretical justification of the appropriate scale for comparing homogeneous elements.

TABLE XIV. TABLE TYPE STYLES

a_{ij}	**When comparing criterion i with criterion j is:**
1	Equally important
3	Slightly more important
5	Moderately more important
7	Strongly more important
9	Extremely more important
2, 4, 6, 8	Intermediate values for compromise judgments

- **Construction of the comparison matrix and normalization**. Once the scale has been established, all the necessary elements are available to begin the pairwise comparisons of the processes. Next, experts compare each process i with each process j, obtaining values a_{ij} that represent the relative importance between processes. For example, if $a_{ij} = 5$, it indicates that process i is strongly more important than process j. These values a_{ij} are then arranged into a square matrix of order n, known as the pairwise comparison matrix A = [a_{ij}].

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

Following the previous example, we obtain the following matrix provided by an expert based on the defined scale, for a small-sized organization.

$$A = \begin{bmatrix} 1 & 4 & 5 \\ 1/4 & 1 & 3 \\ 1/5 & 1/3 & 1 \end{bmatrix}$$

- **Calculation of the weight vector and consistency**. For the judgments provided by experts to be coherent, it is expected that the comparisons established between processes are consistent. This means that if a process i is considered, for example, twice as important as process j, and process j is three times more important than process k, then process i should be six times more important than process k. This property, known as consistency, is desirable; however, the AHP method does not strictly require it, as it reflects the reality that human judgments may contain slight inconsistencies or contradictions. If the judgments were perfectly consistent, the weights could be directly derived; however, in practice, an indicator called the Consistency Ratio is calculated to measure the overall degree of coherence. If this value is sufficiently low (less than 10%), the results are considered acceptable.

  Thus, based on the afore mentioned matrix, and with calculations detailed in the *appendix*, a Consistency Index is calculated:

$$CI = \frac{\delta_{max} - n}{n-1} \quad (1)$$

  where δ_{max} is the dominant eigenvalue, a value obtained by mathematically processing the comparison matrix, which reflects the degree of consistency among the judgments provided. If all comparisons were perfectly consistent, this value would be exactly equal to n, the total number of processes compared.

  We can then determine the so-called Consistency Ratio:

$$CR = \frac{CI}{RCI} \quad (2)$$

where RCI is the Random Consistency Index, whose values are given in Table 6.

TABLE XV. R.C.I.

n	2	3	4	5	6	7	8	9
RCI	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45

If CR < 10%, the consistency is considered acceptable. In summary, the complete AHP method consists of the following three steps:

1. The pairwise comparison matrix A is obtained from the decision-maker.
2. The weight vector w and its Consistency Ratio (CR) are calculated.
3. If CR < 10%, the weight vector w is accepted. Otherwise, the decision-maker is asked to revise some or all the a_{ij} values.

In the example presented, using an R script (see Appendix), the dominant eigenvalue is $\delta_{max} = 3.085767$, and the weight vector for processes PM1, PM2, and PM3 is [0.6738106, 0.2255355, 0.1006539]. The Consistency Index is CI = 0.04288335, which leads to a Consistency Ratio of CR = 0.0739368 < 10%.

Therefore, the weight vector is used to compute the weighted average of the degree of completeness of the processes within each key factor and organization size, as detailed in the Appendix.

C. Applicability and Future Validation

The proposed model offers high applicability, as it enables the assessment of an organization's quality status through a structured evaluation process based on questionnaires and surveys adapted to the size of the organization. Its main outcome is a diagnosis of the achieved quality level, supporting decision-making and process improvement without requiring significant changes to existing operational structures.

Regarding its validation, an empirical evaluation process is currently underway. The approach is being supported by a panel of experts with experience in service management and quality assessment, including both academic researchers and professionals from private organizations. At the current stage, a set of expert evaluations has already been collected, with the objective of expanding the sample to ensure sufficient representation across different organization sizes (micro, small, medium, and large).

To support this process, a software tool has been developed to automatically compute process weights and consistency metrics based on incoming expert judgments. Additionally, a complementary tool enables organizations to input their evaluation data and obtain assessment results, facilitating the practical application of the model in real contexts.

The model is currently being applied in a real organizational environment within a company department, providing an initial empirical basis. This ongoing validation will allow assessing the robustness and adaptability of the model, as well as its capability to provide differentiated quality assessments according to organization size.

XXXV. CONCLUSIONS

This study has addressed the problem of quality assessment in service organizations, highlighting the lack of models that are easy to apply and adaptable to the size of the organization. It has been shown that existing quality frameworks, such as EFQM (2021), ISO 9001:2015, ITIL (Agutter, 2020), and CMMI (CMMI Institute, 2018), are primarily designed for large organizations, which makes their application in smaller structures difficult due to their complexity and the resources required for implementation.

As a solution, LightSM3 has been proposed as a multidimensional model that introduces an adaptive evaluation based on three fundamental dimensions: organizational characteristics (people, organizational governance, customers and consumers, and services), maturity level (immature, basic, improved, and excellent), and organization size (micro, small, medium, and large).

The main contribution of this work is the differentiated assignment of weights to processes according to the size of the organization, using an adaptation of the Analytic Hierarchy Process (AHP). This approach enables a more flexible and realistic evaluation, avoiding the imposition of homogeneous requirements that may be ineffective or even counterproductive for SMEs.

The results suggest that quality in service organizations should not be measured uniformly across companies of different sizes. The adaptation of criteria and the use of weighting methodologies provide a more flexible and context-aware approach to quality assessment across diverse organizational contexts. However, these findings are based on an illustrative application and should be considered preliminary, as the empirical validation of the model is currently ongoing.

As future work, several directions are proposed: validating the LightSM^3 model in a larger number of organizations across different sectors; incorporating an additional dimension related to the sector in which the organization operates; exploring the integration of machine learning techniques to optimize weight assignment; and developing software tools to facilitate the practical application of the model.

With this approach, LightSM³ is positioned as a viable alternative for improving quality management in service organizations, enabling sustainable growth and adaptation to the specific needs of each organization.

ACKNOWLEDGMENT

This work has been partially funded by the Spanish State Research Agency through the SerDigital project (PID2020-117244RB-I00), by the Community of Madrid through the DESAFÍO-CM project (TEC-2024/COM-235), and by the Services Science and Engineering Network 2022 (RED2022-134148-T), funded by the Spanish Ministry of Science and Innovation.

REFERENCES

Agutter, C. (2020), ITIL 4 Essentials: Your essential guide for the ITIL 4 Foundation exam and beyond, IT Governance Publishing.

Agutter, C., van Hove, S., Steinberg, R. and England, R. (2017), VeriSM: A service management approach for the digital age, Van Haren Publishing, Zaltbommel.

Barba-Romero, S. and Pomerol, J.-C. (1997), Decisiones multicriterio, Universidad de Deusto, Bilbao.

CMMI Institute (2018), CMMI Model Version 2.0, ISACA.

Curtis, B., Hefley, B. and Miller, S. (2009), People Capability Maturity Model (P-CMM) Version 2.0, Second Edition, SEI, Carnegie Mellon University.

EFQM (2021), The EFQM Model: Revised 2nd Edition, The European Foundation for Quality Management.

Feversani, D.P. (2023), LightSME: Marco ligero para la evaluación de la gestión de servicios enfocado a pequeñas y medianas empresas, PhD Thesis, Universidad Rey Juan Carlos, España.

FitSM (2016), FitSM-0: Overview and vocabulary, The FitSM Standard Family: Standard for lightweight IT service management, version 2.4, available at: https://www.fitsm.eu (accessed 31 March 2026).

ISO (2015), ISO 9001:2015 – Quality management systems – Requirements, International Organization for Standardization, Geneva, Switzerland.

ISO/IEC (2015), ISO/IEC 33000:2015 – Information Technology – Process Assessment, International Organization for Standardization, Geneva, Switzerland.

ISO/IEC (2018), ISO/IEC 20000-1:2018 – Information technology – Service management – Part 1: Service management system requirements, International Organization for Standardization, Geneva, Switzerland, available at: https://www.iso.org/standard/70636.html (accessed 31 March 2026).

Mendoza, A., Solano, C., Palencia, D. and García, D. (2019), "Aplicación del proceso de jerarquía analítica (AHP) para la toma de decisión con juicios de expertos", Ingeniare. Revista chilena de ingeniería, Vol. 27 No. 3, pp. 348–360.

Saaty, T.L. and Vargas, L.G. (2022), The Analytic Hierarchy Process, 2nd ed., Springer, New York.

UNE-CEN (2015), UNE-CEN/TS 16880:2015 Service excellence – Creating outstanding customer experiences through service excellence.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Professional Training in Lean Six Sigma: An analysis of Yellow Belt and Green Belt Programs

Mónica Puertas[1], Edgar Yalta [2], José Carlos Flores-Molina[3]

[1] Pontificia Universidad Católica del Perú; mpuerta@pucp.edu.pe;
[2] Pontificia Universidad Católica del Perú; edyaltag@pucp.edu.pe;
[3] Pontificia Universidad Católica del Perú; jflores@pucp.edu.pe;

***Abstract*— This study presents an analysis of two Lean Six Sigma training programmes offered by a continuing education institution in Peru: Lean Six Sigma Yellow Belt (LSSYB) programme, aimed at university students and recent graduates, and Lean Six Sigma Green Belt (LSSGB) programme, aimed at practising professionals. Using data collected between 2016 and 2025, statistical analyses (Kruskal-Wallis and Mann-Whitney) and segmentation techniques (k-means clustering) were applied to identify the factors influencing academic performance and to characterise the participants' profiles. In addition, the final projects from the LSSGB programme were analysed to identify the sectors and processes most frequently addressed in improvement initiatives. The results reveal differences between the two programmes in terms of the factors affecting performance, as well as distinct profiles based on gender, age, work experience, roles and specialization. This study aims to provide empirical evidence on the effectiveness and adaptability of Lean Six Sigma training in educational and professional contexts; it also offers recommendations for the design of inclusive training programmes aligned with the principles of continuous improvement and organisational excellence.**

Keywords – Lean Six Sigma; Professional Training; Higher Education; Continuous Improvement; Quality Education

XXXVI. INTRODUCTION

The continued pursuit of operational excellence in organisations has driven the adoption of continuous improvement methodologies. Lean Six Sigma (LSS) has become established as a prominent methodology for advancing quality management and process optimisation by combining Lean principles (focused on waste elimination) with Six Sigma techniques (statistical rigour to reduce variability), thereby achieving significant improvements in cost, quality and customer satisfaction (Antony et al., 2018; Kwak & Anbari, 2006; George, 2002; Munro et al., 2022; Snee, 2010). Since its emergence, the application of LSS has not been confined to the manufacturing industry; rather, it has expanded into other sectors (Makwana & Patange, 2021). This expansion reflects the adaptability of LSS as a quality improvement strategy across different organisational contexts.

With regard to professional training in LSS, it is structured into belt levels (typically Yellow Belt, Green Belt and Black Belt) that represent progressively more advanced competencies and roles in improvement projects (Pyzdek & Keller, 2018). Each level entails expectations regarding a professional's ability to apply LSS within an improvement project. For instance, a Green Belt professional should be prepared to lead medium-scope improvement projects using the DMAIC cycle (Define, Measure, Analyse, Improve and Control) (Munro et al., 2022), whereas a Black Belt professional undertakes projects of high complexity (Antony et al., 2017; Mcshane-Vaughn, 2022; Pyzdek & Keller, 2018). Overall, LSS training programmes aim to ensure that participants develop robust competencies in process analysis, data-driven decision-making and leadership (Antony et al., 2018; DeRuntz & Meier, 2009).

Despite growing interest in LSS training, both in industrial and educational settings, gaps remain in the literature. From a general perspective, prior LSS research has focused on cases of improvement project implementation in companies or on organisational impact—i.e., industrial contexts (Buestan et al., 2025). In the context of professional training, most studies have described the perceptions of certified professionals (Green, 2006), explored the critical success factors for implementing LSS in higher education or other educational and service institutions (Antony et al., 2017; Kuwaiti, 2019; Nadeau, 2017), examined the standardisation of content in advanced training courses (Louzada et al., 2023), or characterised participant profiles and analysed performance within a Black Belt certification programme (Flores-Molina et al., 2025). Against this background, there is a clear need to document the characteristics and outcomes associated with LSS programmes aimed at Yellow Belt and Green Belt audiences, who are typically university students or recent graduates and practising professionals, respectively.

This study seeks to provide empirical evidence on the effectiveness and adaptability of LSS training in educational and professional contexts through the analysis of two training programmes in Peru. This setting is of particular interest due to the coexistence of organisations at different levels of organisational maturity (e.g., small and medium-sized enterprises (SMEs), informal firms, digital transformation initiatives), wide sectoral diversity, and the growing demand for professional certifications (INEI, 2022; OECD, 2025). Specifically, the study examines a Lean Six Sigma Yellow Belt (LSSYB) programme, aimed at final-year university students and recent graduates, and a Lean Six Sigma Green Belt (LSSGB) programme, aimed at practising professionals. The LSSYB programme provides foundational knowledge, whereas the LSSGB programme offers more in-depth training and requires the completion of an improvement project. Delivered by a Peruvian continuing education institution, these training programmes represent two articulated levels within LSS education.

Analysing participants between 2017 and 2025 makes it possible to identify factors that affect participants' academic performance and to characterise typical profiles using segmentation techniques. In addition, the practical application of LSS is assessed by collecting information on the improvement projects developed as part of the LSSGB programme. The findings are intended to inform recommendations for the design of LSS training programmes aligned with the principles of continuous improvement and organisational excellence, and tailored to participants' characteristics.

Accordingly, three objectives are established:

(1) To determine which participant profile factors significantly influence academic performance in each training programme (Lean Six Sigma Yellow Belt programme and Lean Six Sigma Green Belt programme). Specifically, the study evaluates the effect of variables such as modality, gender, age, speciality and—depending on the programme—employment status or role on final marks (academic performance).

(2) To characterise typical participant profiles in both training programmes through cluster analysis, in order to identify distinctive demographic segments.

(3) To explore trends in the practical application of Lean Six Sigma through the improvement projects undertaken within the Lean Six Sigma Green Belt programme, identifying the sectors in which it is most frequently applied and the types of processes most commonly addressed.

XXXVII. METHODOLOGY

A. Design and variables

The study was conducted using a case study design (Case & Light, 2011). The approach is quantitative with a descriptive scope. A Peruvian continuing education institution, recognised for its training programmes in quality management, was selected in order to analyse in detail two of its LSS offerings: the LSSYB programme and the LSSGB programme. Although both programmes are delivered by the same institution, participants come from a wide variety of educational institutions nationwide. For example, LSSYB participants come from 49 higher education institutions and 13 regions of Peru (Lima, Callao, Arequipa, Ica, Piura, Ancash, La Libertad, Ayacucho, Cusco, Junín, Apurímac, Lambayeque and Puno), while LSSGB participants come from 53 higher education institutions and 14 regions (Lima, Arequipa, Callao, La Libertad, Piura, Lambayeque, Cajamarca, Junín, Ancash, Ica, Cerro de Pasco, Apurímac, Tacna and Moquegua). This indicates that the study is not limited to participants from a single organisation or institution; rather, it reflects a multi-institutional and multi-regional context for the adoption of LSS in Peru.

The dependent variable was defined as each participant's final grade (academic performance), measured on a 20-point scale (0–20), with 11 as the minimum passing score. The independent variables including in both training programmes are as follows:

a) *Modality*: a dichotomous variable (face-to-face and online), where each mode differs in terms of content delivery, interaction with the instructor and the assessment system.

b) *Gender*: a variable recording the participant's self-declared gender (male and female), based on the information provided in the institution's enrolment form.

c) *Age Group*: a categorical variable divided into three groups that vary by training programme. For LSSYB, the categories were 19–23 years, 24–26 years and 27 years and above; for LSSGB, the categories were 21–29 years, 30–39 years and 40 years and above. This categorisation allows for the analysis of performance differences across age groups, taking into account each programme's target audience.

d) *Speciality*: a categorical variable classifying participant as either engineering (industrial engineering, civil engineering, software engineering, among others) or non-engineering (disciplines such as economics, social sciences, business administration, among others). This variable is included because the institution predominantly serves an engineering-oriented student population. However, increasing demand has been observed among participants from non-engineering backgrounds.

e) *Type of institution*: a categorical variable classifying the participant's higher education institution of origin as either public or private. This distinction is important given the substantial structural and pedagogical differences between public and private higher education in Peru, which influence students' prior academic preparation. Such disparities may affect the development of analytical and quantitative competencies required for successful learning in LSS.

