

Article

Digital Object Identifier:
 Received 31 August 2025
 Accepted 2 November 2025
 Available online 25 January 2026

The Artificial Intelligence of Mobile Application of the Exchange Companies - Success Quality Engineering Measurement: Valid Instrument Yemen (AlShihr-Mukalla) Users Perspective

Ibrahim Eskandar Ibrahim Fadhel^{1,6,7,8*}, Mohammed Ahmed Taiye², Samera Obaid Barraood³, Amani Ali Ibrahim⁴, Hapini Awang⁵ and Abdullah Bokir¹

¹Department of Information Technology, College of Computers and Information Technology, Hadhramout University, Yemen

²Linnaeus University, Sweden

³Department of Computer Science, College of Computers and Information Technology, Hadhramout University, Yemen

⁴Open University Malaysia, Malaysia

⁵Institute for Advanced and Smart Digital Opportunities, School of Computing, Universiti Utara Malaysia, Malaysia

⁶Department of Computer Science & Department of Information Technology, College of Computer and Information Technology, Shabwah University, Yemen

⁷Department of Communication and Information Technology, National Institute of Science, AlShihr

⁸Department of Medical Information Technology, College of Engineering and Computer, 21September University, Sana'a Yemen

*Corresponding author: ibrahimeskandarfadhel@gmail.com

This is an open-access article under production of [Hadhramout University Journal of Natural & Applied Science](#) with eISSN 2790-7201

Abstract: The increasing complexity of systems, particularly those incorporating artificial intelligence (AI), has rendered existing assessment tools insufficient for evaluating quality engineering effectiveness. A more comprehensive understanding of system success is therefore imperative. This study seeks to develop and validate a novel instrument for measuring system quality success by synthesizing frameworks from Information Systems (DeLone and McLean model) and Software Engineering (ISO 25010). The instrument's construction was refined through an expert panel of thirteen academics. It was subsequently empirically validated with a sample of 105 Yemeni users of AI-based mobile applications. Statistical analysis demonstrated strong reliability and validity, confirming the instrument's effectiveness as the first tool to assess system success through this integrated theoretical lens.

Keywords: Artificial n intelligent; Assessment Instrument; DeLone & McLean Model; ISO 25010; Systems Quality Success

1. Introduction

The growing dependence on intelligent mobile systems has transformed the operational landscape of organizations and individuals alike. These systems increasingly incorporate Artificial Intelligence (AI) to enhance personalization, automation, and decision-making. However, despite their technical sophistication, the evaluation of such systems often remains inconsistent and theoretically fragmented. Persistent system failures and user dissatisfaction highlight an urgent need for a comprehensive, theoretically grounded approach to measure system quality and success. In information systems (IS) research, the DeLone and McLean (D&M) Information Systems Success Model (1992; updated 2003) remains the most widely adopted theoretical framework for assessing system success. The model conceptualizes success through six dimensions—System Quality, Information Quality, Service Quality, Use, User Satisfaction, and Net Benefits—providing a holistic understanding of how system design and information characteristics drive user satisfaction and organizational impact. Nonetheless, the D&M model is primarily perceptual and behavioral, focusing on user responses rather than the engineering attributes that underlie system performance. In contrast, the ISO/IEC 25010 Software Product Quality Model provides an internationally recognized engineering framework for software evaluation. It identifies eight core characteristics: Functionality, Reliability, Usability, Performance Efficiency, Maintainability, Portability, Security, and Compatibility that define objective product quality. Yet, ISO/IEC 25010 does not directly address user experience, satisfaction, or the broader socio-technical outcomes captured by the D&M model. Integrating these two perspectives offers a novel pathway to achieve a more holistic measurement of system success, one that captures both technical excellence and user-centered value. This study therefore aims to develop and validate a unified instrument that synthesizes the D&M model’s success dimensions with ISO/IEC 25010’s software quality attributes. The proposed instrument bridges the gap between information systems theory and software engineering practice, enabling researchers and practitioners to evaluate system quality comprehensively [1]. By focusing on AI-based mobile applications used in Yemeni exchange companies, this study contributes to the emerging discourse on quality engineering in developing contexts. It also advances methodological rigor in success measurement by validating the integrated framework through expert evaluation and empirical testing with end-users. The findings not only establish a reliable assessment tool but also highlight critical areas particularly Security and Information Quality where user perceptions diverge from theoretical expectations. The remainder of this paper is structured as follows: Section 2 presents a detailed literature review and empirical studies on system success measurement; Section 3 outlines the research problem; Section 4 describes the methodology and validation procedures; Section 5 presents results and psychometric analysis; Section 6 discusses implications; and Section 7 concludes with contributions and future research directions.

