

Multi-Criteria Decision-Making (MCDM) Method for Quality Service Assessment: A Systematic Literature Review

Omar Al-Tarawneh¹, Marwan Alseid², Mejhem Altarawneh³, and Serein Al-Ratrout⁴

^{1,2,3} Department of Software Engineering, College of Information Technology, Amman Arab University, Amman, Jordan.

⁴ College of Computing, Michigan Technological University, Michigan, USA.

* Corresponding author: o.husain@aau.edu.jo. Contributing authors: m.alseid@aau.edu.jo; m.tarawneh@aau.edu.jo; salratro@mtu.edu.

Received: 17 May 2026 **Accepted:** 31 August 2026 **Published:** 7 October 2026

DOI: 10.15849/ijasca.119 **License:** Open access under CC BY 4.0. Authors retain copyright.

Publisher: Al-Zaytoonah University of Jordan (ZUJ), Amman, Jordan.

Cite this article: O. Al-Tarawneh, M. Alseid, M. Altarawneh, and S. Al-Ratrout, "Multi-criteria decision-making (MCDM) method for quality service assessment: a systematic literature review," *Int. J. Advance Soft Compu. Appl.*, vol. 18, no. 3, pp. 374–393, 2026, doi: <https://doi.org/10.15849/ijasca.119>.

ABSTRACT

The increasing complexity of digital service ecosystems has led to an increase in the need for reliable Quality Service Assessment Methods (QSAMs) that can assess systems from both technical and experiential perspectives. The current research study demonstrates a systematic literature review (SLR) of QSAMs, which includes Multi-Criteria Decision-Making (MCDM) methods, with guidance by the Evaluation Theory and the PRISMA 2020 protocol. In this regard, around Seventy-eight peer-reviewed studies were published between the period of 2010 and 2025 and were analyzed across seven Evaluation Theory dimensions, which are evaluation criteria, target, measurement scale, data-gathering method, synthesis technique, evaluation process, and validation approach. According to the review, the MCDM techniques — primarily AHP, TOPSIS, Fuzzy MCDM, and hybrid combinations — Improved both the clarity and consistency of QSAM decision-making. Hence, this methodological sophistication has not been matched by equal attention to validation, sensitivity testing, and evaluation-process design. Building on that, the current research paper suggests a set of practical guidelines for sensitivity analysis in QSAMs and outlines a conceptual model, AET-QSAM, in an attempt to address the validation and process gaps identified. In addition, the current paper aims to contribute to a theory-grounded framework for diagnosing the completeness of existing QSAMs and a research agenda for developing more transparent, trustworthy, and adaptable assessment methods for future e-services.

Keywords: e-service quality evaluation; evaluation theory; quality service assessment method; systematic literature review; multi-criteria decision-making.

1 Introduction

Electronic services, also referred to as e-services, have become a fundamental part of how people carry out different financial, governmental, educational, as well as business operations since they started being widely used 10 years ago. These days, organizations depend a lot on several kinds of services like e-banking, e-government, e-learning, and e-commerce services through which they deliver continuous, high-quality, and user-satisfaction access. Such a requirement can be hard to meet in an environment where things are dynamic due to the complicated and multidimensional nature [1], [2].

SERVQUAL and SERVPERF, the traditional evaluation models, were designed for physical or hybrid services and failed to capture the elements of interactivity, technical complexity, and user trust that characterize digital ecosystems [3, 4]. As a result, there has been an increasing move towards multi-criteria decision-making (MCDM methods such as AHP, TOPSIS, VIKOR, and DEMATEL for the evaluation of e-services depending on both quantitative and qualitative factors [5, 6]. Thus, these methods allow decision makers to take many performance indicators (e.g., efficiency, usability, reliability, privacy, responsiveness, etc.) into account simultaneously and make decisions based on their priorities more consistently and transparently.

Even as MCDM has been increasingly promoted as a quality evaluation methodology, there has been no uniformity in the literature. Studies typically address individual components such as criterion weighting, ranking schemes, or information sources, rather than the complete evaluation process

integrating conceptual models, data acquisition, synthesis, and validation. [7]. Consequently, there has been no complete realization of how functioning QSAMs are when applied to real e-service scenarios.

This paper endeavors to fulfill this academic need by conducting a systematic literature review (SLR) of MCDM-based QSAMs designed between 2010 and 2025, grounded in Evaluation Theory [8]. There is an integrated template to evaluate the completeness of the evaluation method in Evaluation Theory in seven dimensions: the evaluation objectives, the criterion, the yardstick, the data collection process, the synthesis process, the evaluation process, and the validation process [9]. In this paper, the theoretical approach allows one to reveal the conceptual roots, methodological trends, and research gaps in the field of e-service with respect to MCDM-based QSAMs.

This study makes three main contributions:

1. To synthesize the trends of existing QSAM research based on the Evaluation Theory framework.
2. To highlight the strengths and limitations of the methodology in the current MCDM approaches towards assessing the quality of e-services.
3. To provide a foundation for developing future QSAMs that combine human judgment, analytical methods, and AI-powered decision support to improve the quality and effectiveness of digital services.

The core principles underlying the formulation of evaluation criteria for e-service quality assessment are outlined in Table 1.

Table 1. Principles of Formulated Evaluation Criteria for E-Service Quality Assessment

Principle	Explanation
Understandability	The criteria for evaluation should also be clearly defined and context-sensitive, to enable interpretation by people tasked with making choices [3].
Decomposability	Each major criterion (such as security, usability, and efficiency) can be decomposed into measurable sub-criteria to better encompass the complexities of each corresponding e-service's performance [5].
Reliability	The criteria used should stem from validated standards such as ISO/IEC 25010 and SERVQUAL extensions and should also be empirically or expertly validated [8], [10].
Relevance	The assessment dimension employed should serve both organizational objectives as well as those of users, such as personalization in e-learning and integrity of transactions in e-banking [11].
Measurability	Enabling consistent assessment through MCDM techniques. Each criterion should be supported by specific, replicable indicators [6], [12].

2 Literature Review

2.1 Quality Service Assessment Method (QSAM)

A systematic evaluation approach used to quantify the effectiveness, consistency, and value perceived of digital or electronic services is called a Quality Service Assessment Method (QSAM). In real-world implementation, QSAMs are decision-making tools that assist firms in evaluating the quality of their services, determining areas for improvement, and redistributing resources as required based on customer demands and operating limitations [13].

The assessment of e-services' quality is by nature multifaceted and involves multiple trade-offs between cost-effectiveness and delivery speed or convenient use and safe access, for instance. Overall, because of this inherent nature, Multi-Criteria Decision-Making (MCDM) techniques may be most apt for assessment. Overall, it is possible that MCDM models can find a compromise between evaluating both numerical and non-numerical criteria [14]. Analytic Hierarchy Process (AHP) techniques, TOPSIS,

VIKOR, and fuzzy methods are extensively utilized for applying MCDM in solving a wide range of problems. Each method has its own priority calculations and combination approaches [5].

The general process for QSAM has five major steps:

1. Criteria for evaluation: for instance, trust, responsiveness, and customization.
2. Assign weights and preferences, either driven by expert knowledge or user input.
3. Collect data using surveys, system metrics, and usage analytics.
4. Use of the model for regrouping alternatives, synthesizing, and ranking.
5. Validation and Feedback: Results of validation by expert judgement or statistical judgement [16].

This strategy can be used in the evaluation of e-learning, online banking, health information technology, and government services. Most importantly, the objective is not only concerned with the technical performance but also considers the ability of the system to satisfy the user's expectations, trust, and perceived value [17] [11].