Because the training programmes target different audiences, programme-specific variables were also included, as follows:

a) *Employment status*: a binary variable indicating whether an LSSYB programme participant was employed at the time of the programme (yes/no). This variable allows the examination of the potential impact of studying and working simultaneously during LSS training.

b) *Role*: a three-category variable capturing the organisational hierarchy level of LSSGB programme participants (senior management and heads of department, supervisors, and technical or operational staff). This classification reflects the typical organisational structure of Peruvian organisations and enables analysis of whether professionals' level of responsibility, often linked to experience and acquired competencies, may influence academic performance.

c) *Project implementation sector*: variable that takes into account the sector in which the LSSGB programme student developed their final project (e. g. mining, food, health, etc.).

d) *Type of project improvement*: variable that takes into account the type of improvement addressed by the student on the LSSGB programme (*time reduction*, *reduction of nonconformities*, or *waste elimination*). Projects focused on *waste elimination* are associated with 'wastage', an inevitable loss that occurs during the process, rather than a non-conformity. In contrast, projects focused on *nonconformity reduction* are associated with the failure to meet an established requirement, without necessarily constituting waste. Lastly, projects focused on *time reduction* are associated with the optimisation of time in processes.

B. Training programmes

The Lean Six Sigma Yellow Belt (LSSYB) programme is a 36-hour introductory programme designed to familiarize participants with the fundamental principles of Lean and Six Sigma. It is primarily targeted at final-year university students and recent graduates, most commonly from engineering or management programmes, although it is open to any field of study. The curriculum covers the Lean philosophy, the DMAIC methodology and basic quality tools. Because it is an entry-level programme, a final improvement project is not required; therefore, assessment is based on coursework and intermediate deliverables.

In contrast, the Lean Six Sigma Green Belt (LSSGB) programme is a more extensive 120-hour programme that provides in-depth training in statistical and Lean techniques. It is aimed at practising professionals seeking to certify their competencies at the Green Belt level. As a final requirement, participants must complete a team-based improvement project applied to a real organisational process, making use of the tools addressed throughout the programme. These Projects are carried out in teams of four to six participants with guidance from a supervisor.

Each training programme comprises a set of modules (LSSYB: 3 modules, LSSGB: 8 modules). Each module is assessed in a different way depending on the assessment method (e.g., exam, group assignment, final project, etc.). The final score (academic performance) is calculated as the average of the scores for each module. Both training programmes were offered in face-to-face modality (classroom-based sessions) and online modality (live synchronous sessions via a digital platform). In recent years, however, there has been a shift towards a greater proportion of online cohorts, driven by increased demand (particularly in the case of LSSGB, where geographical flexibility is valuable for working professionals) and the impact of the COVID-19 pandemic.

C. Sample and variables

A historical dataset of participants and their academic results was compiled for cohorts delivered between 2016 and 2025. For the LSSGB programme, data were gathered for the 2017–2025 editions, whereas for the LSSYB programme, data were gathered for the 2018–2025 editions—periods in which the available information was complete and comparable across cohorts. The final sample comprises 450 LSSGB programme participants and 388 LSSYB programme participants. The frequency distribution of participants' scores (academic performance) is shown in **Figure I**.

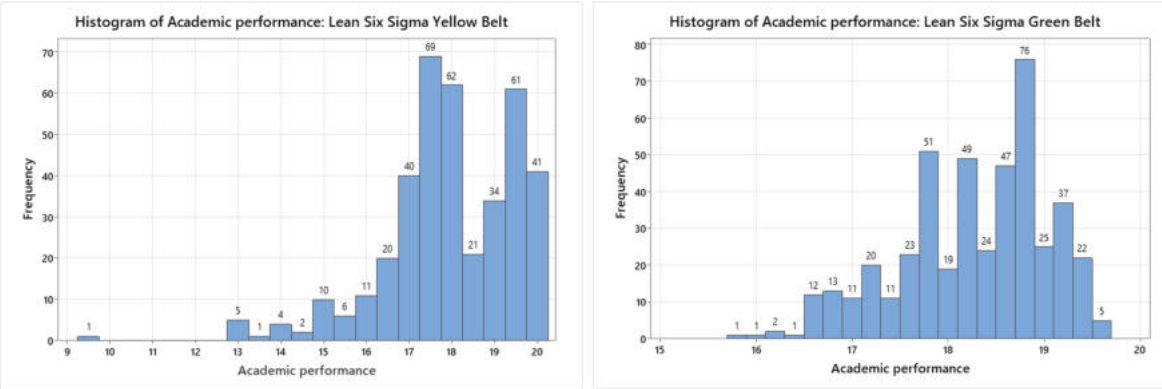


Fig. 1. Histogram: frequency distribution of the dependent variable (academic performance).

Table I presents the demographic data for the Lean Six Sigma Yellow Belt programme. This table provides a clear characterisation of the university-based profile of participants in the LSSYB programme. The data show a relatively balanced distribution across age groups, alongside a marked predominance of students or graduates from engineering programmes (84.54%) and higher participation of men than women. A further notable feature is the prevalence of the online modality, together with a larger proportion of participants who were not employed at the time of the programme (57.73%). It is also important to highlight that most participants were either studying at, or had graduated from, private higher education institutions (62.10%).

Table II presents the demographic data for the Lean Six Sigma Green Belt (LSSGB)programme. In line with the programme´s professional orientation, ages are concentrated in early and mid-career stages, with a higher proportion of male participants, and a greater representation of individuals occupying intermediate organisational roles. The predominance of the online modality suggest that training is integrated with participant's work routines, while the high proportion of engineers (83.75%) reflects a specialisation profile similar to that observed in the LSSYB programme, albeit with a more consolidated focus on professional practice. Unlike the LSSYB programme, the distribution by type of higher education institution reported by participants is balanced.

Tables I and **II** also present an exploratory analysis of the dependent variable (academic performance), specifically the following descriptive statistics for each variable: mean, median and standard deviation. This made it possible to identify patterns of potential differences between groups and to guide subsequent inferential analyses.

Table I. Demographic data and exploratory analysis for the Lean Six Sigma Yellow Belt (2018–2025)[a]

Variable	Category	Frequency	Percentage	Academic performance		
				Mean	*Median*	*Standard deviation*
Modality	Face-to-face	155	39.95%	17.60	18.0	1.51
	Online	233	60.05%	18.21	18.0	1.52
Gender	Male	221	57.00%	18.08	18.0	1.67
	Female	167	43.00%	17.88	18.0	1.35
Age group	19-23 years	143	36.86%	18.02	18.0	1.35
	24-26 years	156	40.20%	17.84	18.0	1.55
	27 years and older	89	22.94%	18.10	18.0	1.79
Speciality	Engineering	328	84.54%	17.97	18.0	1.51
	Non-engineering	60	15.46%	17.93	18.0	1.73
Employment status	Employed	164	42.27%	18.20	18.3	1.58
	Unemployed	224	57.73%	17.80	17.7	1.49
Type of institution	Public	147	37.90%	18.10	18.0	1.40
	Private	241	62.10%	17.88	18.0	1.62

[a]The sample consisted of 388 participants

Table II. Demographic data and exploratory analysis for the Lean Six Sigma Green Belt (2017–2025)[a]

Variable	Category	Frequency	Percentage	Academic performance		
				Mean	*Median*	*Standard deviation*
Modality	Face-to-face	121	26.89%	18.13	18.2	0.80
	Online	329	73.11%	18.33	18.4	0.75
Gender	Male	293	65.11%	18.29	18.4	0.76
	Female	157	34.89%	18.24	18.2	0.78
Age group	21-29 years	206	45.78%	18.39	18.6	0.77
	30-39 years	169	37.55%	18.22	18.2	0.73
	40 years and older	75	16.67%	18.11	18.3	0.80
Speciality	Engineering	376	83.60%	18.28	18.4	0.77
	Non-engineering	74	16.40%	18.26	18.3	0.75
Role	Managers and leaders	78	17.33%	18.37	18.4	0.64
	Supervisors	272	60.45%	18.25	18.3	0.77
	Technicians and operators	100	22.22%	18.25	18.4	0.84
Type of institution	Public	223	49.56%	18.57	18.6	0.64
	Private	237	50.44%	18.98	18.0	0.77

[a]The sample consisted of 450 participants

In summary, Tables I and II confirm expected differences between the two programmes. For example, the LSSYB audience consists largely of young participants, many of whom are not employed at the time of enrolment, whereas all those enrolled in LSSGB are practising professionals holding roles with varying levels of responsibility. The tables also show a notable level of female participation in both programmes, suggesting a growing presence of women professionals pursuing Lean Six Sigma certification.

D. Data analysis

To achieve Objective 1, inferential statistical techniques were employed. For each programme, the dependent variable was the participant's final mark (quantitative scale from 0 to 20). The independent variables for the LSSYB programme included modality (face-to-face vs online), gender (male vs female), age group (varies by training programme), speciality (engineering vs non-engineering), and employment status (employed vs not employed); For the LSSGB programme, the same set of variables were considered with employment status replaced by role within the organisation (technical/operational staff, supervisors, and managers/leaders). Since the data did not follow a normal distribution (Lean Six Sigma Yellow Belt: Kolmogorov-Smirnov, $p < 0.010$; Lean Six Sigma Green Belt: Kolmogorov-Smirnov, $p < 0.010$), non-parametric tests were applied in each case to determine whether there were statistically significant differences in final marks between the groups. A significance level of $p = 0.05$ was adopted. For variables with two categories, the Mann-Whitney U test was used, and for variables with three or more categories, the Kruskal-Wallis H test was used.

To achieve Objective 2, a k-means cluster analysis was performed to identify groups of participants with similar characteristics (Hair et al., 2019). Separate analyses were conducted for the population of each programme. The segmentation variables consisted of key demographic characteristics: gender, age, speciality and employment status for LSSYB; and gender, age, speciality and role for LSSGB. As required by the k-means algorithm, these categorical variables were numerically encoded. In both cases, a three-cluster (k = 3) solution was selected, as it produced clearly interpretable clusters of adequate size, with no cluster representing less than 15% of the sample. Mean profiles were calculated for each cluster, and χ^2 tests confirmed that all variables significantly differentiated the clusters ($p < 0.001$ for all variables), demonstrating the statistical validity of the segmentation. The resulting clusters therefore represent typical participant profiles within the programmes.

To achieve Objective 3, a descriptive analysis of the final projects from the LSSGB programme was conducted. Data were compiled for 44 group projects developed between 2018 and 2025, typically corresponding to one project per group of four to six participants. The variables 'Project implementation sector' and 'Type of project improvement' were used. Absolute and relative frequencies were calculated for each category to identify the most frequently addressed sectors and improvement types. As this analysis was exploratory and illustrative in nature, no hypothesis testing was performed. Its purpose was to illustrate practical trends in the application of Lean Six Sigma tools, complementing the assessment of academic performance with an applied dimension.

All statistical calculations were performed using Minitab 22. Results are presented separately for each programme (LSSYB followed by LSSGB), combining quantitative reporting with interpretative discussion, in line with the study objectives.

XXXVIII. RESULTS

A. Six Sigma Yellow Belt

The results for the Lean Six Sigma Yellow Belt programme are presented below. First, to address Objective 1, non-parametric tests were conducted. The results are shown in the following table:

Table III. Results of Non-parametric tests: Lean Six Sigma Yellow Belt

Variable	Non-parametric tests		p-value
	Mann-Whitney U test	Kruskal-Wallis H test	
Modality	13855.00		< 0.001
Gender	19246.50		0.466
Age group		4.196	0.123
Speciality	9720.50		0.880
Employment status	15099.00		0.003
Type of institution	16563.00		0.281

For the LSSYB programme, significant differences were observed in mean final marks between participants enrolled in face-to-face and online modalities ($p < 0.001$), as well as between those who were employed and those who were not employed during the programme ($p = 0.003$). In contrast, no statistically significant differences in academic performance were identified with respect to gender, age groups, speciality or type of institution. In summary, while academic performance was broadly comparable across participants in this programme, both the instructional modality (online or face-to-face) and the presence of concurrent employment were both associated with differences in their academic outcomes.

To better understand these significant differences in the LSSYB programme, mean comparisons were conducted. **Figure 2** presents the results of comparisons across the categories of the modality and employment status variables:

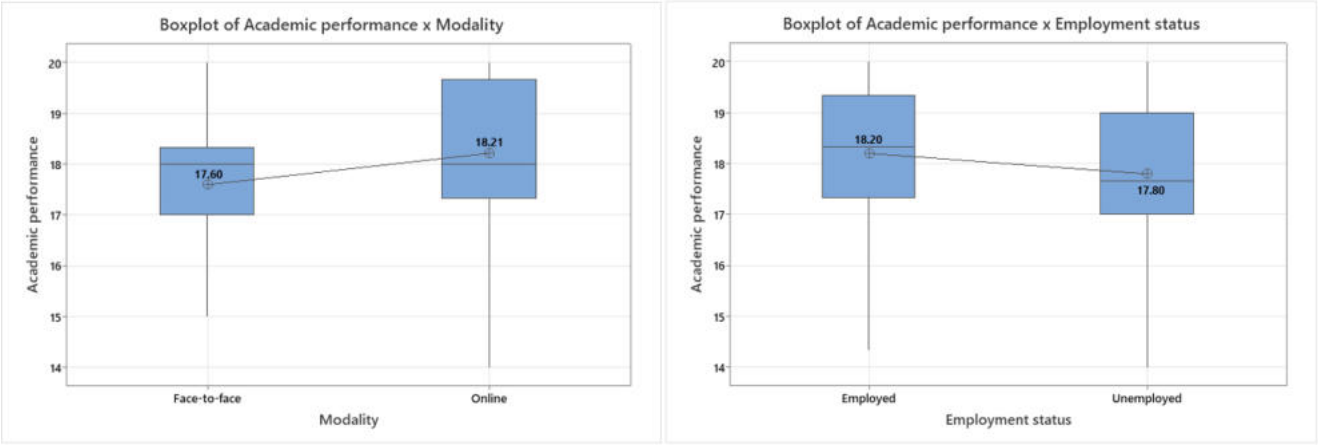


Fig. 2. Boxplot of Academic performance x modality ($p < 0.001$) and academic performance x employment status ($p = 0.003$).

With respect to modality, participants enrolled in the online modality achieved higher marks on average (M = 18.21, SD = 0.099) than participants in the face-to-face modality (M = 17.60, SD = 0.120), with the difference being statistically significant ($p < 0.001$). This finding suggests a performance advantage associated with synchronous online delivery at the Yellow Belt level, possibly due to the greater flexibility afforded by the online modality. In addition, participants who were employed during the programme obtained slightly higher marks on average (M = 18.20, SD = 0.120) compared with those who were not employed (M = 17.80, SD = 0.140), and this difference was also statistically significant ($p < 0.001$). This modest gap may indicate that concurrent workplace experience (i.e., exposure to practical improvement contexts) positively influences learning during the programme.

To address Objective 2, a k-means cluster analysis was conducted. Grouping the 388 LSSYB programme participants into three clusters resulted in clearly differentiated demographic profiles. The characteristics of the identified clusters (Clusters 1, 2 and 3) across the respective variables are shown in the following figure:

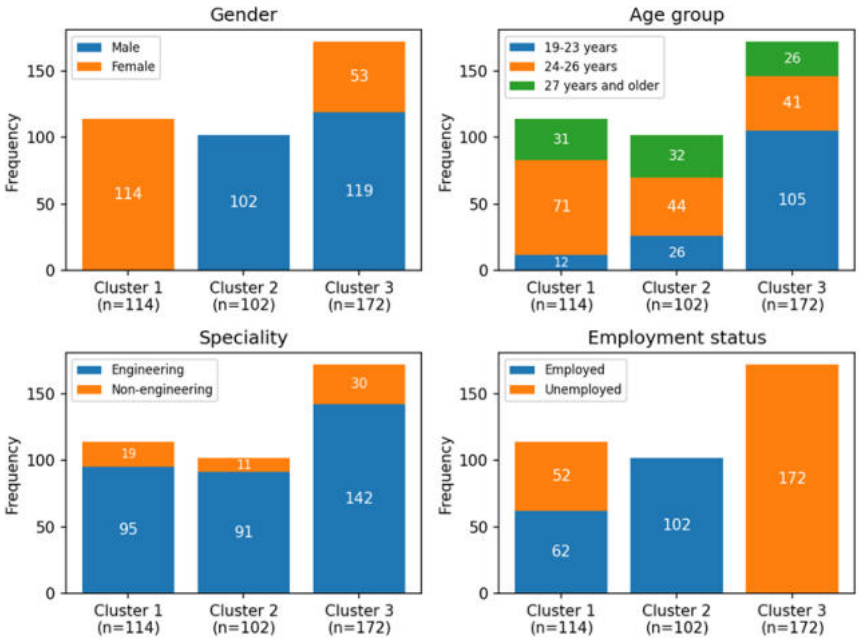


Fig. 3. Analysis of demographic profiles (k-means clusters) of the LSSYB. The distribution in tables was statistically significant (chi-squared test: $p < 0.001$).