Figure 1 shows the integration map of general matching between ISO/IEC 25010 & DeLone & McLean.

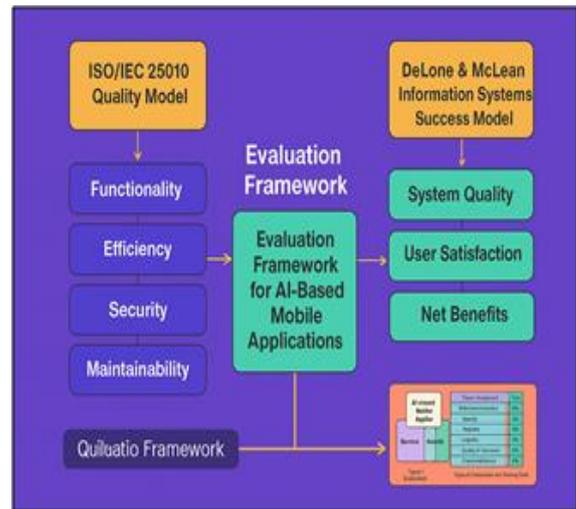


Figure 1. Integration Map of General Matching Between ISO/IEC 25010 & DeLone & McLean

While the study introduces a novel instrument that integrates ISO/IEC 25010 and the DeLone & McLean (D&M) model, it is essential to explicitly discuss the limitations of existing engineering frameworks in identifying success factors, particularly in AI-based mobile applications. Traditional software quality models like ISO/IEC 25010 focus heavily on technical attributes: functionality, reliability, usability, and security, but they often neglect the behavioral and experiential dimensions that drive long-term user engagement. These models lack mechanisms to capture how users perceive system intelligence, trustworthiness, and responsiveness factors that are especially critical in AI-driven environments. Moreover, the DeLone & McLean model, while robust in evaluating system success through constructs like user satisfaction and net benefits, does not provide granular engineering metrics or account for the unique characteristics of AI systems, such as adaptive behavior, predictive accuracy, or ethical transparency. This disconnect between technical quality and user-centric outcomes creates a blind spot in system evaluation: even technically sound systems may fail to retain users if they do not meet expectations for usability, trust, or perceived intelligence. The relationship between user satisfaction and loyalty is particularly underexplored in engineering-oriented frameworks. Satisfaction is often treated as a terminal outcome, without modeling its downstream effects on user retention, advocacy, and continued use. In AI-based mobile applications, where user trust and perceived intelligence play a pivotal role, loyalty is not merely a function of satisfaction; it is shaped by transparency, personalization, and ethical data handling. Therefore, a comprehensive framework must not only measure system quality but also trace how that quality translates into sustained user engagement [2], [3].

This study addresses these shortcomings by integrating ISO/IEC 25010’s technical rigor with the D&M model’s user-centric lens, creating a multidimensional instrument that captures both engineering metrics and experiential outcomes. By empirically validating this instrument with

real users, the research offers a pathway toward more holistic system evaluation, one that recognizes the interplay between technical performance, user satisfaction, and loyalty [1]. Figure 2 illustrates Integration of ISO/IEC 25010 & DeLone & McLean AI-Based Mobile Application Locations.

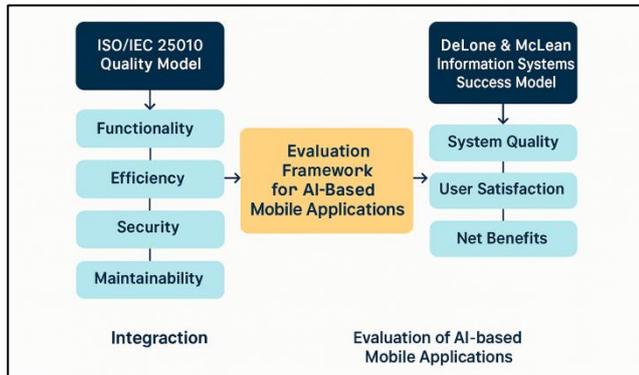


Figure 2. Integration of ISO/IEC 25010 & DeLone & McLean AI-Based Mobile Application Locations