2.2 Evaluation Theory and Its Role in QSAM

This review draws on Evaluation Theory to assess the rigor and completeness of existing QSAMs. Evaluation Theory proposed by Scriven [8] and extended by Shadish [9] contains six key features of a good evaluation process: evaluation target, criteria, yardstick (measurement scale), data collection techniques, synthesis techniques, and evaluation process.

This study analyzes the existing QSAMs and investigates whether these components are clearly defined in these QSAMs, i.e., the evaluation target (e-service type), the criteria (e-service usability, e-service trust), the measurement scale (Likert-scale, ranking function), the data source (user survey, system log), the synthesis method (MCDM models), and the evaluation process. In addition, a 7th dimension validation method is presented to assess the empirical testing and/or application of the methods in the real world.

This framework can be used to inform a deeper understanding of the limitations of existing QSAMs, such as in the criteria validation, data reliability, and sensitivity testing domains, and can provide a foundation for future research focused on developing more adaptive and AI-supported QSAM frameworks.

The ten trustworthiness principles underlying this framework, together with the detailed scoring rubric applied to each principle, are presented in Section 3.4.

2.3 Related Work

E-service quality assessment has remained an important topic for research for more than two decades, increasing along with technological improvements and user expectations. Initial studies were grounded in conceptual frameworks, including SERVQUAL [13] and E-S-QUAL [24], that analyzed the delivery of services by estimating the difference that exists between user expectations and actual perceived user responses to services. Although these frameworks were of great use in understanding conventional services, they lacked the capacity to comprehend the dynamic and complex world of digital services that entail intensive interactions [17].

In addition, the adoption of Multi-Criteria Decision-Making (MCDM) approaches has also kept pace with the complexity that has been evident in the digital world to ensure a systematic assessment. For instance, in assessing the level of service performance in an e-service, the integration of various approaches of MCDM assists in the overall assessment of several conflicting criteria such as responsiveness, usability, and system reliability, although the choice among these approaches remains methodologically challenging [5] [19].

Many systematic review articles attempt to collect and collate with this growing body of literature. For example, the application of MCDM to sustainable service systems was investigated by [14], while the application of MCDM for assessing quality, without any post-decision analysis and robustness checks, was discussed by [15]. More specifically, [16] conducted a systematic review of website quality assessment using MCDM and found that studies typically define overlapping and semantically redundant criteria without examining the interdependencies among them.

However, the prevailing literature only deals with isolated parts like criteria for evaluation or certain algorithms for solving Multi-Criteria Decision-Making problems, instead of covering the whole process from criteria definition, collection, combination, and validation for evaluation. Moreover, little research has attempted to use Evaluation Theory for a systematic analysis related to the fitness and authenticity of the existing Quality Service Assessment Methods (QSAMs).

The most relevant prior literature reviews that have been helpful to the development of QSAMs are summarized in Table 2. The analysis shows that previous studies have given good insights on the service quality framework, while no study thoroughly examined the entire MCDM-based assessment process in the e-service context. This paper, therefore, extends the literature by conducting a more comprehensive and theoretically informed review that embraces seven dimensions of analysis in the context of Evaluation Theory [8] and [9], namely evaluation target, evaluation criteria, evaluation yardstick, data collection, synthesis techniques, evaluation process, and validation approach for validation.

Table 2. Summary of Previous Literature Reviews Related to QSAMs

Study	Type	Focus Area	Objectives	Review Dimensions
Parasuraman et al. (1988) [3]	Conceptual	Service quality measurement	Proposed SERVQUAL as a general framework for evaluating service quality	Tangibles, Reliability, Responsiveness, Assurance, Empathy
Parasuraman et al. (2005) [24]	Empirical	E-S-QUAL framework	Developed and validated the E-S-QUAL multi-item scale for measuring website service quality	Efficiency, Fulfillment, System Availability, Privacy
Khan et al. (2021) [14]	SLR	Sustainable supply chain management	Reviewed MCDM techniques adopted in sustainable supply chain management research	MCDM techniques, Drivers and barriers, Research design
Więckowski and Sałabun (2023) [15]	SLR	Sensitivity analysis in MCDA	Classified sensitivity analysis techniques and frameworks used in MCDA	Sensitivity methods, Robustness testing, ranking stability
Rekik et al. (2018) [16]	SLR	Website quality assessment via MCDM	Identified overlapping criteria and extracted association rules between them using text mining	Criteria extraction, Association rules mining, Web site categories
Dhurkari (2022) [19]	Review	MCDM methodological limitations	Critically reviewed the axiomatic foundations and practical difficulties of existing MCDM methods	Axiomatic foundations, Consistency, Descriptive decision theories
Gefen, Karahanna, and Straub (2003) [21]	Empirical	Trust and TAM in online shopping	Integrated trust beliefs with the Technology Acceptance Model (TAM) to explain online shopping intentions	Trust, perceived usefulness, perceived ease of use

Pamucar et al. (2021) [22]	Empirical	Robust service quality measurement	Proposed a grey SWARA- MARCOS (SW [*] MARCOS-G) model to rank five Spanish airports on 14 service quality criteria, validated via sensitivity analysis of criteria weights and Bonferroni parameters	SWARA weighting, MARCOS ranking, Sensitivity analysis of criteria weights
Tedja et al. (2024) [23]	SLR	Service quality and perceived value	Examined the influence of service quality and perceived value on satisfaction and intention to continue the relationship	Service quality, Perceived value, Satisfaction, Continuance intention
Current Study (2025)	Systematic Literature Review (SLR)	E-service quality evaluation based on MCDM	To evaluate the completeness, validation, and trustworthiness of QSAMs using Evaluation Theory	Target, Criteria, Yardstick, Data Gathering, Synthesis, Process, Validation

The progression in assessing the level of services from conceptual frameworks to computer-based methods using MCDM can be identified from Table 2. SERVQUAL and E-S-QUAL were the initial frameworks, and these frameworks lack rigor in methodology for weight calculations and validation, whereas recent works include multi-criteria decision-making and methods from artificial intelligence for accuracy.

However, gaps can also be identified from the review. The majority of QSAM studies are neither empirically validated nor checked for robustness, making them less generalizable, especially since the domains are various for e-services. Furthermore, there is also a lack of consideration for the principles behind the integration of feedback and sensitivity analysis for improved adaptive services. The study, therefore, presents a review that is both comprehensive and conceptually grounded, thereby filling the identified gaps.

3 Research Methodology

Presented in this section is the methodological framework of the research for the Systematic Literature Review (SLR) on Quality Service Assessment Methods (QSAMs). The review protocol is grounded in SLR guidelines mentioned by Kitchenham and Charters [25] and aligns with the PRISMA 2020 framework guidelines [26] in an attempt to guarantee both transparency and replicability. There are five main stages to this process as follows:

1. Research question formulation
2. Development of search strategy
3. Identification and Selection of Studies
4. Data Extraction and Synthesis
5. Evaluation and validation of findings

The overriding research question for the study is:

“How well do MCDM-based Quality Service Assessment Methods (QSAMs) perform in evaluating the quality of e-services?”

After identifying this problem, it was broken down into several sub-questions. These questions were formed based on the Evaluation Theory [8], [9]. Sub-questions refer to the evaluation object, criteria,

standard, and the method of collecting data. Also involved are methods for synthesizing the evaluation process as well as validation methods.

3.1 Identification of Research Questions

Continuing the theme of Section 3.6, a sequence of seven pivotal research questions (RQs) associated primarily with the evaluation of precision and suitability of existing QSAMs was devised for the purpose of completing the first main objective.

Each question serves the purpose of exploring a distinct aspect of the QSAM design, such as grasping the criteria, sources, and methods used to carry out the synthesis and validation processes in previous research. This breakdown provides a guide that supports the organization of the review and comparison of studies, all done consistently.