Cluster 1 is characterised by young female engineers aged 24–26 who are moderately integrated into the labour market. This cluster (n = 114) is composed exclusively of women, almost all of whom have an engineering background. The predominant age falls within the second age group (24–26 years). Approximately half were employed at the time of the programme, while the remainder were recent graduates or final-year students without employment. This profile represents young women who are either entering the labour market or seeking certification before joining an organisation. From a gender perspective, this cluster highlights the growing interest and participation of women in Lean Six Sigma training.

Cluster 2 represents young, employed male engineers. This group (n = 102) is composed entirely of men, the majority of whom are engineering graduates (n = 91). All members of this cluster were employed while taking the programme. The age distribution is similar to that of Cluster 1, as it is concentrated in the 24–26 age group. This segment corresponds to early-carrier engineers who enrol in LSS training either to apply Six Sigma within their organisations or to strengthen skills relevant to professional advancement.

Cluster 3 represents, predominantly, very young male participants with no work experience. This is the largest cluster (n = 172) and, unlike Clusters 1 and 2, it presents a mixed gender composition, although men predominate and it is distinguished by the absence of employment among its members. They were therefore likely students with no work experience or recent graduates who had not yet entered the labour market. Moreover, this cluster concentrates the youngest participants (19–23 years). With respect to educational background, although most participants come from engineering programmes, this cluster also includes the highest proportion of participants from non-engineering disciplines (n = 30). Overall, Cluster 3 reflects the traditional Yellow Belt participant profile: university students or recent graduates without experience who seek Lean Six Sigma certification to complement their academic training and enhance future employability.

These findings confirm the presence of distinct participant subgroups within the population, ranging from young female engineering graduates (Cluster 1) to inexperience students and recent graduates (Cluster 3). Such heterogeneity suggests the need of differentiated pedagogical approaches. For example, participants in Cluster 3 may benefit from greater practical contextualisation due to their limited work experience, whereas Cluster 2 (employed participants) may contribute real-world examples but also face constraints related to time availability. This finding underscores the importance of segmenting instructional strategies within an apparently homogeneous programme to maximise learning outcomes for each group.

B. Lean Six Sigma Green Belt

The results for the Lean Six Sigma Green Belt (LSSGB) programme are presented below. To address Objective 1, non-parametric tests were conducted. The results are shown in **table IV**.

Non-parametric tests indicated a statistically significant difference between participants undertaking the programme online versus face-to-face (p = 0.021), as well as between age groups (p = 0.006). In contrast, no statistically significant differences in final marks were observed with respect to gender, speciality, organisational role, or participants' institution type of origin.

Table IV. Results of Non-parametric tests: Lean Six Sigma Green Belt

Variable	Non-parametric tests		p-value
	Mann-Whitney U test	Kruskal-Wallis H test	
Modality	17076.50		0.021
Gender	22036.50		0.443
Age group		10.253	0.006
Speciality	13516.00		0.595
Role		0.793	0.673
Type of institution	24804.50		0.713

To interpret these differences, mean comparisons were conducted. **Figure 4** presents the results of comparisons across categories of the modality and age group variables.

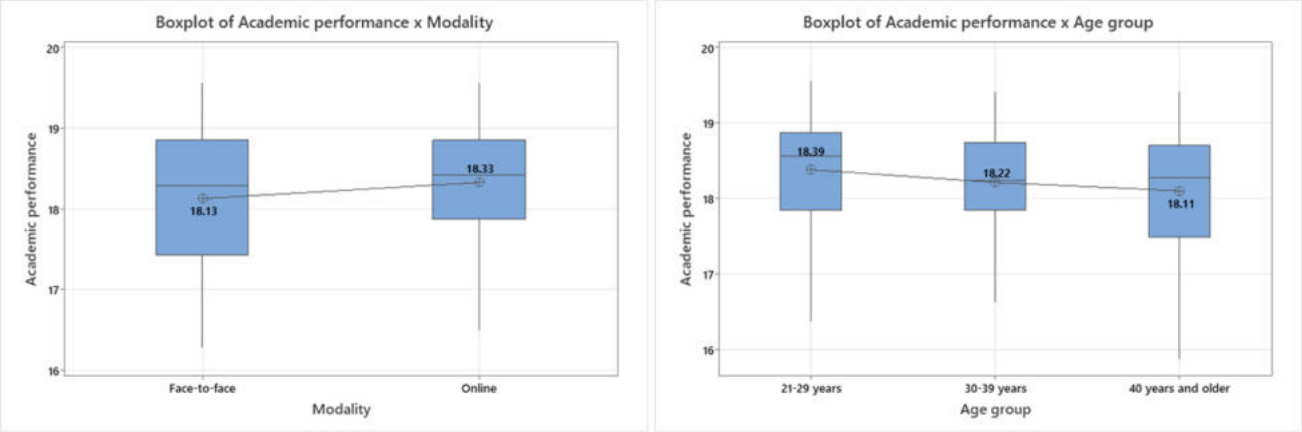


Fig. 4. Boxplot of Academic performance x modality (p = 0.021) and academic performance x age group (p = 0.006).

As observed for the LSSYB programme, participants in the online modality achieved higher performance on average (M = 18.33, SD = 0.751) than those attending the face-to-face sessions (M = 18.13, SD = 0.807), with a statistically significant difference (p = 0.021). This finding may be attributed to the affordances of online training for working professionals, including geographical and scheduling flexibility, which can facilitate programme engagement alongside professional responsibilities. Regarding age group, participants in the youngest group (21–29 years) achieved the highest mean final mark (M = 18.39, SD = 0.799), followed by those in the second age group (30–39 years) (M = 18.22, SD = 0.732) and the third age group (40 years and above) (M = 18.11, SD = 0.793). The difference between age groups were statistically significant difference (p = 0.006), suggesting a modest decline in performance with an increasing age, potentially reflecting differing time constraints or learning contexts among more senior professionals.

To address Objective 2, a k-means cluster analysis was conducted. Since k-means requires 'k' to be set a priori, alternative solutions (e.g., k = 2, k = 3 and k = 4) were evaluated using fit criteria and substantive criteria (interpretability and minimum cluster size). K = 3 was selected as it offered an appropriate balance between group separation, interpretative stability and sufficient cluster sizes. These clusters revealed demographic segments primarily differentiated by combinations of age, gender, speciality and professional role. The resulting profiles (Clusters 1, 2 and 3) and their distribution across the relevant variables, are shown in **Figure 5**.

Cluster 1 represents young technical professionals with a relatively balanced gender composition. This group (n = 106) is characterised by a strong concentration of participants in the youngest age group (21–29 years) and a more gender diversity than the other clusters. Members primarily hold technical or operational roles in their organisations. Another distinctive feature is that participants from non-engineering fields (n = 58) outnumber those from engineering backgrounds. In summary, this profile corresponds to early-career professionals with limited experience who perceive the LSSGB programme as a means to enhance their effectiveness in technical roles.

Cluster 2 represents mid-career supervisors and constitutes the largest group (n = 251), representing the typical participant profile in this program. It comprises professionals predominantly from engineering backgrounds (n = 251), with representation of both young adults (21–29 years) and mid-career adults (30–39 years). The gender distribution is predominantly men (n = 160). Most members (n = 201) hold intermediate or supervisory positions within their organisations, involving responsibility for processes and teams, though without occupying senior executive roles.

Cluster 3 represents experienced senior managers. This group (n = 93) includes the oldest participants (40 years and above), a substantial majority of men (n = 70), and a predominance of engineering backgrounds (n = 77). Roles tend to be managerial and head-of-department positions, including area directors, plant managers, and similar roles. This profile corresponds to senior professionals who pursue Green Belt training to formalise experiential knowledge and to lead continuous improvement initiatives within their organisations. The presence of this cluster highlighted that LSS attracts not only technical and middle-level professionals, but also organisational leaders interested in advancing operational excellence.

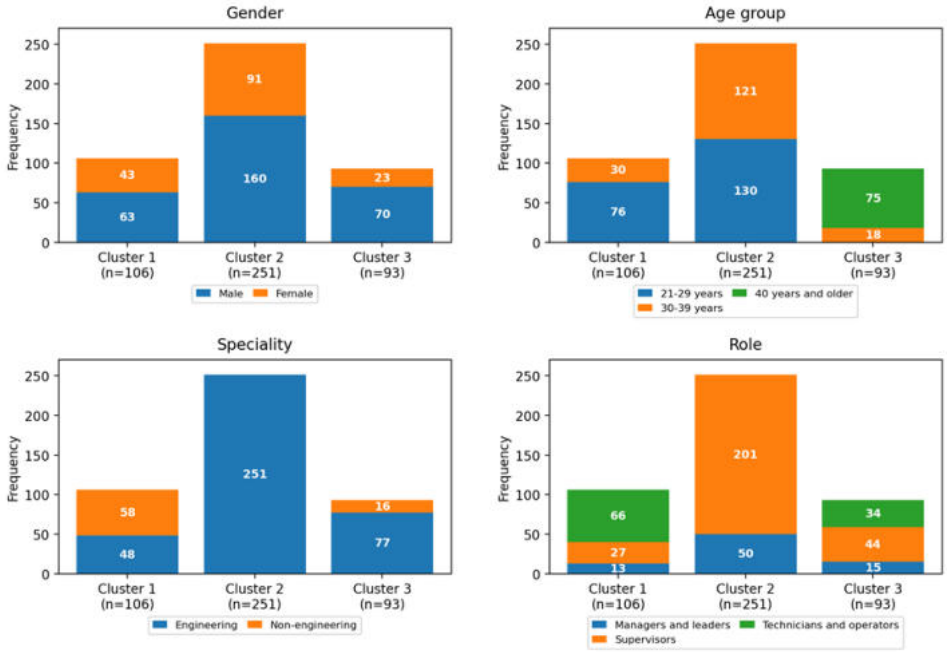


Fig. 5. Analysis of demographic profiles (k-means clusters) of the LSSYB. The distribution in tables was statistically significant (chi-squared test: $p < 0.001$).

To address Objective 3, a descriptive frequency analysis of the final projects developed in the LSSGB programme was conducted. This analysis enabled the identification of patterns related to the organisational sector in which projects were implemented and the types of improvement objectives pursued. **Figure 6** summarises the most frequently represented organisational sectors and the most frequent improvement foci. As shown in **Figure 6**, 40% of projects were developed in organisations operating within the industrial and manufacturing sector, followed by the food sector, which accounted for 31% of the projects. It is also notable that 11% of projects were carried out in mining organisations. This sectorial distribution aligns with Peru's main economic activities, which are strongly oriented towards extractive and processing industries. Smaller proportion of projects were observed in the services, finance and trade sectors (8%), healthcare (6%) and construction (4%). Regarding improvement focus, **Figure 6** shows that the most frequent objective—accounting for more than 50% of projects—was the reduction of nonconformities (53%). The second most frequent objective was waste elimination (26%). Finally, 21% of the projects focused on time reduction. These results suggest a strong emphasis on the systematic improvement of quality and the enhancement of operational efficiency, which are fundamental principles of the Lean Six Sigma methodology.

Overall, project teams sought to implement their projects successfully through measurable and quantifiable performance improvements, which were formally presented and justified during the programme's final assessment. This high level of practical application indicates were able to translate theoretical knowledge into concrete organisational outcomes, providing a key indicator of the effectiveness and success of a training programme.

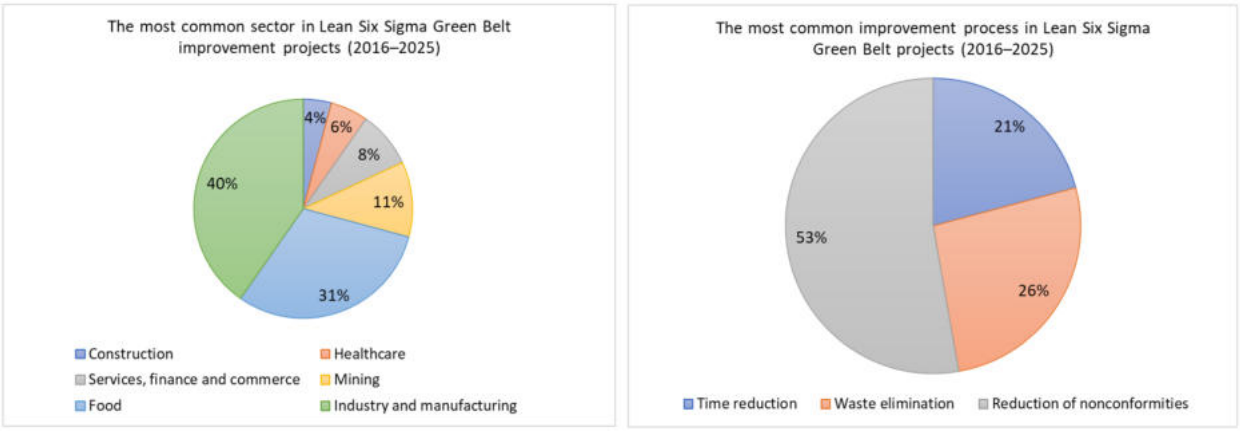


Fig. 6. The most common sector in Lean Six Sigma Green Belt improvement projects (2016–2025) and the most common improvement process in Lean Six Sigma Green Belt projects (2016–2025).

XXXIX. DISCUSSIONS AND IMPLICATIONS

The findings confirm that the effectiveness of LSS training depends not only on the technical content delivered, but also on the training context and the participant's demographic profile. These effects vary by certification level. In both the LSSYB and LSSGB programmes, training modality showed a significant association with academic performance, consistent with previous studies highlighting the importance of flexibility and learner autonomy in continuous improvement programmes designed for adults and early-career professionals (Antony et al., 2017; Nadeau, 2017; Flores-Molina et al., 2025). The study also indicates that other relevant factors differ across training levels. Employment status is a significant variable at the Yellow Belt level, whereas age emerges as a key factor at Green Belt level. This suggests that the influence of individual characteristics is mediated by the level of practical complexity associated with each training programme (McShane-Vaughn, 2022; Munro et al., 2022; Paneerselvam et al., 2025). Moreover, the cluster analysis shows that the participant population is not structurally uniform, reinforcing the need for more adaptive and differentiated approaches (Hollingshed, 2022; Hutwelker & Ott, 2022). Finally, with respect to LSSGB projects, the findings support the argument that evaluating the effective transfer of learning into professional practice (via a final improvement project) is essential for assessing the maturity and practical value of LSS programmes (Antony et al., 2018; Pyzdek & Keller, 2018).

Several practical implications may be considered:

- *Pedagogical differentiation based on profiles within the same certification level*: identifying differences in participant´s educational background, professional experience and academic performance enables the design of teaching strategies and learning supports tailored to each learner profile. This approach enhances learning effectiveness by reducing performance gaps and ensuring instructional methods aligned with empirical evidence. For example, differentiated interactive online materials could be developed for participants who need to reinforce foundational concepts, alongside alternative materials for those with higher prior knowledge or experience.
- *Aligning professionals' work experience with the practical demands of each training programme*: it is advisable to adjust the complexity and applicability of practical exercises and projects according to participants' level of experience, which may enhance the transfer of learning to the workplace and increase the practical value of the certification in that environment. For less experienced professionals, simulated or case-based projects may be more appropriate; by contrast, more experienced professionals can be encouraged to develop projects directly within their organisational contexts.
- *Strategic segmentation of LSS training into differentiated learning pathways*: characterising participant profiles by sector, professional role, speciality and employment status allows the design of specific learning trajectories within LSS training programmes. Such segmentation can improve programme and, consequently, enhance participants' academic performance. For instance, common or core sessions could be interleaved with parallel workshops tailored to specific roles (managers, supervisors, analysts, interns/trainees, among others).
- *Comprehensive evaluation of training impact*: expanding evaluation criteria to include real-world application, implemented improvements and competency development provides a more complete assessment of training effectiveness, facilitating continuous program improvement. For example, in addition to progress-based assessments and participation metrics, rubrics could be introduced to evaluate not only technical proficiency in applying the DMAIC methodology, but also transversal competencies such as critical thinking, leadership, effective communication and teamwork.
- *Prioritising sectors and processes for improvement projects*: Systematic analysis of project characteristics can help to identify sectors and processes with the greatest demand for improvement. This can support training provision aligned with market needs and the creation of strategic partnerships with organisations. For example, representatives from leading companies could be invited to present real organisational challenges, guiding participants toward relevant improvement problems, or to form sector-specific project teams to analyse and addressed targeted issues.