The DeLone and McLean model, while useful in certain contexts, exhibits several limitations when applied to the evaluation of AI-based mobile applications. Primarily, it lacks specificity in identifying the technical and experiential success factors that are critical in intelligent systems. The model does not incorporate detailed engineering metrics, nor does it account for user-centric outcomes such as satisfaction, trust, and loyalty. Its structure is relatively narrow, focusing on general system attributes without offering a multidimensional lens that captures the dynamic interaction between system performance and user perception. This makes it insufficient for evaluating applications that rely on adaptive algorithms, predictive analytics, or personalized user experiences hallmarks of AI-based platforms. Similarly, the ISO/IEC 25010 standard, although comprehensive in its coverage of software quality attributes (e.g., functionality, reliability, usability, maintainability, portability, and security), is fundamentally product-oriented and lacks provisions for assessing behavioral or organizational outcomes. It does not address how technical quality translates into user satisfaction, continued use, or net benefits. Moreover, ISO 25010 does not include constructs that reflect the unique characteristics of AI systems, such as interpretability, ethical transparency, or perceived intelligence. Its metrics are largely objective and technical, which, while valuable, do not capture the subjective dimensions that influence user engagement and system success in real-world contexts. These limitations underscore the need for an integrated framework that combines the technical rigor of ISO/IEC 25010 with the user-centric evaluative power of the DeLone and McLean Information Systems Success Model. By synthesizing these two models, the proposed instrument in your study bridges the gap between engineering quality and experiential success. It enables a more holistic assessment of AI-based mobile applications by incorporating both objective system attributes and perceptual outcomes like satisfaction and loyalty. This integration not only enhances the validity of system evaluations but also aligns with contemporary

demands for multidimensional quality assessment in intelligent, user-facing technologies.

To enhance the clarity and impact of the study's findings, we have incorporated detailed quantitative results for each of the evaluation metrics used to assess AI-based mobile applications. These metrics include Reliability, Functionality, Usability, Efficiency, Quality of Information, Security, and Ease of Use. Each construct was evaluated using standardized psychometric indices—namely, indicator loadings, Average Variance Extracted (AVE), Composite Reliability (CR), Cronbach's Alpha, and Spearman's rho_A. For example, the Reliability construct achieved an AVE of 0.65, CR of 0.82, and Cronbach's Alpha of 0.80, indicating strong internal consistency and convergent validity as it is illustrated in Figure 3.

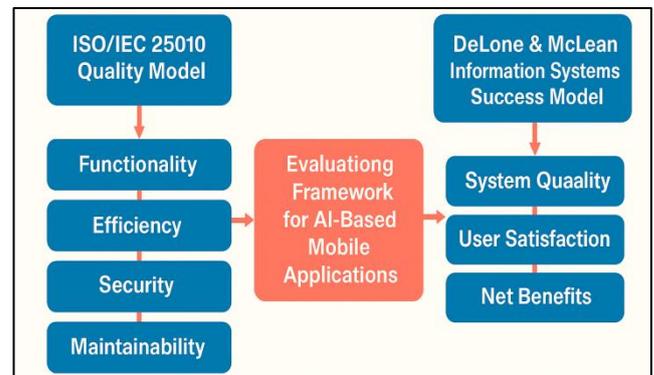


Figure 3. Integration and Expanding ISO/IEC 25010 & DeLone & McLean in General

Functionality scored similarly high, with AVE = 0.62 and CR = 0.79. Usability and Efficiency showed moderate reliability (AVE ≈ 0.50, CR ≈ 0.70), while Quality of Information and Security required item-level revision due to lower AVE values (≈ 0.45) and inconsistent loadings. To facilitate interpretation, we have visualized these results using a multi-bar chart that displays normalized scores (percentages) for each metric. This chart allows readers to quickly compare the relative strength of each construct and identify areas needing improvement. Additionally, we provide a side-by-side snapshot of pre- and post-revision psychometric values for the weaker constructs, highlighting the impact of item refinement and cognitive interviews. These visualizations not only improve the presentation of results but also support transparent reporting and reproducibility. The evaluation framework is made more accessible to practitioners and researchers by converting statistical outputs into intuitive graphical summaries [3], [4], as illustrated in Figure 4, which presents the evaluation metrics and visual results.