3.2 Search Strategy

This review of literature includes the works published over the period between January 2010 and March 2025, within the academic context of QSAM. During that period, works both basic in nature and modern have been looked at to assess e-service from the view of QSAM. We opted for 2010 as a starting year mainly because it is about the time of the implementation of the advanced digital service platforms and cloud-based systems [27], which was considered a major transformation of mindset regarding service quality.

A search was conducted in six major electronic databases:

- Science Direct (Elsevier)
- Springer Link
- IEEE Xplore
- Emerald Insight
- ACM Digital Library
- Google Scholar

Boolean operators and synonyms were used to expand the search. The search terms for evaluation target, criteria, methods, and processes were merged as below:

An e-service quality or an online service assessment (or a combination of the two) was used.

AND (“MCDM” OR “multi-criteria decision-making” OR “AHP” OR “TOPSIS” OR “VIKOR”)

The articles were limited to peer-reviewed journal articles and conference papers in English. Duplicates, editorials, and papers with a focus on SERVQUAL-type conceptual models without multi-criteria framework were not included.

The English-language restriction was applied for practical reasons: the review team's ability to accurately interpret full methodological detail — criteria definitions, scoring formulas, and validation procedures — could not be reliably guaranteed through machine translation of technical decision-analysis text, and English-language coverage is also more consistent across the six databases searched. One can acknowledge, however, that this restriction may include a language bias, particularly under-representing relevant QSAM research published in other languages (e.g., Chinese, Spanish, or Arabic), such as studies situated in non-English-speaking e-service markets. In this article, this is explicitly flagged as a limitation of the present review, and it is recommended that future replications or extensions adopt multilingual search strategies or professional translation support to mitigate this risk.

3.3 Study Selection Process

The initial search returned 1,420 studies. Titles and abstracts were then screened to exclude irrelevant articles, leaving 230 papers. Following the application of the inclusion and exclusion criteria, 78 full-text studies were selected for detailed analysis.

It should be clarified that the final count of 78 studies was not determined by a formal saturation analysis (e.g., iterative forward/backward citation tracking continued until no new eligible studies emerged). Rather, it represents the complete and exhaustive set of studies meeting the inclusion criteria after the PRISMA-based search and screening protocol (Figure 1) was applied across the six databases and date range specified in section 4. This SLR therefore follows the standard systematic-review convention of exhaustive, criteria-bound coverage of the eligible literature, rather than a sample-based saturation approach.

The inclusion criteria were:

1. The study proposed or applied an MCDM-based method for evaluating e-service quality.
2. The method explicitly addressed multi-dimensional assessment criteria.
3. The study provided sufficient methodological detail for replication.

Exclusion criteria included:

1. Studies outside the e-service or digital quality context.
2. Studies that lack any empirical or methodological component.
3. Duplicate publications or incomplete reporting.

3.4 Data Extraction and Synthesis

Data from each of the 78 studies that were identified were compiled using the structured approach to data extraction. The identified documents were then assessed for each of the seven dimensions identified in the Evaluation Theory framework. The process of extracting information helped identify trends, weaknesses, and gaps in the existing QSAM frameworks that were present.

Each of the 78 studies was scored against the ten trustworthiness principles listed in Table 3 (Transparency of Execution, Comprehensive Structure, Mathematical Coherence, Justifiable Axioms, Measurement Validity, Synthesis of Judgements, Ranking Robustness, Handling of Intangibles, Sensitivity Analysis, and Feedback Integration), using the three-point ordinal scale defined in that table (0 = absent/weak, 0.5 = partially present, 1 = fully present/strong). Scoring was carried out by one reviewer. Table 9 in Section 4 reports the resulting aggregate frequencies.

Table 3. Principles for Ensuring the Trustworthiness of MCDM Results in E-Service Quality Assessment

#	Principle	Description	Rating Principle
1	Transparency of Execution (TE)	The MCDM model should have a clearly defined logic that can be easily understood and replicated by decision makers [18].	0 = Opaque; 0.5 = Partially clear; 1 = Fully transparent
2	Comprehensive Structure (CS)	The model should include both breadth (number of criteria) and depth (hierarchical decomposition of sub-criteria) to represent service complexity [6].	0 = Limited; 0.5 = Moderate; 1 = Extensive
3	Mathematical Coherence (MC)	The method should include a formal mathematical framework that justifies weighting, aggregation, and ranking [5].	0 = Weak; 0.5 = Average; 1 = Strong
4	Justifiable Axioms (JA)	The theoretical foundations (e.g., consistency, transitivity, normalization) must be logically justified [19].	0 = Unjustified; 0.5 = Partial; 1 = Fully justified

#	Principle	Description	Rating Principle
5	Measurement Validity (MV)	Scales used (nominal, interval, fuzzy) should align with the nature of data collected and the context of e-service evaluation [10].	0 = Invalid; 0.5 = Partially valid; 1 = Valid
6	Synthesis of Judgements (SJ)	The aggregation process (e.g., weighted sum, fuzzy logic) must combine expert opinions and user data systematically [20].	0 = Weak; 0.5 = Adequate; 1 = Robust
7	Ranking Robustness (RR)	The method should produce consistent rankings even when input data are perturbed or scaled differently [6].	0 = Unstable; 0.5 = Moderate; 1 = Stable
8	Handling of Intangibles (HI)	The framework should be capable of measuring qualitative dimensions such as trust, satisfaction, and perceived value [21].	0 = Ignored; 0.5 = Partial; 1 = Included
9	Sensitivity Analysis (SA)	Variations in input data should be tested to assess model reliability under uncertainty [22].	0 = None; 0.5 = Partial; 1 = Complete
10	Feedback Integration (FI)	The evaluation should include feedback loops for continuous improvement based on user and expert input [34].	0 = None; 0.5 = Limited; 1 = Integrated

Thus, essentially, Table 3 presents a basis for assessing the credibility of QSAMs for e- services that use the MCDM technique. An authentically credible model for evaluating e- services would inevitably cover both the technical and experiential perspectives, apart from having clarity, theoretical soundness, and empirically grounded constructs. For example, the use of fuzzy logic and AHP can improve the reliability and decisiveness of the results, although it still must undergo a sound explanation for its weighting system and appropriate sensitivity analysis to prove its credibility [15], [19]. The integration of actual user inputs can further improve its flexibility and relevance by ensuring that the results remain attuned to the ever-changing expectations and developments in web technology [21], [34].

It should be clarified that the ten trustworthiness principles in Table 3 were derived from, and are grounded in, established Evaluation Theory and MCDM literature [8], [9], [10], [19], rather than validated through a dedicated expert-panel review (e.g., a Delphi study or content-validity assessment) before their application to the 78 reviewed studies. We acknowledge this as a limitation of the present rubric: while its theoretical grounding is well established, its practical comprehensiveness and clarity have not been empirically confirmed by independent domain experts. Formal expert validation of this rubric is included as Stage 1 of the AET-QSAM validation plan proposed in Section 6.

3.5 Comparison with Related Reviews

The study is then contrasted with other important research related to the evaluation of e-service quality, presented in Table 4. This study, unlike other research, fully describes the progression of QSAM through a theory-driven method that outlines seven analytical factors for evaluating its reliability and completeness.