The results derived from the Peruvian case provide evidence from a context characterised by high sectoral diversity and varying levels of organisational maturity—conditions that are typical of emerging economies but remain underrepresented in the Lean Six Sigma literature. Unlike studies focusing on large, highly standardised organisations, this research shows how LSS training can be effectively implemented and can yield consistent outcomes in an environment where extractive industries, manufacturing, service sector, and organisations with heterogeneous resources coexist (OECD, 2025). While the findings are not intended to be generalised, the study demonstrates the adaptability of LSS principles and certification schemes to contexts with diverse organisational structures. This reinforces the relevance of LSS beyond highly industrialised economies (Makwana & Patange, 2021). In this sense, the Peruvian case provides a useful reference point for educational institutions and organisations in similar settings aiming to implement continuous improvement programmes that are closely aligned with their operational realities.

In conclusion, the findings indicate that academic performance is influenced by different factors depending on programme level: modality emerges as a common factor, while employment status and age group are programme-specific factors (LSSYB and LSSGB, respectively). The cluster analysis further reveals the presence of differentiated participant profiles within each programme, while project analysis evidenced effective transfer of learning towards improvements in real organisational processes. Collectively, these findings reinforce the importance of designing Lean Six Sigma training programmes that explicitly consider participant diversity and the that deliberately integrate learning activities with organisational practice.

One of the limitations identified in the research is the potential bias in the method of evaluation, given that the sample comes from a single institution. Another limitation is the lack of comprehensive information on the projects, as only their titles are recorded, which may lead to a bias in the assignment of the 'type of project improvement' categories.

This study highlights the need to move beyond standardised Lean Six Sigma training towards more adaptive and practice-oriented models. Specifically, training programmes should be deliberately designed to account for the diversity of participant profiles, including differences in experience, professional roles, and learning contexts, while ensuring a strong integration between instructional content and real organisational application, thereby enhancing both academic outcomes and practical impact on operational excellence.

REFERENCES

Antony, J., Snee, R. and Hoerl, R. (2017), "Lean Six Sigma: yesterday, today and tomorrow", International Journal of Quality & Reliability Management, Vol. 34 No. 7, pp. 1073–1093, doi: 10.1108/IJQRM-03-2016-0035.

Antony, J., Gupta, S., Sunder M., V. and Gijo, E.V. (2018), "Ten commandments of Lean Six Sigma: a practitioners' perspective", International Journal of Productivity and Performance Management, Vol. 67 No. 6, pp. 1033–1044, doi: 10.1108/IJPPM-07-2017-0170.

Buestan, M., Perez, C.C. and Rodríguez-Zurita, D. (2025), "Lean six sigma for health care: multiple case studies in Latin America", International Journal of Lean Six Sigma, Vol. 16 No. 1, pp. 172–196, doi: 10.1108/IJLSS-10-2023-0169.

Case, J.M. and Light, G. (2011), "Emerging Research Methodologies in Engineering Education Research", Journal of Engineering Education, Vol. 100 No. 1, pp. 186–210, doi: 10.1002/j.2168-9830.2011.tb00008.x.

DeRunt, B. and Meier, R. (2009), "Trainers' perceptions of the relative importance of the ten topics included in the American society for quality's Six Sigma Black Belt certification", Journal of Industrial Technology, Vol. 25 No. 3, pp. 1–13.

Flores-Molina, J.C., A. Puertas, M. and Yalta, E. (2025), "Lean Six Sigma Black Belt training for professionals: analysis of profiles, projects and academic performance", Proceedings of the International Conference on Industrial Engineering and Operations Management, in 10th North American International Conference on Industrial Engineering and Operations Management, IEOM Society International, Orlando, USA, pp. 886-895, doi: 10.46254/NA10.20250256.

George, M.L. (2002), Lean Six Sigma: Combining Six Sigma Quality with Lean Speed, McGraw-Hill, New York.

Green, F.B. (2006), "Six Sigma and the green belt perspective: a study in five companies", International Journal of Six Sigma and Competitive Advantage, Vol. 2 No. 3, p. 291, doi: 10.1504/IJSSCA.2006.011119.

Hair, J.F. (2019), Multivariate Data Analysis, Eighth edition., Cengage, Andover, Hampshire.

Hollingshed, M. (2022), "Standardizing Six Sigma Green Belt training: identification of the most frequently used measure phase DMAIC tools", *International Journal of Lean Six Sigma*, Vol. 13 No. 2, pp. 276–294, doi: 10.1108/IJLSS-12-2020-0220.

Hutwelker, R. and Ott, H. (2022), "A Lean Six Sigma Certification MOOC: Balancing Standard Material and Individual Guidance to Help Participants Implement Real Projects", in Guralnick, D., Auer, M.E. and Poce, A. (Eds.), *Innovations in Learning and Technology for the Workplace and Higher Education*, Vol. 349, Springer International Publishing, Cham, pp. 122–132, doi: 10.1007/978-3-030-90677-1_12.

INEI. (2022), "Perú: estructura empresarial", available at: https://www.gob.pe/institucion/inei/informes-publicaciones/6330149-peru-estructura-empresarial-2022 (accessed 6 March 2026).

Kwak, Y.H. and Anbari, F.T. (2006), "Benefits, obstacles, and future of six sigma approach", Technovation, Vol. 26 No. 5–6, pp. 708–715, doi: 10.1016/j.technovation.2004.10.003.

Kuwaiti, A.A. (2020), "Use of Six Sigma and Lean Six Sigma Methodologies in Higher Education Institutions: A Systematic Review", Lean Six Sigma for Higher Education, WORLD SCIENTIFIC (EUROPE), pp. 1–20, doi: 10.1142/9781786348500_0001.

Louzada, P.D.S., Sigahi, T.F.A.C., Moraes, G.H.S.M.D., Rampasso, I.S., Anholon, R., Antony, J. and Cudney, E.A. (2023), "Critical analysis of Lean Six Sigma black belt certification courses offered in Brazil", The TQM Journal, Vol. 35 No. 7, pp. 1980–2002, doi: 10.1108/TQM-08-2022-0254.

Makwana, A.D. and Patange, G.S. (2021), "A methodical literature review on application of Lean & Six Sigma in various industries", Australian Journal of Mechanical Engineering, Vol. 19 No. 1, pp. 107–121, doi: 10.1080/14484846.2019.1585225.

McShane-Vaughn, M. (2023), The ASQ Certified Six Sigma Black Belt Handbook, ASQ Quality Press, La Vergne.

Munro, R.A., Ramu, G. and Zrymiak, D.J. (Eds.). (2022), The ASQ Certified Six Sigma Green Belt Handbook, Third Edition, Third., ASQExcellence, Milwaukee.

Nadeau, S. (2017), "Lean, Six Sigma and Lean Six Sigma in Higher Education: A Review of Experiences around the World", American Journal of Industrial and Business Management, Vol. 07 No. 05, pp. 591–603, doi: 10.4236/ajibm.2017.75044.

OECD. (2025), Estudios Económicos de la OCDE: Perú 2025, Vol. 2025, OECD Publishing, Paris, doi: 10.1787/626594d0-es.

Paneerselvam, N., Muhammad, N.A., Azhan, A.M. and Muhammad, N. (2025), "Analyzing critical success factors in Lean Six Sigma training", International Journal of Productivity and Performance Management, Vol. 74 No. 4, pp. 1400–1424, doi: 10.1108/IJPPM-11-2023-0627.

Pyzdek, T. and Keller, P.A. (2018), The Six Sigma Handbook, Fifth Edition., McGraw-Hill Education, New York.

Snee, R.D. (2010), "Lean Six Sigma – getting better all the time", International Journal of Lean Six Sigma, Vol. 1 No. 1, pp. 9–29, doi: 10.1108/20401461011033130.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Real-Time Visual Inspection System for Automotive Chassis: development of a framework in a Toyota Plant

Marco Silva [1], Ana Raquel Xambre [2], Helena Alvelos [3] and Carlos Rodrigues [4]

[1] CIDMA, DEGEIT, University of Aveiro; mass@ua.pt
[2] CIDMA, DEGEIT, University of Aveiro; raquelx@ua.pt
[3] CIDMA, DEGEIT, University of Aveiro; helena.alvelos@ua.pt
[4] ESAN, University of Aveiro; cmor@ua.pt

Abstract

Quality inspection in automotive manufacturing, particularly in welding processes, remains a challenging task due to its reliance on manual and subjective evaluation methods. In the context of Quality 4.0, the integration of deep learning and computer vision offers new opportunities to improve inspection efficiency and real-time decision-making. This paper presents an industrial case study conducted in a Toyota manufacturing environment, focusing on the development of a framework for a real-time automated visual inspection system for automotive chassis. The study follows an Action-Research methodology, enabling the identification of process limitations and the iterative design of a solution aligned with real production requirements.

Based on a comprehensive analysis of state-of-the-art approaches, a modular framework is proposed, integrating image acquisition, dataset preparation, deep learning-based detection and segmentation, geometric measurement, and operator-oriented visualization. The approach uses YOLOv8 models for both object detection and weld segmentation, enabling not only the identification of structural components and weld regions but also the extraction of geometric features for quantitative analysis.

The work is currently at a testing stage, with detailed development and validation of the initial framework stages and a structured plan for the implementation of the remaining stages. The proposed approach addresses a gap in the literature by combining deep learning-based inspection with geometric measurement capabilities in a real industrial context. The results highlight the potential of the proposed framework to support the transition towards data-driven and real-time quality control systems. By integrating deep learning models with human-centred design principles, this work contributes to the advancement of Quality 4.0 solutions in manufacturing environments.

Keywords – Computer vision; Quality 4.0; Welding inspection; YOLOv8.

XL. INTRODUCTION

Quality inspection represents a fundamental pillar of the automotive industry, particularly in structural processes such as chassis welding, where defects can compromise vehicle safety, durability, and regulatory compliance (Colledani et al., 2014). Despite increasing levels of production automation, visual inspection tasks are still frequently performed by human operators, a factor that introduces subjectivity and high operational costs. With the evolution of Industry 4.0, technologies such as intelligent sensors, cyber-physical systems, and artificial intelligence have increasingly transformed manufacturing processes. In this context, the Quality 4.0 concept emerges as a digital extension of traditional quality management, promoting the use of advanced technologies to support real-time quality decision-making (Carvalho et al., 2021; Oliveira et al., 2025).

This paper presents an industrial case study conducted at a Toyota manufacturing plant, focusing on the design and implementation of an automated real-time visual inspection system for automotive chassis. The organization belongs to the automotive sector and operates under the "Toyota Way" (TW) which is defined by a set of core principles representing Toyota's organizational culture, incorporating the Toyota Production System (TPS) and the comprehensive management of the company's value chain (Jayamaha et al., 2014; Liker, 2021).

The TPS is recognized as a global standard for lean manufacturing, aiming to maximize value creation by eliminating waste (muda) and promoting a culture of continuous improvement (kaizen) (Monden, 1984; Lander & Liker, 2007). This system is built upon two foundational pillars: Just-in-Time (JIT) and Jidoka. While JIT ensures production efficiency by delivering only what is needed, Jidoka – often described as "automation with a human touch" – empowers both systems and workers to respond immediately to abnormalities, preventing the propagation of defective products (Powell, 2024). In the environment where the presented work was developed, the production floor is structured around key sectors – Logistics, Welding, Painting, Assembly, and Inspection – coordinated by Group Leaders (GLs), Team Leaders (TLs), and Team Members (TMs), who ensure the continuity of the production flow.

In the context of this study, particular attention is given to the welding control process. The initial state of this process was found to be highly manual, relying on visual inspection and sample-based manual measurements to verify weld bead integrity. This traditional approach was highly susceptible to human error, leading to issues ranging from missing structural components (welded brackets) to weld defects such as insufficient or excessive material. To address these challenges, this work proposes a

framework that incorporates computer vision and machine learning techniques. The proposed solution aims to target weld detection and measurement, alongside the confirmation of structural components. It demonstrates how the integration of deep learning, computer vision, and an operator-centered interface can significantly improve inspection performance and process quality.

The main contribution of this work is the development of a real-time visual inspection architecture designed for a high-demand industrial environment and currently in prototype/testing stage. Additionally, beyond addressing the operational challenges identified in the production line, this work also aims to contribute to reducing an existing gap in the literature regarding practical and low-cost implementations of automated weld inspection systems. In many industrial contexts, scalable solutions developed using accessible tools – such as Python-based computer vision frameworks and open-source machine learning models – are still relatively limited in documented industrial applications.

Despite some recent advances, limited research has focused on integrating YOLO-based detection with geometric measurement techniques for quantitative weld analysis in real industrial environments. Furthermore, most existing studies emphasize defect classification rather than combining detection with dimensional verification and process validation. In this context, the present work addresses this gap by leveraging YOLOv8 for both detection and segmentation tasks, enabling not only weld identification but also the extraction of geometric features for measurement and quality assessment.

XLI. Theoretical Background

A. Quality 4.0

The concept of Quality 4.0 (Q4.0) emerges from the digital transformation of Traditional Quality Management (TQM), reflecting the increasing integration of digital technologies into quality-related processes (Carvalho et al., 2021). Technologies such as data analytics, the Internet of Things (IoT), and Artificial Intelligence (AI) enable continuous monitoring of production systems, support predictive quality management, and enhance the effectiveness of problem-solving activities (Colledani et al., 2014). Recent systematic literature reviews indicate a clear transition from traditional inspection-based quality approaches toward data-driven quality management systems, proposing conceptual frameworks that integrate established quality methodologies with Industry 4.0 (I4.0) technologies. In this context, Q4.0 can be defined as the application of I4.0 technological capabilities to quality management, representing a new stage in the evolution of quality practices (Carvalho et al., 2021; Oliveira et al., 2025).

Manufacturing systems are progressively evolving toward increased flexibility, higher quality standards, and improved productivity (Zhong et al., 2017). The adoption of Quality 4.0 principles contributes to a wide range of organizational benefits, including the reduction of costs and quality-related issues such as defects, waste, and rework; improvements in product, service, and process quality; and faster, data-driven decision-making. Additionally, Q4.0 enhances customer satisfaction, increases transparency and traceability, and supports the identification of continuous improvement opportunities. It also fosters the development of new business models and strengthens workforce capabilities, ultimately contributing to improved overall operational performance (Mahin et al., 2024).

B. Real-Time Quality Control Using Computer Vision

Computer vision has emerged as a transformative technology, enabling computers to interpret and analyse visual data across a wide range of industrial applications, from medical diagnostics to manufacturing quality control (Ettalibi et al., 2024). In the context of Industry 4.0 and smart manufacturing, the integration of computer vision with machine learning and deep learning techniques has significantly accelerated the development of automated systems for real-time defect detection and classification (Islam et al., 2024; Patrashko & Gurau, 2026). This approach is crucial for enhancing the adaptability of quality control systems, allowing them to respond rapidly to evolving production requirements (Mijailović et al., 2023).

This technological evolution is largely driven by the need to overcome the limitations of traditional inspection methods, which often rely on intensive human intervention and struggle to keep pace with modern production rates, thereby compromising efficiency and accuracy in quality control processes (Kim et al., 2025; Sharma et al., 2023). Consequently, the growing demand for real-time quality control systems has directed research toward computer vision-based solutions that reduce dependence on manual judgment and improve decision-making processes (Sharma et al., 2023).

The adoption of artificial intelligence and real-time computer vision is redefining quality control paradigms, shifting from time-consuming manual inspections to automated systems that minimize errors and subjectivity while enhancing consistency and reliability (Ettalibi et al., 2024; Kim et al., 2024). Practical applications demonstrate that these technologies play a key role in optimizing and organizing material flow within production environments, improving overall efficiency and reducing operational errors (Medina et al., 2024).

In particular, recent advances in algorithms such as the YOLO (You Only Look Once) family have revolutionized visual inspection by providing cost-effective solutions that significantly reduce the need for human intervention (Ashourpour & Azizpour, 2025; Sapkota et al., 2024). This capability enables proactive defect detection, preventing defective components from progressing through the production process, improving the overall process efficiency (Weiss et al., 2024). Furthermore, the application of advanced deep learning techniques such as YOLO has demonstrated high accuracy in real-time defect detection and classification tasks, outperforming conventional inspection methods in both efficiency and scalability (Qi et al., 2024).

C. Visual Inspection using YOLO

Among deep learning approaches, the YOLO (You Only Look Once) family has become a leading solution due to its ability to perform object detection as a regression problem, enabling real-time localization and classification of defects directly from images without the need for separate region proposal stages (Ramos & Sappa, 2025). More recent versions, such as YOLOv8, further improve this capability by incorporating enhanced feature extraction mechanisms and optimized architectures, allowing higher accuracy and faster inference suitable for real-time industrial applications.