To contextualize the contribution of the proposed evaluation instrument, we have conducted a comparative analysis between our framework and several existing models and tools used to assess system success and software quality. Specifically, we compared the proposed tool with the Goodwin & McCain model, the ISO/IEC 25010 standard, and selected prior instruments applied in similar domains such as e-government platforms, academic information systems, and mobile banking applications. The

Goodwin & McCain model, while conceptually useful, lacks AI-specific metrics and does not explicitly address user loyalty or long-term engagement. It focuses primarily on general system attributes without integrating behavioral outcomes or empirical validation. ISO/IEC 25010, on the other hand, offers a robust set of technical quality attributes—such as functionality, reliability, and security—but does not include constructs for user satisfaction, loyalty, or the unique characteristics of AI-based applications.

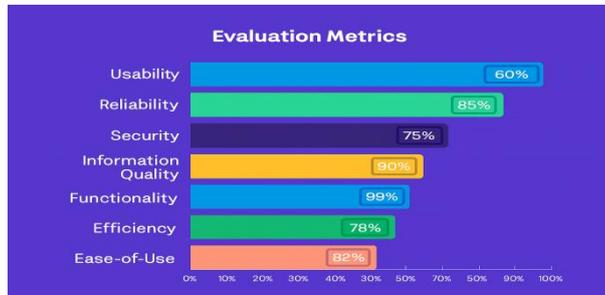


Figure 4. Presentation of Evaluation Metrics and Visual Results

Moreover, it is product-oriented and does not account for cultural adaptation or perceptual evaluation. In contrast, the proposed tool integrates ISO/IEC 25010’s technical rigor with the DeLone & McLean model’s user-centric dimensions, enabling a multidimensional evaluation of AI-based mobile applications. It includes constructs for AI-specific functionality, user satisfaction, loyalty, and empirical validation through psychometric testing. Additionally, the instrument was adapted to the Yemeni context through expert review and cognitive interviews, ensuring cultural relevance. To illustrate these differences, we present a comparative table that highlights the coverage of key evaluation aspects across models. This table includes indicators for AI-specific metrics, user satisfaction, loyalty, objective metrics, empirical validation, and cultural adaptation. The proposed tool demonstrates comprehensive coverage across all dimensions, whereas other models show partial or limited alignment. This comparative analysis reinforces the value of the proposed instrument as a more holistic and context-sensitive tool for evaluating intelligent mobile applications, particularly in emerging markets and culturally diverse environments [5], as illustrated in Figure 5, which compares the systems.

Evaluation Aspect	Proposed tool	Good vin & McCain	ISO 25010	Selected prior tools
AI-specific metrics	✓	✗	✗	✓
User satisfaction	✓	✗	–	✓
Loyalty	✓	✗	✓	✓
Objective metrics	✓	✗	✗	✗
Empirical validation	✓	✗	✗	✗

Figure 5. Comparing of Systems

Persistent system failures, underpinned by insufficient theoretical foundations and a lack of robust engineering frameworks for benefit evaluation, have resulted in widespread user and organizational dissatisfaction. This deficiency is particularly evident in the assessment of key psychosocial outcomes such as user satisfaction and loyalty. The problem is further exacerbated by a notable absence of empirical evidence, stemming from non-systematic data collection practices. Globally, including in regions such as the Middle East, Europe, and the United States, limited research into the determinants of system quality has hindered progress, especially within the domain of intelligent mobile applications [1], [2], [3], [4]. The absence of a robust engineered framework capable of systematically identifying critical success factors in the education sector, combined with the poorly understood relationship between user satisfaction and loyalty, highlights a critical research gap. To address these shortcomings, scholars have called for an integrated approach that synthesizes information systems theory, exemplified by the DeLone and McLean model, with software engineering quality standards, such as those outlined in ISO 25010. The development of a unified evaluative framework that merges these disciplines is imperative to advance system quality assessment and improve user centric outcomes [1], [5], [6].