Table 4. Comparison with Most Related Reviews on QSAMs

Study	Search Sources	Papers Reviewed	Focus on MCDM	Time Range	Dimensions Analyzed	Evaluation Theory (ET) Integration
Aazagreyir et al. (2022) [33]	5 databases	60	✓	2010–2021	QoS Factors, Application Domains, Fuzzy MCDM Methods, Sensitivity Analysis	✗
Rekik et al. (2018) [16]	1 database (Scopus)	532	✓	2009–2015	Criteria Extraction, Association Rules Mining, Web Site Categories	✗

Pamucar et al. (2021) [22]	N/A (empirical study)	N/A (5 airports)	✓	N/A	SWARA weighting, MARCOS ranking, Sensitivity analysis of criteria weights	✗
Tedja et al. (2024) [23]	1 database (Scopus)	22	✗	2008–2024	Service Quality, Perceived Value, Satisfaction, Continuance Intention	✗
Current Study (2025)	6 databases	78	✓	2010–2025	Target, Criteria, Yardstick, Data Gathering, Synthesis, Process, Validation	✓

3.6 Summary of Research Questions

The seven research questions of this review are listed in Table 5. Each question is a part of the evaluation theory and thus, all methodological and theoretical aspects related to QSAM development are comprehensively covered in the study.

Table 5. Summary of Research Questions

RQ	Sub-Research Questions
RQ1: What evaluation criteria are used by QSAMs to assess e-service quality?	RQ1.1: Which quality criteria are most frequently addressed? RQ1.2: Are the criteria definitions explicit and context-specific? RQ1.3: Are the criteria decomposed into measurable sub-criteria? RQ1.4: What verification approaches are used to validate criteria?
RQ2: What are the primary targets of QSAMs?	RQ2.1: Which e-service domains (e.g., e-learning, e-banking, e-health) are most frequently studied? RQ2.2: Which stakeholders (users, providers, managers) receive the most attention?
RQ3: What yardsticks or scales are used in QSAMs?	Which measurement systems (Likert, fuzzy, numerical) are applied to capture judgements?
RQ4: What data-gathering techniques are employed?	What sources (surveys, expert input, analytics) are used to collect assessment data?
RQ5: What synthesis techniques are adopted in QSAMs?	Which MCDM-based methods (AHP, TOPSIS, VIKOR, Fuzzy-MCDM) are most common? RQ5.1: How trustworthy are the results produced by these techniques?
RQ6: What evaluation processes are applied?	Do QSAMs follow structured, repeatable assessment workflows or ad hoc procedures?
RQ7: What validation approaches are used?	RQ7.1: What sources and datasets are used for validation? RQ7.2: Are results tested for reliability and sensitivity?

This systematic literature review (SLR) was planned to produce an analytical, clear, and repeatable method, along with full coverage and sufficient depth of analyses. The limitations of both methods can be overcome by combining the knowledge from the Evaluation Theory [8, 9] and the methodological advantages of Kitchenham's SLR process [25].

In contrast to previous research, this study employs a multidimensional approach to assess QSAM performance. The criteria considered are not only the algorithmic efficiency but also the theoretical integrity, validity, and clarity. Furthermore, it intends to provide a clear and consistent foundation for future research efforts and to guide continuous improvement in quality management of e-services using MCDM techniques.

4 Results and Analysis

This section summarizes the findings from the comprehensive review of 78 works on the use of QSAMs that combine MCDM methods to evaluate e-services. The studies have been categorized according to the seven aspects of the evaluation theory. The findings highlight the growing preference for frameworks based on data, users, and AI technologies for decision-making in e-service evaluation. However, several challenges remain, especially in terms of sensitivity analysis, validation procedures, and empirical testing of such frameworks.

The results are presented according to the seven research questions (RQ1–RQ7), covered in Sections 4.1 to 4.7. Table 10 at the end of this section summarizes the main findings for each research question.

Search and Selection Overview

The initial screening identified a total of 1,420 publications. These were progressively filtered through different stages by removing duplicates, non-English publications, and studies not pertinent to the research questions. As a result, the final set of documents left for screening at the abstract stage was 230, of which, 78 articles met the inclusion criteria and were carried forward for detailed assessment.

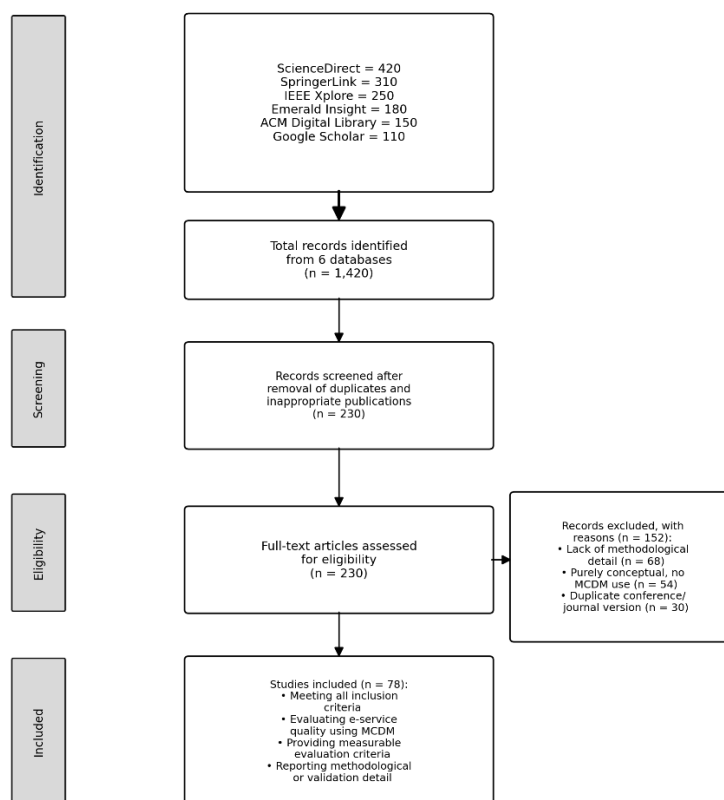


Figure 1. PRISMA flow diagram (Source: Adapted from Page et al. [26])

Figure 1 summarizes the selection process, while Table 6 presents the detailed search string used across databases.

Table 6. Search String and Databases Applied

Database	Search String Applied	Results Retrieved	After Screening
ScienceDirect	("e-service quality" OR "digital service assessment") AND ("MCDM" OR "AHP" OR "TOPSIS" OR "VIKOR") AND ("evaluation" OR "ranking")	420	28
SpringerLink	("service quality" AND "multi-criteria decision-making")	310	15
IEEE Xplore	("quality assessment" AND "decision-making")	250	14
Emerald Insight	("digital service quality" AND "evaluation methods")	180	8
ACM Digital Library	("user satisfaction" AND "multi-criteria evaluation")	150	6
Google Scholar	("MCDM service quality evaluation")	110	7
Total	—	1,420	78

4.1 Answering RQ1: Evaluation Criteria Used in QSAMs

Criteria specification is an important feature when it comes to QSAM design. The study identified 92 different criteria from the reviewed literature, grouped under seven theme dimensions, namely usability, efficiency, reliability, trust and security, responsiveness, perceived value, and innovation and personalization (see Table 7).

The criteria that were most frequently cited were related to user experience, namely, usability and trust, emphasizing the shift in evaluation from technology to user experience in modern e-service systems.

Table 7. Major Evaluation Criteria Used in QSAMs

Dimension	Representative Criteria	Frequency of Criteria
Usability	Interface clarity, Ease of use, Navigation simplicity	18
Efficiency	Response time, System speed, Resource utilization	14
Reliability	Accuracy, Availability, Error-free operation	13
Trust & Security	Data privacy, Confidentiality, Integrity, Transparency	17
Responsiveness	Feedback timeliness, Support quality, Adaptiveness	11
Perceived Value	Cost-benefit, Satisfaction, Willingness to reuse	10
Innovation & Personalization	AI-driven recommendations, Service customization	9
Total	—	92

Many researchers used a complete set of evaluation criteria, but few of them have been validated by testing or experts' opinions. Most studies relied mainly on existing literature to establish their evaluation criteria. Fewer studies used formal expert validation, and only a small number applied more rigorous confirmatory techniques, such as confirmatory factor analysis or inter-rater reliability testing.