In industrial contexts, YOLO-based models have demonstrated strong performance in defect detection tasks, including weld inspection. Performance is typically evaluated using metrics such as mean Average Precision (mAP), which measures the accuracy of object detection models by averaging precision across different recall levels and Intersection over Union (IoU) thresholds. Specific metrics such as AP50 and AP75 correspond to Average Precision calculated at IoU thresholds of 0.50 and 0.75, respectively, where higher thresholds indicate stricter localization requirements. Studies report high accuracy levels, such as YOLOv3 achieving mAP values of 98.25% (AP75) and 84.36% (AP50) in wheel weld inspection scenarios, confirming its effectiveness in controlled environments (Ali & Zhang, 2024; Mao & Hong, 2025). Additionally, YOLO has been widely applied to surface defect detection across production lines, reinforcing its suitability for high-throughput industrial inspection systems (Ali & Zhang, 2024).

Survey studies highlight YOLO as a robust one-stage detector with strong generalization capabilities and low background error in complex industrial environments. However, limitations arise when detecting small or subtle defects, such as micro-cracks or pores in weld seams, due to the grid-based detection mechanism. These challenges have driven the evolution of YOLO architectures, including YOLOv8, which incorporates improvements such as multi-scale feature extraction and more efficient feature fusion strategies to enhance small object detection (Yi et al., 2024).

Recent advances in industrial inspection also include the integration of YOLO with transformer-based architectures, such as Swin-Transformer-YOLO, which improve the detection of fine-grained defects on metallic surfaces. Earlier variants such as YOLOv4 and YOLOv5 remain widely used as baselines due to their balance between accuracy and computational efficiency, while newer versions introduce enhancements such as feature pyramids and attention mechanisms to address challenges associated with small-defect detection (Hussain, 2024; Jiang et al., 2022).

Overall, YOLO-based approaches achieve high detection performance, often exceeding 90% mAP while maintaining real-time processing capability, making them suitable for integration into automated quality control systems. Nevertheless, challenges remain regarding the detection of low-contrast defects, dataset availability, and domain generalization, particularly when models are applied across different weld configurations or industrial environments (Hussain, 2024; Jiang et al., 2022; Wang & Liao, 2024).

XLII. Methodology

This study follows an Action-Research methodology, enabling the iterative development and evaluation of a real-time visual inspection system within a real industrial environment. This approach is particularly suitable for industrial contexts, as it combines problem-solving with continuous improvement through cycles of planning, implementation, evaluation, and refinement (Penaranda et al., 2010).

The research was conducted in collaboration with a Toyota manufacturing plant, focusing on the welding control process identified as a critical area for improvement. The initial diagnosis phase revealed a highly manual inspection process, characterized by operator-dependent visual verification and sample-based measurements, leading to potential quality risks.

The development of the proposed system was structured into iterative cycles aligned with PDCA principles, including: (i) problem identification and requirements' definition; (ii) development of a framework that includes computer vision and deep learning models; (iii) validation of a prototype industrial application in a real industrial context. However, it is important to note that, at this stage of the research, no final implementation cycles of the prototype have yet been carried out, with the work remaining at the validation and refinement phase.

A. Industrial Context and System Requirements

The proposed inspection system was developed and validated in a real automotive chassis production line within a Toyota manufacturing facility. The inspection process focuses on verifying weld integrity and confirming the presence of structural components along the chassis structure. These inspections must comply with quality standards while respecting the cycle-time constraints imposed by the production line.

Traditionally, weld inspection in this environment relied on manual visual verification performed by operators, complemented by occasional manual measurements. This approach introduces variability due to operator subjectivity and increases the risk of undetected defects or missing components.

During the initial process analysis, several functional requirements were identified for the automated inspection system: i) Real-time inspection capability, compatible with the production cycle time; ii) Automatic detection and counting of welds, ensuring that all required weld points are present; iii) 2D Geometric measurement of weld regions, allowing dimensional verification; iv) Detection of structural brackets and components, preventing missing or incorrectly assembled parts; v)

Automatic identification of chassis type, based on the configuration of detected structural components; vi) Operator-oriented visualization interface suitable for use on the production floor.

The identification of the chassis type was particularly important because the production line handles different vehicle configurations. Specifically, the system needed to distinguish between single-cabin and double-cabin chassis, which differ in the number and spatial arrangement of welded brackets. Detecting these brackets allows the system to infer the chassis configuration and verify whether the correct set of components has been assembled.

XLIII. Results and Discussion

Given that the system is currently at a prototype stage, this section focuses on the framework design, initial implementation steps (namely image acquisition and dataset preparation), and implementation planning, rather than on full experimental validation. Therefore, the "results" presented in this section are primarily conceptual and methodological, reflecting the structured synthesis of insights derived from the literature and their translation into a practical inspection framework.

The development of the proposed framework was supported by a structured review of state-of-the-art visual inspection systems, particularly those based on computer vision and deep learning approaches for industrial applications. The selection of relevant studies was conducted using scientific databases (Scopus and Web of Science), based on keywords such as "visual inspection systems", "industrial computer vision", "weld inspection", "YOLO", "defect detection", and "real-time quality control". The review focused on systems addressing manufacturing contexts, with particular emphasis on approaches integrating detection, segmentation, and measurement capabilities.

The selected works were analysed according to key dimensions associated with the design of inspection systems, including: (i) image acquisition setups and environmental constraints; (ii) dataset preparation and annotation strategies; (iii) model architectures for detection and segmentation; (iv) evaluation metrics and performance considerations; (v) visualization and operator interaction; and (vi) integration with industrial production lines. Particular attention was given to the integration of geometric measurement with defect detection, as these aspects are consistently highlighted in the literature as critical for enabling application-oriented inspection solutions. This multi-dimensional analysis directly informed the definition of the proposed framework presented in the following section.

A. Inspection Framework

The literature consistently highlights that effective real-time inspection systems rely on the integration of multiple interconnected components, including image acquisition, dataset preparation, model training, defect detection, and result visualization (Kim et al., 2024). In particular, studies on YOLO-based inspection systems emphasize the importance of modular architectures that enable scalability, flexibility, and integration with existing industrial processes (Ramos & Sappa, 2025; Yi et al., 2024).

Based on this analysis, a modular framework was defined to guide the development of the proposed inspection system. The framework consists of five main stages: (i) image acquisition, (ii) dataset annotation and preparation, (iii) deep learning-based detection and segmentation, (iv) geometric measurement, and (v) operator-oriented visualization, as indicated in figure 1.

Fig. 8. Proposed inspection framework for weld detection and measurement

The adoption of a modular structure is strongly supported in the literature, as it allows each stage of the inspection process to be independently adapted to specific industrial requirements. For instance, previous works highlight that the performance of deep learning models is highly dependent on dataset quality, which justifies the separation between data acquisition and annotation stages (Hussain, 2024; Jiang et al., 2022). Similarly, the integration of detection and measurement modules is increasingly reported as a key requirement for moving from qualitative defect detection toward quantitative quality assessment (Wang & Liao, 2024).

Furthermore, real-time industrial inspection systems require not only high detection accuracy but also efficient processing pipelines and intuitive interfaces for operator interaction. The inclusion of a visualization stage in the framework is therefore aligned with recent research emphasizing human-centred design and the role of operators in supervising automated quality control systems (Sharma et al., 2023).

Overall, the proposed framework reflects a synthesis of best practices identified in the literature, providing a structured approach for the development of real-time visual inspection systems capable of meeting industrial requirements related to accuracy and usability.

B. Detailed Analysis of Initial Stages and System Planning

This section presents a detailed analysis of the first stages of the proposed framework, focusing on image acquisition and dataset preparation, as well as the planning of the subsequent stages.

1) Image Acquisition

The image acquisition stage represents a critical component of the inspection system, as the quality of the captured data directly influences the performance of subsequent stages. In industrial environments, several challenges must be considered when designing this stage.

One of the main challenges is lighting variability, particularly in metallic surfaces such as welded chassis components. Reflections, shadows, and inconsistent illumination can significantly affect the visibility of weld regions and compromise accuracy. To mitigate these issues, careful positioning of lighting sources and the use of controlled illumination conditions were considered.

Another important factor is the selection and positioning of the camera system. The resolution of the camera must be sufficient to capture fine details of weld regions, especially when small defects are of interest. At the same time, the camera must be positioned to ensure consistent framing of the inspection area, minimizing variability between captured images.

Additionally, factors such as camera angles, distance to the object, and vibration in the production environment must be taken into account. Inconsistent positioning can introduce noise in the dataset and reduce model robustness. Therefore, maintaining a fixed acquisition setup is considered to be essential.

2) Dataset Preparation and Annotation

Dataset preparation is a critical stage in the development of deep learning-based inspection systems. In this work, particular attention was given to ensuring that the dataset accurately represents real production conditions.

One of the main challenges in this stage is the annotation of weld regions, which often present irregular shapes and undefined boundaries. Unlike rigid objects, welds can vary significantly in geometry, making consistent annotation more difficult. The use of polygon-based annotations is therefore essential to capture the true shape of weld regions and support segmentation tasks.

Figures 2 and 3 illustrate examples of the annotation process using Roboflow, a cloud-based platform for dataset management, annotation, and organization in computer vision projects. Figure 2 presents an example of weld annotation performed using the polygon tool, which is essential for capturing the irregular geometry of weld regions required for segmentation tasks. Figure 3 shows an example of a specific bracket annotated using the bounding box annotation tool, which is more suitable for object detection tasks. The combination of these two annotation approaches is particularly important, as it enables the system to support both precise segmentation of welds and reliable detection of structural components.

Fig. 9. Example of weld annotation using polygon masks in Roboflow

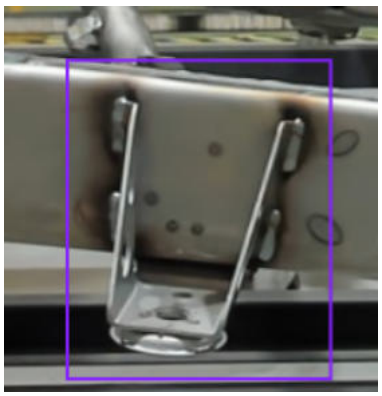


Fig. 10. Example of bracket annotation using bounding boxes in Roboflow

The dataset was structured into different classes to support different stages of the inspection system. Specifically, structural components (brackets) were annotated to train the detection model responsible for identifying chassis configuration, enabling the distinction between different chassis types. In parallel, weld regions and reference points were annotated to support segmentation and subsequent geometric measurement.

The inclusion of reference points with known physical dimensions represents a key aspect of the dataset design. These points serve as calibration markers, allowing the system to estimate real-world dimensions by comparing pixel distances. In particular, weld measurements are obtained by relating the pixel dimensions of weld regions to the nearest reference point, enabling the conversion from image space to physical measurements.

Additionally, it is important to ensure dataset variability, including different lighting conditions, weld appearances, and chassis configurations. This variability is essential for improving model generalization and reducing overfitting.

3) Planning of Detection and Segmentation Models

The third stage of the framework consists of implementing deep learning models for detection and segmentation, based on the YOLOv8 architecture. Two complementary models are considered in the proposed approach.

The YOLOv8-Detect model is designed to identify structural components, particularly brackets, which are directly associated with the chassis configuration. By detecting these components, the system is able to infer the type of chassis and verify the correct presence of required elements.

In parallel, the YOLOv8-Seg model is intended to detect and segment weld regions using the annotated polygon masks. This segmentation output provides detailed spatial information, which is essential for enabling geometric analysis and measurement in subsequent stages.

This dual-model approach allows the system to combine structural verification (detection) with weld analysis (segmentation), supporting both qualitative and quantitative inspection tasks.

4) Planning of Geometric Measurement

The fourth stage focuses on geometric measurement, which is based on the segmentation outputs generated by the YOLOv8-Seg model. The goal is to extract geometric features of weld regions and convert them into real-world dimensions.

This process relies on the reference points introduced during the annotation stage. Since these points have known physical dimensions, they enable pixel-to-distance calibration. By comparing the pixel measurements of weld regions with the corresponding reference points, it becomes possible to determine dimensions such as weld length, width, and area.

The expected outcome of this stage is the ability to perform quantitative weld evaluation, moving beyond simple detection toward measurable quality verification.

5) Planning of Visualization Interface

The final stage involves the integration of the inspection system into a visualization interface using Streamlit, an open-source Python framework for building interactive web-based applications. Streamlit enables the rapid development of user interfaces that can display data, images, and model outputs in real time, making it particularly suitable for industrial applications where ease of use and quick deployment are required.

This interface is intended to provide real-time feedback to operators by displaying detection results, segmented weld regions, and measurement outputs. The design prioritizes usability and intuitive interaction, ensuring that operators can easily interpret inspection results without requiring advanced technical knowledge. This supports a transition from manual inspection toward supervised automated inspection, where operators focus on decision-making and process control.

The interface is expected to provide real-time visualization of captured images together with detected weld regions, structural components, and corresponding measurement outputs. Specifically, it should display processed images with annotated weld regions and detected components, as well as inspection results and measurement indicators. This visual representation enables operators to quickly interpret the inspection outcome and identify potential anomalies in the production process.

XLIV. Conclusions

This study presented the development of a structured framework for real-time automated visual inspection of automotive chassis welding processes, grounded in both industrial requirements and the literature reviewed. Through the adoption of an Action-Research methodology, the work enabled the systematic identification of process limitations and the definition of a solution aligned with real production constraints.

The results demonstrate that existing approaches in the literature can be effectively consolidated into a modular inspection framework composed of five key stages: image acquisition, dataset preparation, deep learning-based detection and segmentation, geometric measurement, and operator-oriented visualization. This structured approach provides a clear and replicable pathway for the development of computer vision-based inspection systems, particularly in high-demand industrial environments.

A key contribution of this work lies in the integration of YOLOv8-based detection and segmentation models with a geometric measurement strategy supported by reference points. This combination enables not only the reliable identification of weld regions and structural components but also the extraction of quantitative information for dimensional verification. As such, the proposed approach advances beyond traditional defect detection toward a more comprehensive, data-driven quality control paradigm.

From an industrial perspective, the proposed framework is fully aligned with the principles of Quality 4.0, supporting the digitalization of inspection processes, improving consistency, and reducing reliance on subjective manual evaluations. Furthermore, the planned integration of a Streamlit-based interface reinforces the importance of human-centered system design, enabling operators to interact with and supervise automated inspection processes in real time.

Despite its potential, some limitations must be acknowledged. The system is currently at a prototype stage and requires further validation under real production conditions to assess its robustness and scalability. Additionally, the exclusive reliance on YOLO-based models may introduce limitations in detecting small or low-contrast defects, highlighting the need for future enhancements and possible hybrid approaches.

Future work will focus on completing the implementation of all stages defined in the proposed framework, ensuring full system integration and validation in a real industrial environment. This includes the deployment of detection and segmentation models, the implementation of geometric measurement capabilities, and the development of the visualization interface, followed by comprehensive performance evaluation under varying operational conditions. Further improvements will also address dataset expansion, model optimization, and the exploration of advanced measurement approaches, such as multi-view or 3D analysis.

Ultimately, this study demonstrates that the integration of deep learning technology and computer vision with structured industrial frameworks can significantly improve quality control processes. By combining deep learning, computer vision, and the principles of the Toyota Production System, it is possible to achieve substantial improvements in operational performance while maintaining a strong human-centered approach, fully aligned with the values of the Toyota Way.

Acknowledgments

This work was supported by CIDMA – Center for Research and Development in Mathematics and Applications (https://ror.org/05pm2mw36), through the Portuguese Foundation for Science and Technology (FCT, https://ror.org/00snfqn58), under projects UID/04106/2025 (https://doi.org/10.54499/UID/04106/2025) and UID/PRR/04106/2025 (https://doi.org/10.54499/UID/PRR/04106/2025). This work was also supported by the PhD scholarship with reference 2024.04472.BDANA, granted by FCT.

References

Ali, M. and Zhang, Z. (2024) "The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection", Computers, 13, p. 336. https://doi.org/10.3390/computers13120336.

Ashourpour, M. and Azizpour, G. (2025) "Edge-deployed deep learning for automated quality control in industrial assembly: A case study in real-time defect detection", Procedia CIRP, 134, pp. 735–740. https://doi.org/10.1016/j.procir.2025.02.180.

Carvalho, A.V., Enrique, D.V., Chouchene, A. and Charrua-Santos, F. (2021) "Quality 4.0: An overview", Procedia Computer Science, 181, pp. 341–346. https://doi.org/10.1016/j.procs.2021.01.176.