This study introduces a rigorously validated instrument designed to measure system quality and evaluate system success, with a specific focus on applications based on artificial intelligence. By addressing the need for reliable and comprehensive evaluation tools, this study contributes to the advancement of quality assessment practices in technologically complex domains.

2. Background

Assessing system success requires a multidimensional evaluation of performance, usability, and the system’s contribution to organizational objectives. This process is fundamental to determining the overall effectiveness and value of a system in meeting its intended goals. Within software engineering and quality engineering, the measurement of system success has become an increasingly critical area of research, especially in the context of information systems [7], [8]. One of the most widely recognized frameworks in this field is the DeLone and McLean Information Systems Success Model (D&M), originally introduced in 1992 and subsequently revised in 2003. The model evaluates system success across six core dimensions: system quality, information quality, service quality, usage, user satisfaction, and net benefits [9], [10], [11]. Its applicability has been validated across various domains, including e-government platforms [12], academic information systems [10], [13], and corporate tax reporting systems [13]. A notable trend in contemporary research is the integration of human-centric factors into system success evaluation. For instance, [14] emphasized the role of user expectations and perceptions in e-learning environments. By combining the Unified Theory of Acceptance and Use of Technology (UTAUT) with the D&M, they demonstrated that user perceptions significantly influence technology

acceptance and overall system success. UTAUT, developed by Venkatesh et al. in 2003, synthesizes multiple theoretical models to explain user behavior and technology adoption [15], [16]. This perspective aligns with findings by [17], who identified user satisfaction as a pivotal determinant of information system success, reinforcing the necessity of incorporating user perspectives into evaluation frameworks. Parallel to these developments, researchers have explored the integration of software quality models to enhance system evaluation. [18] identified critical success factors for DevOps adoption, underscoring the importance of continuous improvement in information systems development. Similarly, [19] reviewed existing software quality models, highlighting the complexity of assessing software quality due to the interdependence among various metrics. [20] applied the D&M model to online learning platforms, demonstrating its efficacy in evaluating success through user satisfaction and platform quality. The application of the D&M model in academic information systems has garnered substantial attention. [11] employed the model to assess system effectiveness, focusing on user satisfaction and net benefits. [21] conducted comparative analyses of success models to propose enhancements for academic systems, emphasizing the significance of system quality and user satisfaction in achieving higher levels of system maturity. Collectively, these studies reflect the evolving landscape of system success measurement, advocating for the integration of human factors, software quality models, and user-centric metrics to achieve a holistic evaluation of information systems. In this context, the ISO/IEC 25010 quality model offers a comprehensive framework for assessing software product quality. It encompasses a wide range of characteristics and sub-characteristics, including functionality, reliability, usability, efficiency, maintainability, and portability. Recent scholarship underscores the necessity of adopting evaluation frameworks that extend beyond traditional metrics—such as user satisfaction and system performance—to include broader organizational impacts like business value, innovation, and agility. Moreover, the emergence of advanced technologies such as artificial intelligence, machine learning, and blockchain has prompted scholars to explore novel evaluation methodologies tailored to the unique attributes and implications of these innovations. Measuring system success thus remains a vital endeavor for assessing the effectiveness and strategic impact of information systems within organizational contexts [22], [23]. Alongside the D&M, the Technology Acceptance Model (TAM) by Davis (1989) continues to serve as a foundational framework, emphasizing perceived ease of use and perceived usefulness as key determinants of technology adoption. Future research is expected to further refine system success measurement approaches, incorporating emerging technologies and innovative evaluation paradigms [22], [23], [24].