Overall, these findings provide a direct answer to RQ1. The reviewed QSAMs use a wide range of criteria, but fewer than a quarter have undergone empirical or expert validation. This suggests that the development of comprehensive criteria has progressed faster than their verification.

4.2 Answering RQ2: Primary Targets of QSAMs

The 78 studies were classified based on the type of methodology used, data source, area of application, and validation approach (see Table 8). The main areas of application of the studies covered were e-Learning (28%), E-banking (24%), and E-Government (17%). Most of the studies used either AHP or

Fuzzy TOPSIS. This shows that a methodology that presented a balance between accuracy and understanding was much preferred in decision-making [15, 16].

Table 8. Classification and Categorization of Selected Studies

Dimension	Categories (number of reviewed studies)	Total
Application Domain	E-learning (22), E-banking (19), E-government (13), E-commerce (10), E-health (8), Others (6)	78
MCDM Technique Used	AHP (25), TOPSIS (18), VIKOR (10), Fuzzy MCDM (15), Hybrid (AHP–TOPSIS/VIKOR) (10)	78
Data Source	Expert Judgement (32), User Survey (25), System Metrics (12), Mixed Sources (9)	78
Validation Approach	None (34), Expert Verification (20), Empirical Testing (15), Simulation (9)	78

Regarding RQ2.2, Table 8 shows that studies are classified mainly by application domain; explicit stakeholder-level classification (users, providers, or managers) was rarely reported, which is itself an identified gap for future QSAM research.

4.3 Answering RQ3: Yardsticks and Measurement Scales

In response to RQ3, the reviewed QSAMs predominantly rely on 5- or 7-point Likert-type scales to capture expert or user judgements, with a smaller subset (mainly the Fuzzy-MCDM and Hybrid studies) adopting fuzzy linguistic scales to represent uncertainty in the assessment. Very few studies, however, reported testing the sensitivity of results to the choice of scale, which is reflected in the low Sensitivity Analysis (SA) coverage of 32% in Table 9.

4.4 Answering RQ4: Data-Gathering Techniques

For RQ4, Table 8 indicates that data for QSAMs is gathered primarily through expert judgement (32) and user surveys (25), while system-generated metrics (12) and mixed sources (9) remain comparatively underused. This suggests that QSAMs still depend heavily on subjective, one-time data collection rather than continuous, system-level analytics.

4.5 Answering RQ5: Synthesis Techniques and Their Trustworthiness

The average methodological completeness across the ten trustworthiness principles in Table 9 (calculated as the mean of the ten individual percentages) was 50%, indicating a medium level of theoretically and empirically advanced maturity. While most studies have focused on technical expertise in the application of MCDM algorithms, relatively little attention has been given to user feedback, evaluation, and learning. These elements are important for developing trustworthy frameworks for assessing e-services.

However, several of the reviewed studies tend to use a combination of AHP and Fuzzy TOPSIS for weighting and ranking. However, this review did not extract a direct comparative metric to establish whether this combination provides better coherence or validation than other approaches. For RQ5, Table 8 shows that AHP (25) and TOPSIS (18) are the most used synthesis techniques, followed by Fuzzy MCDM (15) and AHP-TOPSIS/VIKOR hybrid methods (10). This indicates a preference for methods that are relatively easy to apply computationally and straightforward to explain to decision makers.

All the reviewed studies were judged on 10 principles of trustworthiness adapted from Table 3 to assess the overall methodological robustness. The results (Table 9) indicate that transparency, coherence, and measurement validity scores were relatively high, while feedback integration and sensitivity analysis were the least covered parts of the development of QSAM.

Table 9. Trustworthiness Evidence of Proposed Evaluation Criteria in QSAMs

Principle	Frequency (out of 78)	% of Studies
Transparency of Execution (TE)	55	71 %
Comprehensive Structure (CS)	47	60 %
Mathematical Coherence (MC)	51	65 %

Justifiable Axioms (JA)	34	44 %
Measurement Validity (MV)	49	63 %
Synthesis of Judgements (SJ)	41	53 %
Ranking Robustness (RR)	39	50 %
Handling of Intangibles (HI)	33	42 %
Sensitivity Analysis (SA)	25	32 %
Feedback Integration (FI)	19	24 %

4.6 Answering RQ6: Evaluation Processes

With respect to RQ6, the reviewed QSAMs are, in the large majority, applied as manual, one-time assessments rather than as structured, repeatable workflows: none of the 78 studies described an automated or continuously updated evaluation cycle. This is consistent with the low Feedback Integration (FI) score of 24% reported in Table 9 and indicates that QSAM evaluation is still largely treated as a static exercise rather than an ongoing process.

4.7 Answering RQ7: Validation Approaches

Finally, for RQ7, Table 8 shows that 34 of the reviewed studies did not report any formal validation of their QSAM, while 20 relied on expert verification, 15 used empirical testing, and only 9 applied simulations. Among the studies that included empirical testing, only a few reported using rigorous confirmatory techniques, such as confirmatory factor analysis or inter-rater reliability testing. These findings suggest that validation, especially reliability and sensitivity testing, remains one of the weakest stages in the QSAM development cycle.

Cross-Question Summary

The table below (Table 10) gives a summary of the findings reported across Sections 4.1–4.7, mapping each research question to its corresponding evaluation dimension, main finding, and identified gap.

Table 10. Summary Table Mapping RQ1–RQ7 to Key Findings and Identified Gaps

RQ	Evaluation Dimension	Main Findings	Identified Gaps
RQ2	Evaluation Target	Most QSAMs focus on single-sector e-services (e-learning, e-banking).	Limited cross-domain validation.
RQ1	Evaluation Criteria	Comprehensive but rarely verified empirically.	Lack of standardization and expert validation.
RQ3	Yardstick / Measurement	Preference for 5- or 7-point Likert or fuzzy scales.	Limited scale sensitivity testing.
RQ4	Data-Gathering Techniques	Primarily surveys and expert inputs.	Low integration of real-time system analytics.
RQ5	Synthesis Techniques	AHP, TOPSIS, and Fuzzy MCDM dominate.	Few adaptive or hybrid models with AI/ML.
RQ6	Evaluation Process	Often manual and one-time.	Absence of dynamic feedback loops.
RQ7	Validation Approach	Mostly theoretical; empirical validation rare.	Need for CFA, simulation, and multi-user testing.

The outcomes indicate that there is a shift from concept-based studies to modeling research in QSAM, but it lacks methodology integration. The majority perceive evaluation as a static process when, in reality, it is dynamic, varying depending on user activity, system updates, and other factors that alter e-service quality.

Therefore, it can be inferred that the integration process for analytics using AI and user feedback mechanisms will constitute the future direction for QSAM. The future direction for research will relate to both accuracy in the calculation process for MCDM and, more importantly, its interpretation through mechanisms that ensure the measurement of digital services' quality will form a 'living' process, and not a 'snapshot,' concept [28, 29].

5 Discussion and Implications

The analysis of 78 studies on Quality Service Assessment Methods (QSAMs) that apply Multi-Criteria Decision-Making (MCDM) has important implications for understanding how the process has been used for assessing the quality of e-services. This section will explore the implications of the obtained results.

5.1 Evolution of QSAM Research

The outcome clearly reflects an evolutionary process for QSAM research during the past decade.