Colledani, M., Tolio, T., Fischer, A., Iung, B., Lanza, G., Schmitt, R. and Váncza, J. (2014) "Design and management of manufacturing systems for production quality", CIRP Annals, 63(2), pp. 773–796. https://doi.org/10.1016/j.cirp.2014.05.002.

Ettalibi, A., Elouadi, A. and Mansour, A. (2024) "AI and computer vision-based real-time quality control: A review of industrial applications", Procedia Computer Science, 231, pp. 212–220. https://doi.org/10.1016/j.procs.2023.12.195.

Hussain, M. (2024) "YOLOv5, YOLOv8 and YOLOv10: The go-to detectors for real-time vision", arXiv. https://doi.org/10.48550/arxiv.2407.02988.

Jayamaha, N.P., Wagner, J.P., Grigg, N.P., Campbell-Allen, N.M. and Harvie, W. (2014) "Testing a theoretical model underlying the 'Toyota Way': An empirical study involving a large global sample of Toyota facilities", International Journal of Production Research, 52(14), pp. 4332–4350. https://doi.org/10.1080/00207543.2014.883467.

Jiang, P., Ergu, D., Liu, F., Ying, C. and Ma, B. (2022) "A review of YOLO algorithm developments", Procedia Computer Science, 199, pp. 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135.

Kim, B., Shin, M. and Hwang, S. (2024) "Design and development of a precision defect detection system based on a line scan camera using deep learning", Applied Sciences, 14, p. 12054. https://doi.org/10.3390/app142412054.

Kim, S., Nguyen, T. and Yoon, J. (2025) "A systematic review of machine vision applications in factory and manufacturing processes: From quality control to predictive diagnostics", International Journal of Precision Engineering and Manufacturing – Green Technology. https://doi.org/10.1007/s40684-025-00769-2.

Lander, E. and Liker, J. (2007) "The Toyota Production System and art: Making highly customized and creative products the Toyota way", International Journal of Production Research, 45, pp. 3681–3698. https://doi.org/10.1080/00207540701223519.

Liker, J. (2021) The Toyota Way. New York: McGraw-Hill.

Mahin, M., Kadasah, N., Alsabban, A. and Albliwi, S. (2024) "Exploring the landscape of Quality 4.0: A comprehensive review of its benefits, challenges, and critical success factors", Production and Manufacturing Research, 12(1). https://doi.org/10.1080/21693277.2024.2373739.

Mao, M. and Hong, M. (2025) "YOLO object detection for real-time fabric defect inspection in the textile industry: A review of YOLOv1 to YOLOv11", Sensors, 25. https://doi.org/10.3390/s25072270.

Medina, A., Bradley, R., Xu, W., Ponce, P., Anthony, B. and Molina, A. (2024) "Learning manufacturing computer vision systems using tiny YOLOv4", Frontiers in Robotics and AI, 11, p. 1331249. https://doi.org/10.3389/frobt.2024.1331249.

Mijailović, Đ., Đorđević, A., Stefanović, M. and Erić, M. (2023) "Quality control in the manufacturing industry based on the application of computer vision", Journal of Industrial and Business Intelligence, 1, pp. 161–166. https://doi.org/10.61552/JIBI.2023.04.001.

Monden, Y. (1984) Toyota Production System. Atlanta: Institute of Industrial Engineers.

Oliveira, D., Alvelos, H. and Rosa, M.J. (2025) "Quality 4.0: Results from a systematic literature review", The TQM Journal, 37(2), pp. 379–456. https://doi.org/10.1108/TQM-01-2023-0018.

Patrashko, D.Y. and Gurau, V. (2026) "Machine learning-powered vision for robotic inspection in manufacturing: A review", Sensors, 26, p. 788. https://doi.org/10.3390/s26030788.

Penaranda, N., Mejia, R., Romero, D. and Molina, A. (2010) "Implementation of product lifecycle management tools using enterprise integration engineering and action-research", International Journal of Computer Integrated Manufacturing, 23(10), pp. 853–875. https://doi.org/10.1080/0951192X.2010.495136.

Powell, D.J. (2024) "Artificial intelligence in lean manufacturing: Digitalization with a human touch?", International Journal of Lean Six Sigma, 15(3), pp. 719–729. https://doi.org/10.1108/IJLSS-05-2024-256.

Qi, Z., Ding, L., Li, X., Hu, J., Lyu, B. and Xiang, A. (2024) "Detecting and classifying defective products in images using YOLO".

Raisul Islam, M., Zakir Hossain Zamil, M., Eshmam Rayed, M., Mohsin Kabir, M., Mridha, M.F., Nishimura, S. and Shin, J. (2024) "Deep learning and computer vision techniques for enhanced quality control in manufacturing processes", IEEE Access, 12, pp. 121449–121479. https://doi.org/10.1109/ACCESS.2024.3453664.

Ramos, L. and Sappa, A. (2025) "A decade of You Only Look Once (YOLO) for object detection", IEEE Access, 13, pp. 192747–192794. https://doi.org/10.1109/ACCESS.2025.3630988.

Sapkota, R., Qureshi, R., Flores-Calero, M. et al. (2025) "YOLO advances to its genesis: A decadal and comprehensive review of the You Only Look Once series", Artificial Intelligence Review, 58. https://doi.org/10.1007/s10462-025-11253-3.

Sharma, H., Kumar, H., Gupta, A. and Shah, M. (2023) "Computer vision in manufacturing: A bibliometric analysis and future research propositions".

Wang, C. and Liao, H. (2024) "YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems", arXiv. https://doi.org/10.48550/arxiv.2408.09332.

Weiss, E., Caplan, S., Horn, K. et al. (2024) "Real-time defect detection in electronic components during assembly through deep learning", Sensors, 24, p. 3721. https://doi.org/10.3390/s24123721.

Yi, L., Akbar, M., Wahab, M., Rosdi, B., Fauthan, M. and Shrifan, N. (2024) "The prospect of artificial intelligence-based wood surface inspection: A review", IEEE Access, 12, pp. 84706–84725. https://doi.org/10.1109/ACCESS.2024.3412928.

Zhong, R.Y., Xu, X., Klotz, E. and Newman, S.T. (2017) "Intelligent manufacturing in the context of Industry 4.0: A review", Engineering, 3(5), pp. 616–630. https://doi.org/10.1016/J.ENG.2017.05.015.

7th International Conference on Quality Engineering and Management

2-3 July 2026, Universidade NOVA de Lisboa, Lisbon, Portugal

Complementarity of Concepts: The Synergies between Society 5.0, Industry 5.0 and Quality 5.0

Flaviane Moblicci [1], Paulo Sampaio [2] and André M. Carvalho [3]

[1] Mestrado em Engenharia e Gestão da Qualidade – Universidade do Minho; flamoblicci@gmail.com
[2] Departamento de Produção e Sistemas; Centro ALGORITMI, Universidade do Minho; paulosampaio@dps.uminho.pt
[3] UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal; andremc@fct.unl.pt

***Abstract*—This article explores the interconnection and synergies between the concepts of society 5.0, industry 5.0, and quality 5.0, highlighting the importance of integrating technology and innovation with human-centered principles and sustainability concerns. It concludes that, in a complementary way, these concepts drive innovation, efficiency, and collaborative responses to social challenges, promoting sustainable development and continuous improvement in quality of life.**

Keywords – Society 5.0, Industry 5.0, Quality 5.0, Human-Centered Approach.

XLV. Introduction

We are currently living in the era of Industry 4.0, characterized by the pursuit of business automation and the integration of emerging technologies such as Artificial Intelligence and Machine Learning. While this reality remains distant for some industries, it has already been surpassed by certain scholars in the field. Industry 4.0 addresses human centrality, sustainability, and resilience as consequences and follows a technological approach. Unlike Industry 4.0, Industry 5.0 shifts the focus from individual technologies to a systematic approach. This approach enables the industry to achieve goals beyond employment and growth, placing the well-being of industrial workers at the center of the production process. This may explain why Industry 5.0 is considered a different type of Industrial Revolution compared to previous ones (Xu et al., 2021), one that is framed ina. Broader societal framing – that of Society 5.0. Simultaneously, the evolution of Quality Management has seen similar calls for a fifth stage - Quality 5.0 - where quality management extends beyond the product/service and customer focus to encompass a broader perspective, incorporating sustainability, customer satisfaction, employee engagement, and social responsibility.

This article explores the interconnections between these concepts and how they complement each other in terms of objectives, technologies, and human values. It examines potential intersections and synergies among the three concepts, technological barriers, social and cultural challenges that may arise from this synergy, as well as opportunities for innovation, growth, and future visions. Focusing on the key values - Human-Centeredness, Sustainability, Resilience, and Technology - this study establishes a framework for understanding the synergies between these concepts.

This study also seeks to answer the following research questions:

1. Can the concepts and values of Society 5.0, along with the Japanese experience, complement the principles of Industry 5.0?
2. Can Quality contribute in any way to enabling industries to achieve the Industry 5.0 standard?
3. Is it possible to establish a synergy between Society 5.0, Industry 5.0, and Quality 5.0? Do they complement each other?

The originality of the proposed conceptual model lies not merely in the synthesis of Society 5.0, Industry 5.0, and Quality 5.0, but in the establishment of an integrative and relational framework that explicitly connects their underlying principles, operational dimensions, and quality-oriented implications. While previous studies have typically examined these concepts in isolation or through partial associations, the present framework advances the literature by positioning Quality 5.0 as a mediating and enabling dimension between technological transformation and societal value creation. In this sense, the model contributes by clarifying how human-centricity, sustainability, resilience, digitalization, and stakeholder value can be systematically aligned within a unified quality management perspective. Furthermore, the framework provides a conceptual basis for future empirical studies and managerial applications aimed at assessing organizational readiness and maturity in the transition toward Society 5.0-oriented systems. Unlike existing frameworks that address these constructs separately, the proposed model emphasizes their interdependence and proposes a quality-centered integration logic capable of supporting both organizational transformation and societal outcomes.

A. Societies from 1.0 to 5.0

In 2016, the concept of Society 5.0 was introduced by the Japanese government. It envisions a society driven by scientific and technological innovation, with Human-Centeredness, Super Smart, and Lean Society as its core values. The concept was proposed to balance economic advancement with addressing Japan's social challenges, such as low birth rates, lack of competitiveness, and an aging population (Huang et al., 2022). However, to introduce the concepts of Society 5.0, it is necessary to first present the preceding societies throughout history, along with the Industrial Eras and Quality Evolutions.

Society 1.0, also known as the Hunting Society, began when humans used their knowledge to seek food and survive (BÖLÜKBAŞI, 2021). Humans were nomads, constantly migrating in search of food, with an unproductive economy (Narvaez Rojas et al., 2021). Society 2.0, or the Agricultural Society, was marked by the advent of agriculture, which enabled society to reach a new level of evolution (BÖLÜKBAŞI, 2021). Agricultural production led to the formation of tribes, and humans stopped being nomads, becoming self-sufficient through cultivation and manual labour. (Narvaez Rojas et al., 2021). Society 3.0, or the Industrial Society, was marked by the replacement of manual labour with machines and factories. Technological advancements brought mass production, reducing costs and production time while increasing wages and resources (Narvaez Rojas et al., 2021). Society 4.0, or the Information Society, is characterized by the pursuit of information and knowledge. (BÖLÜKBAŞI, 2021). The internet revolutionized society, providing faster access to information and greater interaction among people. Companies further automated their processes, always seeking efficiency and increased profits, while individuals had to invest in new skills to succeed in the market (Narvaez Rojas et al., 2021). In January 2016, the Japanese government released details about the Fifth Basic Plan for Science and Technology (2016-2020), in which the initiative known as "Society 5.0" gained recognition. Huang et al. (2022) define its core values as Human-Centered, Super Smart, and Lean, aiming to balance economic growth with societal well-being. Society 5.0 integrates technology and society to meet individual needs, enhance quality of life, and address social issues through a human-centered, sustainable approach combining physical and virtual domains (Carayannis & Morawska-Jancelewicz, 2022).

B. Industrial Eras from 1.0 to 5.0

Different industrial eras are clearly marked by major transformations in production contexts, and often branded as "revolutions". The First Industrial Revolution began in the 18th century in England, and was characterized by the transition from manual production methods to machine-based manufacturing, new chemical processes, and iron production (Groumpos, 2021).The Second Industrial Revolution began around 1870 with electricity and assembly lines. Innovations like the telegraph, railroads, and electricity drove rapid industrial growth and infrastructure development in various sectors. (Groumpos, 2021). The Third Industrial Revolution began around the 1950s, shortly after World War II. The use of computers and the invention of the internet brought automation to production processes, reducing human assistance, and robots began being developed and utilized (Groumpos, 2021). The Fourth Industrial Revolution began in 2011 following a technology initiative developed by the German government, with its main feature being the concept of the Smart Factory (Xu et al., 2021). The Smart Factory concept has as mission to develop, apply, and disseminate industrial technologies through both research and practical implementation (Zuehlke, 2010). With regard to the application of technology in Industry 4.0, the Boston Consulting Group identified nine core technologies essential to its implementation: Big Data, Autonomous Robots, Simulation, Horizontal and Vertical System Integration, the Internet of Things (IoT), Cybersecurity, Cloud Computing, Additive Manufacturing (systems that control all stages of the production process in real time), and Augmented Reality (Xu et al., 2021).

As Industry 4.0 (I4.0) has focused on advanced technologies to optimize processes and increase efficiency, there have been some concerns about the lost focus on human, social and environmental aspects of this approach (Rad et al., 2022; Raja Santhi & Muthuswamy, 2023). Gaining prominence around 2020, the concept of Industry 5.0 was advanced, addressing issues such as sustainability, human-orientation and resilience (Breque et al., 2021). Reiforced by the European Commission, which recognized the need for a model that aligns industrial growth with ecological and social responsibilities (Xu et al., 2021), the concept of industry 5.0 (I5.0) I5.0 involves a shift towards a more collaborative, customized, and flexible production processes that values human intelligence and creativity, pursuing optimization between technological progress and human well-being (Longo et al., 2020). Industry 5.0 is thus a concept that represents a new phase in the Industrial Revolution, with a focus on a human-centered, sustainable, and resilient manufacturing system (Figure 1). It aims to integrate human creativity and skills with advanced technologies—such as Collaborative Robots, Artificial Intelligence (AI), and Big Data—to create a more efficient and adaptable production environment (Huang et al., 2022). Unlike previous industrial models, Industry 5.0 promotes synergy between humans and machines, highlighting the value of personalization, sustainability, and innovation (Nahavandi, 2019).

Fig. 11. Industry 5.0 values. Source: (Breque et al., 2021).

C. Evolution of Quality Management Eras From 1.0 to 5.0

Frick & Grudowski (2023) refer to quality even before the Industrial Revolution, labeling it as "Quality 1.0." According to the authors, the concept of quality was centered on the characteristics or attributes of a good or service, focusing on craftsmanship, raw materials, and functionality. However, other studies cite quality beginning with the First Industrial Revolution, a period initially marked by a lack of concern for product quality (Hattinger & Stylidis, 2023).

The Second Industrial Revolution marked a turning point in the history of quality (Hattinger & Stylidis, 2023). Concepts still used today, such as Statistical Process Control (SPC) and Production Process Control through waste elimination (Deleryd & Fundin, 2020), significantly advanced the status of quality.

The third stage of quality evolution was significantly influenced by digital technologies and automation (Hattinger & Stylidis, 2023). The emergence of Total Quality Management (TQM), Six Sigma, and Lean Production emphasizing customer satisfaction and operational excellence (Deleryd & Fundin, 2020), helping companies reduce production costs and improve product and service quality (Hattinger & Stylidis, 2023). Business excellence models such as the EFQM (European Foundation for Quality Management) Model and ISO 9000 were introduced based on these new quality concepts (Deleryd & Fundin, 2020).

Quality 4.0, covering the period from 1995 to 2020, is characterized by total customer satisfaction, innovation, and sustainability. Frick & Grudowski (2023) argue that merely meeting specifications is insufficient; a product must fulfill customer needs and exceed expectations through user experience, convenience, after-sales service, and environmental impact.

Quality 5.0 emerges as an evolution of Quality 4.0, aligning with the transformations brought by Industry 5.0 and Society 5.0 (Djordjevic et al., 2024). According to the authors, Quality 5.0 is not merely an extension of Quality 4.0 but represents a paradigm shift requiring a reassessment of Quality Management practices, with a renewed focus on people, sustainability, and social responsibility. Its key characteristics include Human-Centric Focus, Integration of Advanced Technologies, Sustainability and Social Responsibility, Knowledge-Based Innovation (Maljugić et al., 2024), Data Management and Security, Development of New Competencies (Djordjevic et al., 2024), and Proactive Quality Control (Frick, 2023).