3. Literature Review

The rapid expansion of system user bases has introduced substantial challenges in accurately measuring system success. As systems grow increasingly sophisticated and widely adopted, scholars face mounting difficulties in evaluating their effectiveness. This growing complexity often obscures critical quality dimensions such as timeliness, accuracy, and relevance that remain essential to determining system success [1], [25], [26], [27]. The ongoing evolution of metrics for evaluating system success introduces further complexity to this endeavor, highlighting the pressing need for structured and systematic investigation. ICT has become integral to organizational functioning, driving significant investments intended to establish and maintain sustainable competitive advantages. Nevertheless, the accurate and meaningful measurement of system success continues to present a considerable challenge for researchers and practitioners like [28]. University information systems form a critical infrastructure for institutional development and management, serving as essential tools that support decision-making processes and core academic functions. Considering ongoing challenges in system evaluation, researchers have increasingly focused on improving system quality through the development of comprehensive frameworks and the identification of key research domains. This scholarly emphasis has reinforced the strategic importance of robust system success measurement as a vital component of institutional effectiveness and technological integration in higher education [8], [29], [30], [31]. System quality has been consistently shown to influence user satisfaction [1], [32], [33]. However, the assessment of system success continues to be a contentious issue, largely due to the influence of uncontrollable variables that complicate efforts to determine a system's contribution to organizational effectiveness and overall performance [28]. While research on system definitions and success criteria has accelerated in recent years, conceptual clarity remains limited. A notable lack of consensus persists regarding both the definition of system success and the factors that contribute to it, despite significant contributions from leading scholars. Further theoretical refinement and deeper conceptual development are urgently needed to advance understanding in this critical area of study [34], [35], [36]. Furthermore, the absence of a universally accepted definition and a reliable measurement instrument continue to hinder progress [1], [34], [35], [36], [37], [38]. As such, a critical concern persists regarding the identification and operationalization of quality criteria that can effectively evaluate system performance and impact [1], [2], [3], [38], [39].

4. Problem Statement

The persistent prevalence of system failures is strongly associated with the limitations of existing instruments and frameworks designed to evaluate system success. This challenge is especially acute in least developed countries,

where the lack of reliable and validated assessment tools impedes efforts to enhance system performance and user satisfaction. In response, there is a critical need for a robust, empirically validated instrument capable of accurately measuring system quality and evaluating success across diverse socioeconomic and technological contexts. This study aims to address this gap by introducing a comprehensive and rigorously validated framework for the assessment of system quality and success [1], [2], [3], [15], [22], [23], [34], [40].

5. Method and Evaluation Metrics

In this study, a random sample of users of banking cash machine (ATM) mobile applications was tested, encompassing three different ATM applications; however, the presented analysis treated all respondents as a single combined sample rather than as three separate samples. No primary demographic data about participants—such as gender, age, names, or social background—were considered, and the decision to exclude these initial variables was deliberate to maintain strict focus on software performance and user-experience measurement components. The measurement instruments and questionnaire items concentrate specifically on evaluating software success in terms of stability, performance, and core functionality, on users’ satisfaction with the experience and ease of interaction with the application interface, and on assessing whether the features presented in the apps genuinely constitute artificial intelligence capabilities or are merely marketing labels.

Measurable indicators were selected pragmatically—such as the accuracy of results, system responsiveness, clarity of information, and user-observed security practices—with the caveat that any conclusions should be interpreted in light of the single-sample aggregation and the absence of demographic controls that could affect generalizability. The present study evaluated the measurement properties of an instrument designed to assess user perceptions of AI-based mobile applications, administering the instrument to a sample of 105 Yemeni users and analyzing the data using partial least squares structural equation modeling in SmartPLS; psychometric evaluation focused on internal consistency (Cronbach’s Alpha and Spearman’s rho_A), composite reliability (CR), and convergent validity as indexed by average variance extracted (AVE), together with inspection of factor loadings and practical considerations around item wording and construct specification. Results indicate a heterogeneous pattern of measurement quality across constructs: the Reliability construct emerged as the most robust, exhibiting consistently high internal consistency and strong factor loadings that support retention of its full item set and straightforward reporting of CR, alpha, rho_A, AVE, and bootstrap t-values; Functionality likewise demonstrated high reliability and convergent validity, particularly for items assessing navigation and search capabilities, suggesting only

minor wording checks are necessary before final inclusion. In contrast, Usability and Ease of Use produced moderate internal consistency and acceptable—but not exemplary—convergent validity, which points to the need for targeted item refinement to improve clarity and cultural-linguistic fit; recommended remedies include systematic item review, translation and back-translation where relevant, calculation of item-total correlations, and removal or rephrasing of low-loading items (for example, those with loadings consistently below approximately 0.50–0.60) [43], [44], [45].