The initial attempt to conduct research has traditionally involved the use of SERVQUAL and E-SQUAL frameworks that relied mainly on user perspectives, mainly through dimensions such as tangibles, reliability, and responsiveness [3], as well as efficiency, fulfillment, system availability, and privacy [24]. Although the conceptual bases were good, the conceptual structures did not include attributes that started to describe what these environments would be like, including such attributes as system security, flexibility, and user interface. A milestone was the introduction of MCDM approaches [36]. AHP, TOPSIS, and VIKOR provided an organized approach to using logical weightings and balancing multiple evaluation dimensions [15, 5]. This change enabled the use of measurable, repeatable decision systems rather than one-dimensional satisfaction questionnaires by academics and practitioners.

As the use of e-services is becoming more sophisticated and user-centric, it will be increasingly adopted fuzzy and hybrid approaches which will be able to process uncertain and judgmental information effectively, while multi-combined methods like AHP-TOPSIS and AHP-VIKOR will be used to handle such information. This will enable, for instance, the user's perception- the level of security that a payment gateway has, the credibility that an e-Learning website has, etc.- to pass through mathematical models.

But there are lingering weaknesses, too. Most studies correctly identify the criteria used for the evaluation and apply the appropriate model to make decisions, but fewer studies are on the topics of feedback mechanisms, empirical testing, and applicability to other domains. This means that many QSAMs still have limited scope in guiding decision-making in various e-service sectors and are context-specific and non-adaptive [28] [29]. This result is consistent with the conclusions of Alabool et al. [35], which suggest that methodological completeness and not algorithmic sophistication is the key to long-term credibility of an evaluation method.

5.2 Theoretical Implications

From a theoretical standpoint, this review extends Evaluation Theory [8], [9] to the domain of e-service quality. Evaluation Theory proposes that a sound assessment framework must clearly specify seven interdependent components:

- (1) Evaluation target,
- (2) Evaluation criteria,
- (3) Yardstick or measurement scale,
- (4) Data-gathering method,
- (5) Synthesis technique,
- (6) Evaluation process, and
- (7) Validation method.

Results portray imbalanced consideration for various factors. The majority of the QSAMs exert intense efforts on formulating criteria and choosing an MCDM methodology, without consideration for the rationale for using a 5-point scale instead of a fuzzy interval, reliability of the data, and validation steps.

Few studies explain how findings are validated through expert judgement, statistical analysis, or longitudinal analysis. With the application of the structure for Evaluation Theory in its entirety, it is

possible to diagnose any quality assessment model for its completeness. It also proves that it is necessary to integrate both analytical and human elements in a theoretically well-developed QSAM, aimed at ensuring that any decisions arrived at using the concept, MCDM, must exhibit clarity.

The Evaluation Theory framework adopted here also maps meaningfully onto established software- and service-quality standards. ISO/IEC 25010 (already cited as a source of validated criteria in Table 1) defines a product quality model of eight characteristics — functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability. The evaluation criteria synthesized in Table 7 (usability, efficiency, reliability, trust and security, responsiveness, perceived value, and innovation) correspond closely to this model, confirming that the criteria dimension of QSAMs (RQ1) is broadly consistent with ISO/IEC 25010, while extending it with e-service-specific constructs (e.g., perceived value, innovation) not explicitly covered by the standard. ISO/IEC 25040, in turn, defines a generic quality-evaluation process (establish evaluation requirements, specify the evaluation, design the evaluation, and execute the evaluation), which parallels the evaluation-process dimension identified under RQ6.

The reviewed QSAMs extend beyond these standards in two respects:

1. Multi-criteria synthesis: neither ISO/IEC 25010 nor 25040 prescribes how to aggregate or prioritize competing quality characteristics; the MCDM techniques identified under RQ5 (AHP, TOPSIS, Fuzzy MCDM) fill this gap by providing explicit weighting and ranking mechanisms.
2. Trustworthiness and sensitivity reporting: the 10-principles rubric in Table 3 and the sensitivity-analysis guidelines proposed in Section 5.3 extend beyond the explicit requirements of ISO/IEC 25040. They provide a more detailed basis for assessing the robustness of evaluation results, rather than focusing only on compliance with the evaluation process.

We therefore view QSAMs, together with the Evaluation Theory framework used to analyze them, as complementary to ISO/IEC 25010 and ISO/IEC 25040. They help operationalize and extend the quality characteristics and process structure defined by these standards by incorporating explicit multi-criteria synthesis and trustworthiness assessment.

5.3 Methodological Insights

Regarding the methodology, the review outlines a clear shift from the initial use of one-stage evaluation to a more comprehensive assessment process. Early studies have used MCDM as a static decision tool for opinion gathering, weighting, and ranking.

Recent research, on the other hand, views the process of appraisal differently – as one that is continuous, with inputs from users, as well as analysis of system traces and data.

It would seem to be the natural course of evolution for man and machine to collaborate in making assessments about decisions, in the spirit of increasing collaboration between humans and AI systems in decision making [31]. The use of an AI analysis tool to assist in the context of MCDM enables the inclusion of behavioral data such as click streams and usage statistics in the QSAM tools and methods.

Despite the presence of developments, however, the review reveals some flaws in methodology, which were also discussed in the study on the evaluation of clouds [35]:

1. Limited sensitivity analysis: only about one-third of studies test how input variation affects ranking stability.
2. Inconsistent weighting schemes: subjective and objective weights are rarely compared.
3. Lack of empirical replication: most evaluations are one-off case studies rather than multi-context experiments.

The low uptake of sensitivity analysis (32% of studies, Table 9) is unlikely to reflect a deliberate methodological choice, but rather a combination of practical barriers:

1. Limited awareness: many QSAM authors adopt AHP, TOPSIS, or Fuzzy MCDM as ready-made tools without formal training in decision-analysis practice, where sensitivity testing is a standard step.
2. Complexity of perception: methods such as weight perturbation, Monte Carlo analysis, and global sensitivity indices require additional statistical knowledge and computing tools that are needed for basic MCDM.
3. Data availability: Sensitivity testing often requires repeated expert elicitation under different weighting scenarios or access to raw, disaggregated judgment data. However, many studies do not collect or report this level of data.
4. Lack of guidelines for reporting: In contrast to research areas like clinical and environmental decision analysis, no sensitivity analysis is required in e-service quality journals as a necessity for the methodology to be rigorous.

Towards standardization of sensitivity analysis in QSAM and to bridge this gap, we recommend the following tiered approach for future QSAM work:

1. Baseline (minimal) practice: provide a one-at-a-time (OAT) weight sensitivity test, where each criterion weight is perturbed by a constant amount (e.g., $\pm 10\%$), and the results show whether the top-ranked alternatives switch.
2. Ranking stability reporting: always includes a measure of the stability of the ranking procedure, for example, the Spearman rank correlation coefficient between the initial ranking and other rankings resulting from OAT perturbations of the weights.
3. Model-specific analyses: if the QSAM uses AHP, provide the consistency ratio along with the weight-interval sensitivity test; for fuzzy logic models, test the sensitivity to the choice of membership function or defuzzification procedure.
4. Advanced practice in case of high stakes: when QSAM results influence resource allocations or regulatory compliance decisions, use global sensitivity analysis (e.g., Monte Carlo sampling over the whole weight range).
5. Transparency in reporting: journal editors and peer reviewers in this area could encourage authors to include a short section entitled "Sensitivity and Robustness," as they do in OR/decision science journals.

Addressing these issues would improve the trustworthiness of QSAM outcomes and support the development of benchmark datasets for validation, which is an important area for future research.

5.4 Managerial and Practical Implications

Implications in terms of practice involve the recognition that QSAMs are not just theoretical concepts but are also valuable instruments in the process of quality improvement and strategic management.

The industries that confront such compromises are e-banking, e-learning, and e-health [37], where the issues of trustworthiness, privacy, costs, and customer satisfaction are the major concerns. The incorporation of QSAM into the MCDM paradigm would enable decision makers to weigh these aspects rationally based on facts instead of intuition.