Figure 2 illustrates the evolution of quality models in accordance with the Industrial Revolutions.

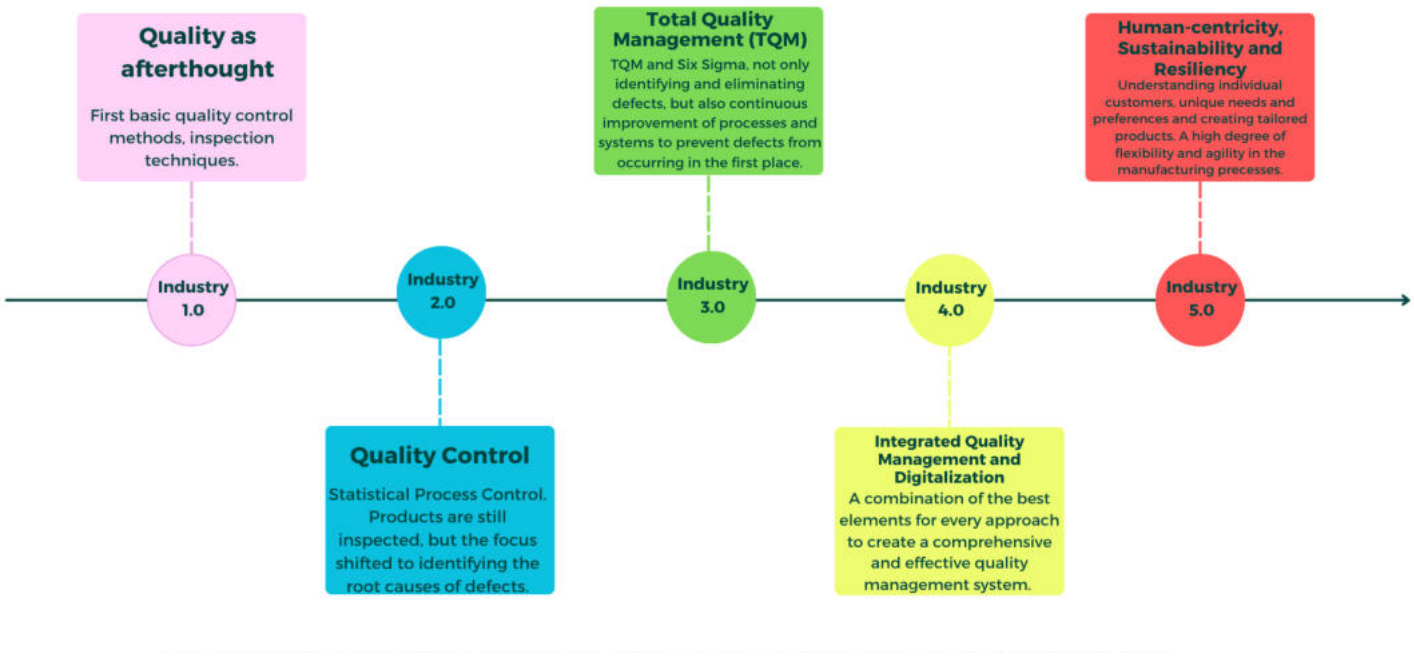


Fig. 12. Evolution of Quality models (Adapted from Hattinger & Stylidis, 2023)

XLVII. Synergies between concepts

The evolution of Societies, Industrial Eras, and Quality has influenced Industry 5.0, Society 5.0, and Quality 5.0 by establishing a historical and social context that emphasizes the importance of sustainability, human well-being, and social responsibility. Each phase has provided lessons that have shaped how, particularly, technology is applied to meet the needs of its contemporary society.

Since 2017, the structures of Society 5.0, Industry 4.0/5.0, and Quality 4.0/5.0 have evolved to highlight the integration of technological advancements with human-centered values. This evolution represents a shift towards systems that are not only technologically advanced but also demonstrate a commitment to social responsibility and sustainability (Maljugić et al., 2024).

A. Human-Centric Approach

Society, Industry, and Quality 5.0 highlight the importance of human-centered design, aiming to improve the quality of life and ensure that technological innovations meet human needs. Society 5.0 is mainly concerned with establishing a superior quality of life, while Industry 5.0 focuses on the well-being of workers involved in manufacturing. Society 5.0 promotes solving social problems, such as inclusion, equality, and quality of life. This approach can guide Industry 5.0 in developing meaningful products that address these needs, ensuring that technological innovations are relevant and beneficial to society (Ulgen & Muneeb, 2028).

The relationship between Society 5.0 and Industry 5.0 is intrinsic, as both approaches aim to integrate technology and innovation to meet human and social needs while fostering sustainable and human-centered development (Barata & Kayser, 2028).The synergy between Industry 5.0 and Society 5.0 refers to the interconnection and mutual support of these concepts. This synergy is evident when industrial advancements contribute to broader social goals. Collaboration between different sectors is essential for creating solutions that meet society's needs, drive industrial innovation, and result in more relevant products for communities (Ulgen & Muneeb, 2024).

Quality 5.0 is defined as the creation, development, and implementation of definitions, principles, methods, tools, and metrics that meet the needs of Society 5.0. This implies a more holistic and human-centered approach that considers not only efficiency and effectiveness but also people's well-being and happiness (Saraiva, 2023). Society 5.0 will transform TQM and enhance Quality Management Systems by focusing on human-centered approaches, advancing technological integration, fostering innovation, promoting sustainability, encouraging collaboration, and developing competencies (Arıcı & Kitapci, 2021).

Industry 5.0 emphasizes collaboration between humans and machines, adopting a sociotechnical approach. Quality Management, by considering both social aspects (such as organizational culture and employee skill development) and technological aspects (such as processes and tools), aligns perfectly with this vision. Industry 5.0 also values active employee participation and team collaboration, which Quality Management can facilitate by encouraging open communication and valuing employee contributions. This synergy is essential for addressing future challenges and ensuring sustainable and high-

quality performance in organizations (Ali & Johl, 2024). International frameworks such as EFQM (European Foundation for Quality Management), OHSAS (Occupational Health and Safety Assessment Series), and ISO 45001 are fundamental to achieving operational excellence, as they provide guidelines that enhance performance while adhering to sustainability and social responsibility principles. These frameworks offer a structured approach that enables organizations to improve efficiency, risk management, and employee well-being, aligning with the objectives of Society 5.0 (Maljugić et al., 2024).

Industry 5.0 also fosters innovation and continuous adaptation, and Quality Management, with its emphasis on ongoing improvement, helps organizations implement innovations effectively (Ali & Johl, 2024). Innovation must go beyond technical efficiency and consider social and cultural impacts. It should be reimagined in light of Society 5.0 principles, which prioritize human well-being and the resolution of social issues. Innovation should not only aim to increase efficiency or productivity but also to enhance quality of life and address social challenges such as inequality and sustainability (Troisi et al., 2023). Synergy occurs when these innovations are applied to create social value and improve quality of life while maintaining high-quality standards.

B. Sustainability & Resilience

Both Society 5.0 and Industry 5.0 emphasize the importance of sustainability. Industry 5.0 must operate within environmental limits, while Society 5.0 aims to create resilient communities that can adapt to changes and crises by utilizing technology to improve people's lives (Saniuk et al., 2022).

Industry 5.0 and Society 5.0 work together to build a more resilient society capable of addressing global crises such as pandemics and climate change. Industrial innovation must be aligned with social needs to ensure that solutions are both effective and sustainable. The synergy between these two spheres is also evident in the necessity of collaboration among different sectors of society, including businesses, governments, and communities. This collaboration is crucial for developing innovative solutions that meet both social and economic needs (Tavares et al., 2022).

Industry 5.0 seeks not only efficiency but also sustainability, and Quality Management can contribute by optimizing processes to reduce waste and improve resource efficiency. Technological advancements require quality professionals to acquire advanced skills, and Quality Management must include training programs to prepare employees for these new challenges (Ali & Johl, 2024). The synergy between Quality 5.0 and Industry 5.0 results in a more efficient, sustainable, and human-centered production system, where quality is integrated into all stages of the production process, benefiting both companies and consumers (Frick & Grudowski, 2023).

Quality 5.0 broadens the interpretation of quality to encompass the satisfaction of all stakeholders, considering social and environmental requirements (Deleryd & Fundin, 2020). This broader vision stems from Society 5.0's commitment to involving all sectors responsible for different aspects of social life (Ülgen & Muneeb, 2024). Society 5.0 must be capable of creating resilient communities that can adapt and respond to crises, such as natural disasters or pandemics. This requires improved data and resource sharing, as well as more effective collaboration across different sectors (Ülgen & Muneeb, 2024).

The interconnection between Industry 5.0, Quality 5.0, and Society 5.0 lies in the fact that sustainable industrial practices cannot be dissociated from social needs. Industry 5.0 provides the necessary tools and technologies to create a productive environment that respects the environment, while Society 5.0 ensures that these innovations are utilized for the well-being of society (Tavares et al., 2022). With a holistic vision that considers not only product quality but also customer satisfaction, social impact, and sustainability, Quality 5.0 aligns with the principles of Industry 5.0, which seeks a balance between industrial production and social and environmental well-being (Frick & Grudowski, 2023). It is important to establish metrics to assess the social impact of Industry 5.0, Society 5.0, and Quality 5.0 initiatives. These metrics may include indicators of well-being, social inclusion, reduction of inequalities, and environmental sustainability. Continuous measurement and evaluation allow for strategic adjustments and ensure that social objectives are met (Mourtzis et al., 2022).

Together, these approaches aim to create a future where economic development does not compromise the health of the planet or people's quality of life, resulting in a truly sustainable development model (Tavares et al., 2022).

C. Technology as an Enabler

The integration of technologies such as Artificial Intelligence (AI), Big Data, and the Internet of Things (IoT) is crucial for the advancement of Society 5.0 and Industry 5.0, as they enable the development of more intelligent systems. The convergence of these technologies creates automatic and adaptive systems that optimize production and services based on real-time data, aiming to enhance quality of life, operational efficiency, reduce waste, and expand access to services (Ülgen & Muneeb, 2024). This technological integration, stemming from both paradigms, is essential for achieving the goals associated with a Super Smart Society and a resilient industrial infrastructure (Ülgen & Muneeb, 2024).

For instance, AI can enhance efficiency in both production and public services through machine learning, pattern recognition, and autonomous decision-making. Big Data enables the collection and analysis of large volumes of data, assisting industries and governments in understanding consumer behavior and identifying areas that require intervention. IoT connects devices and facilitates real-time monitoring of industrial and social processes (Ülgen & Muneeb, 2024).

Society 5.0 reflects changing societal expectations, such as the increasing demand for sustainability, social responsibility, and technological innovation. Industry 5.0, by incorporating these values, enables industries to adapt to these trends through the development of products and services that meet the emerging needs of consumers and society at large. This may include, for example, the creation of technologies that promote health and well-being, or that reduce the carbon footprint of industrial operations (Mourtzis et al., 2022).

Industry 5.0, with its focus on advanced technologies such as AI and Big Data, provides the necessary tools for Society 5.0 in its pursuit of social problem-solving. The fusion between cyberspace and physical space, promoted by Society 5.0, can be further enhanced by technological innovations from Industry 5.0, generating more effective solutions to societal challenges (Huang et al., 2022).

Total Quality Management (TQM) is significantly strengthened by Industry 4.0 technologies, especially IoT and AI. These technologies align with Society 5.0's goals of improving service delivery and everyday life, and they enhance quality processes through real-time monitoring, predictive maintenance, better customer insights, automation, and continuous improvement. This synergy aids in defect elimination and in enhancing service efficiency, thereby enabling organizations to effectively meet the diverse and evolving needs of the population (Arıcı & Kitapci, 2021).

In the context of Quality 5.0, technologies such as Big Data, AI, IoT, Cobots, and Digital Twins drive efficiency, precision, flexibility, and effectiveness (Maljugić et al., 2024), while also supporting the improvement of quality control processes (Frick, 2023). The integration of these technologies aligns quality practices with the human-centered and sustainable principles of Society 5.0 (Maljugić et al., 2024). These tools also assist organizations in collecting and analyzing data more effectively, enabling a better understanding of stakeholder needs and expectations (Deleryd & Fundin, 2020).

Innovation should be a central focus, with an emphasis on technologies that not only increase efficiency but also generate social value. This includes the adoption of emerging technologies to address social challenges such as public health, education, and digital inclusion (Mourtzis et al., 2022).

D. Conceptual Model Proposal

The integration of Industry 5.0 and Society 5.0 concepts can lead to innovations that not only meet market demands but also promote social justice and environmental sustainability (Troisi et al., 2023). Quality, in turn, is driven by continuous improvement and defect elimination. The combination of these approaches can result in innovative solutions that meet societal needs effectively and efficiently (Arıcı & Kitapci, 2021).

To ensure the application and advancement of technologies that are intended to serve as tools for enhancing quality of life, sustainability, and addressing social and environmental challenges, investment in education and training is essential. Society 5.0 and Industry 5.0 require a skilled and educated workforce capable of managing new technologies and approaches. Quality in education and professional training is critical to ensure that individuals can adapt and thrive in this new environment. This implies a commitment to quality at all levels of education and professional development (Arıcı & Kitapci, 2021).

For the present study, we developed a conceptual model that represents the synergy among the three concepts. The creation of a conceptual model is important, as it simplifies communication and provides a streamlined representation of the key elements of an idea or process. A conceptual model is a graphical and structured representation used during the requirements analysis phase of information systems. It serves purposes such as documenting business processes, improving business operations, software development, and database management (Davies et al., 2005). The employment of conceptual frameworks demonstrates its efficacy as a robust instrument for the comprehension and articulation of intricate domains by delineating unequivocally the entities implicated and the interrelations that exist among them (Pastor et al., 2021). Conceptual models enhance organizational communication among stakeholders, clarify ambiguities, identify gaps in requirements, and promote a shared understanding of a system's or process's objectives. Furthermore, a well-developed conceptual model facilitates system maintenance, evolution, and interoperability, thereby improving organizational knowledge management (Mineau et al., 2000).

Figure 3 presents the Conceptual Model of the synergy among the three concepts. This model is structured as follows: at the center, the concepts of Society 5.0, Industry 5.0, and Quality 5.0 are represented, along with the values resulting from their synergy: Human-Centricity, Sustainability & Resilience, and Technology & Innovation. The second circle highlights the objectives or necessary characteristics for this synergy to occur. The synergy among these three concepts aims to improve ecological, economic, and social aspects by enhancing quality of life, making society increasingly intelligent through the promotion of education and human value. The emergence of new technologies, along with those applied in Industry 4.0, will support the achievement of these goals, with a stronger focus on human well-being, the creation of new jobs, and the development of products and services that address human needs without neglecting profitability. Resilience must be embedded in all organizations and communities to ensure preparedness for future adversities, and the Sustainable Development Goals (SDGs) must be considered and fulfilled to reduce resource consumption, minimize the carbon footprint, and uphold social responsibility. The overarching goal of this synergy is the Creation of Value for the betterment of society.

The development of the conceptual model followed an exploratory and integrative approach grounded in the literature on Society 5.0, Industry 5.0, and Quality 5.0. As the objective was to structure and articulate relationships among emerging concepts rather than to test causal relationships, the model was not subjected to empirical validation. Instead, its conceptual consistency, coherence, and applicability were examined through expert assessment involving specialists familiar with the thematic domains addressed in the study. This procedure supported the refinement of the framework and the evaluation of its logical adequacy within the scope of the proposed research design.