The key managerial implications are as follows:

1. Strategic Prioritization: QSAMs assist managers in deciding which qualities are essential to establish overall quality. For instance, when considering e-banking, user trust would be more dependent on security and availability rather than the interface.

2. **Resource Optimization:** The weighting technique in the suggested MCDM methodology ensures that resource allocation is done effectively based on technological as well as other user-related aspects.
3. **BI and Continuous Monitoring:** The repetitive nature of QSAMs supports continuous benchmarking. This is similar to the approach taken by cloud service providers who constantly compare their performance levels to SLAs.
4. **Transparency and Governance:** The transparency of MCDM-based methods enables their justification in meeting quality assurance and regulatory requirements [30].

All these factors, together, support the significance of QSAMs for management, as these QSAMs offer more objective improvements on past subjective appraisal methods.

5.5 Research Gaps and Future Pathways

However, it seems that some gaps have not been filled by different studies on the topic, such as:

- **Lack of empirical depth:** Only up to a quarter of the examined literature makes use of QSAMs on actual traces.
- **Lack of Standardization:** E-service quality indicators lack a standard classification system, as dimensions differ significantly from sector to sector.
- **Integration of analysis of real-time analytics data is not effective** since analysis is mainly done by conducting surveys.
- **Minimal Use of Feedback:** The model doesn't incorporate constant feedback from either users or experts.
- **Lack of Focus on Explainability and Ethics:** Although there has been growth in the use of artificial intelligence and automation in evaluation services, the area of explainability and ethics remains an uncharted territory [32].

The presence of such gaps suggests the creation of a new research approach that merges the concepts of Evaluation Theory, MCDM, and decision support technologies of AI to provide flexible and transparent QSAMs for next-generation e-services.

The seven Evaluation Theory dimensions synthesized in this review — evaluation target, criteria, yardstick, data-gathering, synthesis, process, and validation — offer a domain-agnostic structure [38] that should, in principle, extend beyond the e-learning, e-banking, and e-government contexts that dominate the reviewed literature (as shown in Table 8). However, none of the 78 studies reviewed addressed AI-as-a-Service, edge-computing services, or metaverse-based services directly, so applying this framework to these paradigms would require non-trivial adaptation rather than being a finding substantiated by the present review:

1. **AI-as-a-Service:** the criteria layer would need to expand beyond the dimensions in Table 7 to include model transparency and explainability, performance drift/degradation monitoring, and fairness—directly extending the "Explainability and Ethics" gap already identified above.
2. **Edge-computing services:** criteria such as latency, on-device resource consumption, and offline resilience would need to be added, and data-gathering would likely need to shift from the survey-dominant approach identified under RQ4 toward real-time telemetry, consistent with the "real-time analytics integration" gap above.
3. **Metaverse-based services:** intangible, experiential qualities (immersion, sense of presence, avatar trust) fall outside the largely functional criteria synthesized in this review and would likely require new yardsticks beyond the Likert-type and fuzzy scales identified under RQ3.

We therefore present this extension as a concrete direction for future research — consistent with the AET-QSAM model proposed in Section 6 — rather than a claim supported by the present dataset.

6 Conclusion

This paper performed a systematic literature review (SLR) to discuss, compare, identify, and integrate the unsolved problems and limitations of Multi-Criteria Decision-Making (MCDM) approaches for Quality Service Assessment Methods (QSAMs). One of the important contributions of the present SLR is to open new avenues for research that can inspire researchers to venture into the field of e-service quality assessment. Evaluation Theory was chosen as the theoretical approach to use to evaluate and compare the comprehensiveness of the existing QSAMs. All the major stages of the evaluation process were established and utilized when establishing the sub-research questions for this review.

A rigorous review protocol was developed and applied to select 78 relevant studies published between 2010 and 2025 that addressed the use of MCDM methodologies for service-quality assessment. Signs from the SLR are that QSAM scholarship remains incomplete and disjointed. Use of Evaluation Theory in this review was found to be beneficial in promoting insight into the nature of the components and procedures involved with service-quality assessment. QSAM studies are now confronted with some theoretical and methodological shortcomings that require reconsidering the design of quality evaluation methods, elements, and processes for virtual service contexts.

To be useful and valid in real contexts, QSAMs must have the capacity to address different forms of e-service requirements and forms of appraisal, particularly implementation, progress, and impact studies. Future QSAMs must also have the capacity to (1) identify and justify the criteria for evaluation correctly, (2) calculate the yardsticks and measurement scales, (3) specify data-gathering and combining methods, (4) operationalize the appraisal process, and (5) provide high credibility and validity evidence.

We anticipate that this study will provoke further QSAM research towards the creation of new assessment frameworks or the improvement of current ones according to the theory of Evaluation. In future work, we plan to extend this SLR to other forms of QSAMs such as optimization-based, fuzzy logic-based, and AI-enhanced hybrid approaches and broaden the coverage of digital libraries searched. We also intend to apply the lessons learned through this review to develop and test a generic conceptual QSAM model, comparing it to other existing models to empirically validate the advantages of the proposed approach.

More specifically, we propose to develop the Adaptive Evaluation-Theory QSAM (AET-QSAM): a conceptual model that operationalizes the seven Evaluation Theory dimensions used throughout this review (RQ1–RQ7) and directly targets its three most significant gaps — weak validation practice (RQ7), the largely manual and one-time nature of existing evaluation processes (RQ6), and the absence of standardized sensitivity testing (Section 5.3). The model will combine a criteria layer grounded in the most-cited evaluation dimensions identified in Table 7 (usability, trust and security, and efficiency), a hybrid AHP–Fuzzy TOPSIS synthesis engine consistent with the dominant techniques identified in Table 8, and a built-in feedback loop and sensitivity-reporting module that operationalizes the baseline guidelines proposed in Section 5.3.

In terms of validation approach, we plan to validate AET-QSAM through comparative case studies in the three e-service domains identified in this review as the most frequently studied (Table 8): e-learning, e-banking, and e-government. In each domain, AET-QSAM will be applied alongside a conventional AHP–Fuzzy TOPSIS baseline, and the two will be compared using the following protocol:

1. Stage 1 – Expert panel review: five to seven domain experts per sector will assess criteria completeness and clarity using the rubric in Table 3.
2. Stage 2 – Empirical application: both models will be applied to real user and expert judgement data collected from each domain, and resulting rankings will be compared for consistency.
3. Stage 3 – Sensitivity and robustness testing: rank-stability will be assessed via one-at-a-time weight perturbation and Spearman rank correlation, following the guidelines proposed in Section 5.3.

4. Stage 4 – Trustworthiness scoring: both models will be scored against the ten principles in Table 3 to quantify the improvement in methodological completeness relative to the 50% average identified in this review.

DECLARATIONS

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval. This study is a systematic literature review of published sources and did not involve human participants, human data, or animals. Therefore, ethics approval is not applicable.

Informed consent. Not applicable because no human participants were involved in this study.

Data and code availability. This study is based exclusively on previously published literature. All included studies are cited in the References section, and the extracted and classified data are presented in the tables of this paper.

Author contributions. All authors contributed to the study in terms of conception, design, literature search, screening, and data extraction and analysis. The first draft of the manuscript was written by O. Al-Tarawneh and M. Alseid, and all authors commented on and revised previous versions. All authors read and approved the final manuscript.

Acknowledgements. The authors have no acknowledgements to declare.

Use of artificial intelligence. The authors used Grammarly for proofreading.

REFERENCES

- [1] O. S. Ighomereho, A. A. Ojo, O. S. Omoyele, and O. S. Olabode, "From service quality to e-service quality: measurement, dimensions and model," *Journal of Management Information and Decision Sciences*, vol. 25, no. 1, pp. 1–15, 2022.
- [2] W. J. Kettinger and C. C. Lee, "Perceived service quality and user satisfaction with the information services function," *Decis. Sci.*, vol. 25, no. 5-6, pp. 737–766, 1994.
- [3] A. Parasuraman, V. Zeithaml, and L. Berry, "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality," *J. Retail.*, vol. 64, no. 1, pp. 12–40, 1988.
- [4] J. J. Cronin and S. A. Taylor, "Measuring Service Quality: A Reexamination and Extension," *Journal of Marketing*, vol. 56, no. 3, pp. 55–68, Jul. 1992, doi: 10.1177/002224299205600304.
- [5] T. L. Saaty, "Decision making with the analytic hierarchy process," *Int. J. Serv. Sci.*, vol. 1, no. 1, pp. 83–98, 2008.
- [6] C. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications*. Berlin, Germany: Springer, 1981.
- [7] A. Mardani, A. Jusoh, E. K. Zavadskas, Z. Khalifah, and K. Md Nor, "Application of multiple-criteria decision-making techniques and approaches to evaluating service quality: a systematic review of the literature," *J. Bus. Econ. Manag.*, vol. 16, no. 5, pp. 1034–1068, 2015.
- [8] M. Scriven, *Evaluation Thesaurus*, 4th ed. Newbury Park, CA, USA: Sage, 1991.
- [9] W. Shadish, "Evaluation theory is who we are," *Amer. J. Eval.*, vol. 19, no. 1, pp. 1–19, 1998.
- [10] ISO/IEC 25010, *Systems and software quality models*. Geneva, Switzerland: ISO, 2011.
- [11] N. Singh, K. Kaur, and R. S. Deol, "Transition of E-Service Quality Dimensions from Diverse Business Settings to E-Learning: A Review from Learners' Perspective," *Indian Journal of Marketing*, vol. 52, no. 7, p. 8, Jun. 2022.
- [12] A. Mardani, A. Jusoh, K. Md Nor, Z. Khalifah, N. Zakwan, and A. Valipour, "Multiple criteria decision-making techniques and their applications – a review of the literature from 2000 to 2014," *Econ. Res.-Ekon. Istraživanja*, vol. 28, no. 1, pp. 516–571, 2015.
- [13] A. Parasuraman, V. Zeithaml, and L. Berry, "A conceptual model of service quality and its implications for future research," *J. Mark.*, vol. 49, no. 4, pp. 41–50, 1985.
- [14] S. A. R. Khan, Z. Yu, H. Golpira, A. Sharif, and A. Mardani, "A state-of-the-art review and meta-analysis on sustainable supply chain management: Future research directions," *Journal of Cleaner Production*, vol. 278, p. 123357, Jan. 2021.
- [15] J. Więckowski and W. Sałabun, "Sensitivity analysis approaches in multi-criteria decision analysis: A systematic review," *Appl. Soft Comput.*, vol. 148, p. 110915, 2023.
- [16] R. Rekik, I. Kallel, J. Casillas, and A. M. Alimi, "Assessing websites quality: A systematic literature review by text and association rules mining," *Int. J. Inf. Manag.*, vol. 38, no. 1, pp. 201–216, 2018.
- [17] S. J. Barnes and R. T. Vidgen, "An integrative approach to the assessment of e-commerce quality," *J. Electron. Commer. Res.*, vol. 3, no. 3, pp. 114–127, 2002.
- [18] C. Kahraman, Ed., *Fuzzy Multi-Criteria Decision Making: Theory and Applications with Recent Developments*. New York, NY, USA: Springer Science & Business Media, 2008.

- [19] R. K. Dhurkari, "MCDM methods: Practical difficulties and future directions for improvement," *RAIRO - Operations Research*, vol. 56, no. 4, pp. 2221–2233, 2022.
- [20] S. Opricovic and G.-H. Tzeng, "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS," *European Journal of Operational Research*, vol. 156, pp. 445–455, 2004.
- [21] D. Gefen, E. Karahanna, and D. W. Straub, "Trust and TAM in online shopping: An integrated model," *MIS Q.*, vol. 27, no. 1, pp. 51–90, 2003.
- [22] D. Pamucar, M. Yazdani, M. J. Montero-Simo, R. A. Araque-Padilla, and A. Mohammed, "Multi-criteria decision analysis towards robust service quality measurement," *Expert Syst. Appl.*, vol. 170, p. 114508, 2021.
- [23] B. Tedja, M. Al Musadieq, A. Kusumawati, and E. Yulianto, "Systematic literature review using PRISMA: exploring the influence of service quality and perceived value on satisfaction and intention to continue relationship," *Futur. Bus. J.*, vol. 10, no. 39, 2024.
- [24] A. Parasuraman, V. A. Zeithaml, and A. Malhotra, "E-S-QUAL: A multiple-item scale for assessing electronic service quality," *Journal of Service Research*, vol. 7, no. 3, pp. 213–233, 2005.
- [25] B. Kitchenham and S. Charters, *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Keele, U.K.: Keele Univ., 2007.
- [26] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021.
- [27] R. Buyya et al., "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Gener. Comput. Syst.*, vol. 25, no. 6, pp. 599–616, 2009.
- [28] R. Ladhari, "Developing e-service quality scales: A literature review," *J. Retailing Consum. Serv.*, vol. 17, no. 6, pp. 464–477, 2010.
- [29] G. Gkioka, T. Bothos, B. Magoutas, and G. Mentzas, "Data analytics methods to measure service quality: A systematic review," *Intell. Decis. Technol.*, vol. 17, no. 4, pp. 1007–1029, 2023.
- [30] D. J. Power and R. Sharda, "Model-driven decision support systems: Concepts and research directions," *Decis. Support Syst.*, vol. 43, no. 3, pp. 1044–1061, 2007.
- [31] D. Van Rooy, "Human-machine collaboration for enhanced decision-making in governance," *Data & Policy*, vol. 6, p. e60, 2024.
- [32] L. Nannini, M. Marchiori Manerba, and I. Beretta, "Mapping the landscape of ethical considerations in explainable AI research," *Ethics Inf. Technol.*, vol. 26, no. 3, Art. no. 44, 2024.
- [33] P. Aazagreyir, P. Appiahene, O. Appiah, S. Boateng, W. L. Brown-Acquaye, and G. Koi-Akrofi, "An integrated fuzzy multi-criteria decision-making methods for service selection: A systematic literature review and meta-analysis," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 1, 2022.
- [34] A. Birch-Jensen, I. Gremyr, and Á. Halldórsson, "Digitally connected services: Improvements through customer-initiated feedback," *Eur. Manag. J.*, vol. 38, no. 5, pp. 814–825, 2020.
- [35] H. Alabool, A. Kamil, N. Arshad, and D. Alarabiat, "Cloud service evaluation method-based Multi-Criteria Decision-Making: A systematic literature review," *J. Syst. Softw.*, vol. 139, pp. 161–188, 2018.
- [36] Al-Akash, Omar M., et al. "Multi-criteria decision-making strategies pilot study." *AIP Conference Proceedings*. Vol. 3406. No. 1. AIP Publishing LLC, 2026.
- [37] L. R. Alzabin, "Big Data application for healthcare management," *Emerging Technologies in the Healthcare System for Sustainable Development*, pp. 165–196, 2026.
- [38] O. Alharasees, H. A. Al Hasanat, and D. Alothman, "Adoption of autonomous and shared autonomous vehicles: a multi-criteria decision-making approach," *International Journal of Transportation Science and Technology*, 2026.