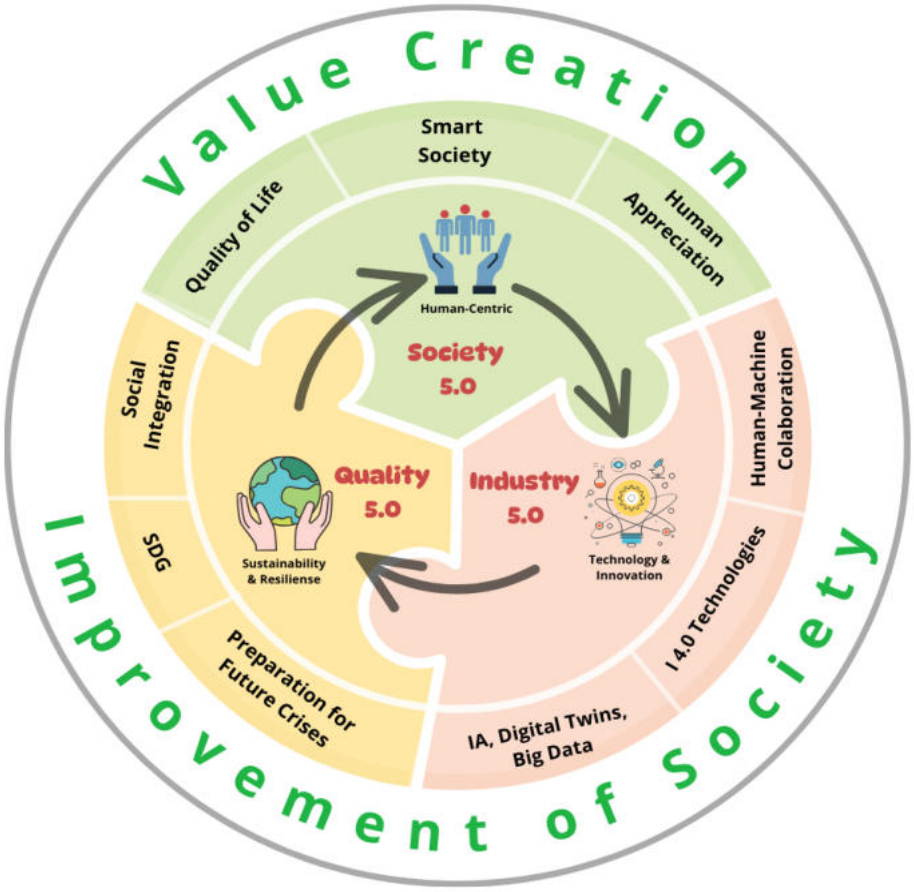


Fig. 13. Conceptual Model of the Synergy between the Concepts

XLVIII. Challenges and Opportunities in Implementing Synergies

Society 5.0 faces several challenges regarding the integration of advanced technologies, as the effective application of tools such as AI and Big Data requires robust infrastructure and the resolution of technological constraints (Onday, 2019). In the domain of social inequality, it is imperative to ensure equitable access to technological benefits across all segments of society, in order to prevent the widening of social disparities (Bölükbaşı, 2021). About sustainability, it is essential to reconcile economic growth with environmental management and social inclusion, advocating for sustainable methodologies that meet current demands without compromising the needs of future generations (Onday, 2019). In terms of cultivating an innovation culture, it is necessary to foster an environment that prioritizes innovation and collaboration across different sectors of society—a task that can prove challenging in more traditional contexts (Holroyd, 2022).

Industry 5.0 faces its own set of challenges, including a shortage of skilled labour, as the demand for a workforce proficient in advanced technologies entails substantial investment in training, resources, and time before yielding tangible results (Adel, 2022). The adoption of new technologies may encounter technical obstacles, such as issues of interoperability and cybersecurity. Additionally, a paradigm shift is required, demanding organizational mindsets to evolve and prioritize human-machine collaboration and process personalization. The need for resilience in times of crisis has also become evident—for instance, the COVID-19 pandemic exposed vulnerabilities in supply chains, emphasizing the necessity for industries to develop greater resilience to future disruptions (Xu et al., 2021).

Quality 5.0 introduces further challenges, such as the adaptation of management systems. There is a need to update Quality Management Systems (QMS) to incorporate approaches centered on human values and sustainability (Maljugić et al., 2024). Stakeholder engagement is also crucial, ensuring that all relevant parties—including employees and customers—are actively involved and their needs effectively addressed (Deleryd & Fundin, 2020). Furthermore, it is essential to foster an organizational culture that values continuous evaluation and improvement, which may be difficult to establish in more rigid institutional environments (Djordjevic et al., 2024). Finally, ethical and privacy concerns must be addressed, as the increasing reliance on advanced technologies raises critical issues such as data privacy and cybersecurity. These concerns must be resolved to earn and maintain consumer trust (Ali & Johl, 2024).

The opportunities associated with the concepts of Society 5.0, Industry 5.0, and Quality 5.0 are equally significant and reflect the transformative and innovative potential these paradigms present. Notable opportunities include the creation of new products and services (Onday, 2019), the emergence of a favourable environment for innovation (Narvaez Rojas et al., 2021),

improvement in quality of life (Djordjevic et al., 2024), the development of personalized products and services (Grabowska et al., 2022), and the establishment of intelligent infrastructures that are more efficient and sustainable. Further opportunities include alignment with the Sustainable Development Goals (SDGs) (Onday, 2019), enhanced access to technology (Narvaez Rojas et al., 2021), cross-sectoral cooperation (Ülgen & Muneeb, 2024), skills development (Narvaez Rojas et al., 2021), and preparedness for future crises, with an emphasis on resilience in both organizations and society (Onday, 2019). By seizing these opportunities, it is possible not only to confront current challenges but also to build a more prosperous and sustainable future for all.

XLIX. Conclusion

This study has demonstrated that the synergy between Society 5.0, Industry 5.0, and Quality 5.0 is essential for the development of a more sustainable and balanced future. These concepts are not isolated but rather complementary, as they all share the common goal of promoting sustainable, human-centered development.

Considering the three research questions, the model was able to address each one. The first research question was addressed through the development of a conceptual model integrating Society 5.0, Industry 5.0, and Quality 5.0 within a common framework oriented toward human-centered, sustainable, and resilient transformation. The model demonstrated that these concepts can be articulated in a complementary and interdependent manner rather than as isolated paradigms. The second research question was addressed by identifying, through the literature review, the main conceptual relationships among the three dimensions, particularly the convergence around digital transformation, human-centricity, sustainability, resilience, innovation, and value creation for society. The analysis also highlighted that Society 5.0 and Industry 5.0 share common objectives, while Quality 5.0 contributes as a managerial and integrative dimension capable of operationalizing these principles. Finally, the third research question was addressed by demonstrating that quality management plays a strategic role in aligning technological innovation with societal well-being and sustainable development. In the context of Era 5.0, Quality 5.0 was identified as an enabling element capable of supporting ethical, human-centered, and socially oriented organizational transformation.

Both Society 5.0 and Industry 5.0 advocate a human-centric approach, leveraging technology not only to increase efficiency but also to enhance quality of life and address social challenges. Industry 5.0 can lead to the creation of new jobs and the development of skills by promoting a more responsible stance on sustainability and organizational culture. It complements Industry 4.0 and represents an opportunity to foster technological advancements that support improved quality of life, social inclusion, and equity, while offering competitive advantages for companies that adopt socially responsible practices.

Quality 5.0, in turn, embraces a proactive management approach that goes beyond customer satisfaction by prioritizing social and environmental needs to generate positive outcomes for all stakeholders. It represents a significant evolution in quality management practices by integrating digital technologies with human elements. This concept aims to promote more effective quality control and a comprehensive approach where sustainability and well-being are central, emphasizing the importance of human values and sustainable practices in generating positive social impact.

Sustainability, human-centricity, and digital innovation represent the core values of this synergy and suggest that industrial practices must evolve to address issues such as climate change. This work also highlights the importance of cross-sector collaboration and inclusive stakeholder engagement through open dialogue to achieve the Sustainable Development Goals (SDGs).

The synergy among these concepts can support not only industries but also governments, which can benefit from their practices and insights in working toward the SDGs—objectives that are crucial for our collective future. The adoption of these concepts requires a cultural shift within organizations, where responsibilities must be redefined to prioritize employee well-being. The integration of these three pillars is fundamental to building a more equitable and sustainable future, in which technology serves social well-being.

However, as demonstrated throughout this study, several challenges must be overcome. The implementation of Society 5.0 in Japan, even if successful, does not guarantee the same outcomes in other nations. Social, economic, and cultural conditions may directly influence different contexts. Industry 5.0 and Quality 5.0 will also face numerous barriers. The technological reality of many industries, as well as their concern for environmental and human factors, is still far behind, and the cost for industries to evolve in these areas is substantial. Investing in cutting-edge technology, in worker knowledge and well-being, and in environmental sustainability is not always a priority for companies and business leaders.

It should be noted that the proposed conceptual model was not empirically validated. Rather, its consistency, coherence, and relevance were assessed through the evaluation of specialists with expertise in the fields of Quality Management, Industry 5.0, and Society 5.0. Given the exploratory and conceptual nature of this study, expert assessment was considered an appropriate approach to examine the logical structure and integrative capacity of the framework. Nevertheless, this limitation also represents an opportunity for future research, particularly through empirical applications capable of testing the model in different organizational and societal contexts.

Regarding further future research, it is recommended that further studies be conducted, particularly to investigate the practical challenges of implementing Industry 5.0 and Quality 5.0 in industrial settings, considering that many companies remain distant from or unaware of these concepts. Moreover, there is a scarcity of studies that analyses the practical results of Society 5.0 implementation in Japan and, especially, in other countries. Additional research supported by statistical data, real-

world examples, and financial results of implementation would help make the topic more tangible and less utopian—especially in relation to the SDGs.

References

Adel, A. (2022) 'Future of Industry 5.0 in society: human-centric solutions, challenges and prospective research areas', *Journal of Cloud Computing*, 11(1). Available at: https://doi.org/10.1186/s13677-022-00314-5.

Barata, J. and Kayser, I. (2023) 'Industry 5.0 – past, present, and near future', *Procedia Computer Science*, 219, pp. 778–788. Available at: https://doi.org/10.1016/j.procs.2023.01.351.

Bölükbaşı, F. (2021) *Revolution and Society 5.0: Japanese human-centric approach and sectoral changes*.

Carayannis, E.G. and Morawska-Jancelewicz, J. (2022) 'The futures of Europe: Society 5.0 and Industry 5.0 as driving forces of future universities', *Journal of the Knowledge Economy*, 13(4), pp. 3445–3471. Available at: https://doi.org/10.1007/s13132-021-00854-2.

Davies, I., Green, P., Rosemann, M., Indulska, M. and Gallo, S. (2006) 'How do practitioners use conceptual modeling in practice?', *Data & Knowledge Engineering*, 58(3), pp. 358–380. Available at: https://doi.org/10.1016/j.datak.2005.07.007.

Deleryd, M. and Fundin, A. (2020) 'Towards societal satisfaction in a fifth generation of quality – the sustainability model', *Total Quality Management & Business Excellence*. Available at: https://doi.org/10.1080/14783363.2020.1864214.

Djordjevic, A., Stefanovic, M., Wawak, S., Popkova, E., Eric, M., Mitrovic, S. and Djordjevic, M.Z. (2024) 'From Quality 4.0 to Quality 5.0 – the transition roadmap', *International Journal for Quality Research*, 18(4), pp. 1–12. Available at: https://doi.org/10.24874/IJQR18.04-18.

Breque, M., De Nul, L. and Petridis, A. (2021) *Industry 5.0: Towards a sustainable, human-centric and resilient European industry*. Luxembourg: Publications Office of the European Union. Available at: https://doi.org/10.2777/308407.

Frick, J. (2023) 'Future of industrial asset management: a synergy of digitalization, digital twins, Maintenance 5.0, Quality 5.0, Industry 5.0 and ISO 55000', *International Journal of Business Marketing and Management*, 8. Available at: https://orcid.org/0000-0002-3204-1574.

Frick, J. and Grudowski, P. (2023) 'Quality 5.0: a paradigm shift towards proactive quality control in Industry 5.0', *International Journal of Business Administration*, 14(2), p. 51. Available at: https://doi.org/10.5430/ijba.v14n2p51.

Fundin, A., Lilja, J., Lagrosen, Y. and Bergquist, B. (2020) 'Quality 2030: quality management for the future', *Total Quality Management & Business Excellence*. Available at: https://doi.org/10.1080/14783363.2020.1863778.

Grabowska, S., Saniuk, S. and Gajdzik, B. (2022) 'Industry 5.0: improving humanization and sustainability of Industry 4.0', *Scientometrics*, 127(6), pp. 3117–3144. Available at: https://doi.org/10.1007/s11192-022-04370-1.

Groumpos, P.P. (2021) 'A critical historical and scientific overview of all industrial revolutions', *IFAC-PapersOnLine*, 54(13), pp. 464–471. Available at: https://doi.org/10.1016/j.ifacol.2021.10.492.

Hattinger, M. and Stylidis, K. (2023) 'Transforming Quality 4.0 towards resilient Operator 5.0 needs', *Procedia CIRP*, 120, pp. 1600–1605. Available at: https://doi.org/10.1016/j.procir.2023.12.002.

Holroyd, C. (2022) 'Technological innovation and building a "super smart" society: Japan's vision of Society 5.0', *Journal of Asian Public Policy*, 15(1), pp. 18–31. Available at: https://doi.org/10.1080/17516234.2020.1749340.

Huang, S., Wang, B., Li, X., Zheng, P., Mourtzis, D. and Wang, L. (2022) 'Industry 5.0 and Society 5.0 – comparison, complementation and co-evolution', *Journal of Manufacturing Systems*, 64, pp. 424–428. Available at: https://doi.org/10.1016/j.jmsy.2022.07.010.

Maljugić, B., Ćoćkalo, D., Bakator, M. and Stanisavljev, S. (2024) 'The role of the quality management process within Society 5.0', *Societies*, 14(7), p. 111. Available at: https://doi.org/10.3390/soc14070111.

Mineau, G.W., Missaoui, R. and Godinx, R. (2000) 'Conceptual modeling for data and knowledge management', *Data & Knowledge Engineering*, 33(2), pp. 137–168. Available at: https://doi.org/10.1016/S0169-023X(99)00049-X.

Mourtzis, D., Angelopoulos, J. and Panopoulos, N. (2022) 'A literature review of the challenges and opportunities of the transition from Industry 4.0 to Society 5.0', *Energies*, 15(17). Available at: https://doi.org/10.3390/en15176276.

Nahavandi, S. (2019) 'Industry 5.0 – a human-centric solution', *Sustainability*, 11(16), p. 4371. Available at: https://doi.org/10.3390/su11164371.

Narvaez Rojas, C., Alomia Peñafiel, G.A., Loaiza Buitrago, D.F. and Tavera Romero, C.A. (2021) 'Society 5.0: a Japanese concept for a superintelligent society', *Sustainability*, 13(12). Available at: https://doi.org/10.3390/su13126567.

Onday, O. (2019) 'Japan's Society 5.0: going beyond Industry 4.0', *Business and Economics Journal*, 10(2). Available at: https://doi.org/10.4172/2151-6219.1000389.

Pastor, Ó., León, A.P., Reyes, J.F.R., García, A.S. and Casamayor, J.C.R. (2021) 'Using conceptual modeling to improve genome data management', *Briefings in Bioinformatics*, 22(1), pp. 45–54.

Rad, F.F., Oghazi, P., Palmié, M., Chirumalla, K., Pashkevich, N., Patel, P.C. and Sattari, S. (2022) 'Industry 4.0 and supply chain performance: a systematic literature review of the benefits, challenges, and critical success factors of 11 core technologies', *Industrial Marketing Management*. Available at: https://doi.org/10.1016/j.indmarman.2022.06.009.

Raja Santhi, A. and Muthuswamy, P. (2023) 'Industry 5.0 or Industry 4.0S? Introduction to Industry 4.0 and a peek into the prospective Industry 5.0 technologies', *International Journal on Interactive Design and Manufacturing*. Available at: https://doi.org/10.1007/s12008-023-01217-8.

Saniuk, S., Grabowska, S. and Straka, M. (2022) 'Identification of social and economic expectations: contextual reasons for the transformation process of Industry 4.0 into the Industry 5.0 concept', *Sustainability*, 14(3). Available at: https://doi.org/10.3390/su14031391.

Saraiva, P. (2023) 'Quality 5.0 in turbulent times', *Quality Progress*, pp. 18–25.

Tavares, M.C., Azevedo, G. and Marques, R.P. (2022) 'The challenges and opportunities of Era 5.0 for a more humanistic and sustainable society – a literature review', *Societies*, 12(6). Available at: https://doi.org/10.3390/soc12060149.

Troisi, O., Visvizi, A. and Grimaldi, M. (2024) 'Rethinking innovation through Industry and Society 5.0 paradigms: a multileveled approach for management and policy-making', *European Journal of Innovation Management*, 27(9), pp. 22–51. Available at: https://doi.org/10.1108/EJIM-08-2023-0659.

Ülgen, S. and Muneeb, M. (2024) 'The synergy of Society 5.0 and Industry 5.0 transforming work, society, and collaboration', *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pp. 2126–2132. Available at: https://doi.org/10.46254/an14.20240564.

Xu, X., Lu, Y., Vogel-Heuser, B. and Wang, L. (2021) 'Industry 4.0 and Industry 5.0 – inception, conception and perception', *Journal of Manufacturing Systems*, 61, pp. 530–535. Available at: https://doi.org/10.1016/j.jmsy.2021.10.006.

7TH International Conference on Quality Engineering and Management

Universidade NOVA de Lisboa, Portugal