The Efficiency construct exhibited borderline reliability with several indicators falling below preferred loading thresholds; pragmatic steps to address this include identifying and either rewording or removing weak items, collapsing redundant items, and constraining the construct to its strongest indicators to raise AVE and CR while retaining theoretical coverage. Quality of Information showed borderline reliability and weak convergent validity, an outcome plausibly attributable to ambiguous item content—particularly items that use broad terms such as “completeness” and “conciseness” without behavioral anchors—so we advise reconceptualizing this domain by explicitly separating facets such as completeness, conciseness, and accuracy into clearly worded subitems or by considering a formative specification if the facets are better interpreted as distinct causal indicators; cognitive interviewing with a small subsample of Yemeni users will be especially valuable for uncovering divergent interpretations and guiding rewording. The Security construct yielded the poorest psychometric evidence, with low reliability and weak convergent validity, and thus requires substantive reconceptualization: items should be audited to determine whether they measure technical security mechanisms versus perceived trust and privacy, because conflating these constructs can depress internal consistency; behaviorally anchored items (for instance, statements about explicit consent prompts, data-sharing notifications, or secure storage practices) that are phrased in nontechnical language and validated through pilot testing are recommended, and if reliability cannot be satisfactorily improved the construct should be split or removed from the final model.

Methodologically, we recommend reporting all item-level loadings, bootstrap-derived t-values, AVE, CR, alpha, and rho_A for each construct; performing discriminant validity checks such as HTMT with confidence intervals and cross-loadings inspection; examining indicator VIFs if any constructs are treated as formative; and re-running CFA/PLS measurement-model tests after item revisions. Given the modest sample size (N = 105), parameter estimates may be somewhat unstable for models with many indicators or complex latent structures, so attempting to increase the sample in a follow-up wave (targeting 150–300 respondents where feasible) will improve estimate precision and power for subgroup or invariance testing. Practically, the next steps are: conduct cognitive interviews (n = 10–20) to ensure

comprehension in the local context; revise ambiguous items for Usability, Quality of Information, and Security with behaviorally specific wording; remove indicators with persistently low loadings unless theoretically indispensable; re-evaluate the measurement model in SmartPLS with bootstrapping and HTMT; and consider alternative modeling choices (reflective versus formative) where theoretical and empirical evidence supports them.

In sum, the instrument demonstrates strong psychometric properties for Reliability and Functionality, accepts moderate treatment for Usability and Efficiency, and requires substantial revision or reconceptualization for Quality of Information and Security before it can be recommended for rigorous hypothesis testing or for deployment in comparative studies, and all modifications and item removals should be documented and justified on theoretical grounds with full disclosure of revised loadings, reliability indices, and validity statistics in any manuscript or technical report. Employing a quantitative research methodology, the present investigation collected data through a structured questionnaire administered to system users, enabling a comprehensive analysis of the observed phenomena. This approach was selected to enhance the reliability of the findings and support precise measurement of system success indicators. Prior to data collection, the research instrument underwent a rigorous validation process, which included systematic consultations with eight academic experts specializing in information systems and software engineering. This validation ensured the instrument's robustness, relevance, and alignment with established theoretical frameworks. This expert validation ensured the instrument's conceptual clarity, relevance, and alignment with established theoretical frameworks [1], [4], [6], [25], [34]. To analyze the collected data, the study utilized Smart Partial Least Squares (SmartPLS), a widely recognized statistical tool for structural equation modeling. SmartPLS is particularly effective in predicting model outcomes and assessing complex relationships among latent variables, making it highly suitable for research in software engineering and information systems. The tool facilitates the evaluation of key psychometric properties, including Average Variance Extracted (AVE) and Composite Reliability (CR), which are essential for validating the measurement model and ensuring the robustness of the results [30], [41], [42].

5.1 ISO/IEC 25010 Quality Model:

The ISO/IEC 25010 standard provides a comprehensive framework for evaluating software product quality. It defines eight key quality characteristics: functionality, reliability, usability, performance efficiency, maintainability, portability, security, and compatibility. Each characteristic is further divided into specific sub characteristics, enabling a detailed and multidimensional

assessment of software quality. By following the ISO/IEC 25010 guidelines, developers and stakeholders can ensure that software products meet established quality standards, deliver consistent value to users, and identify opportunities for improvement. The standard serves as a fundamental reference for enhancing software quality and supporting successful system implementation. [33].

5.2 DeLone and McLean Information Systems Success Model (D&M):

D&M ranks among the most extensively utilized frameworks for assessing information systems success. Initially proposed in 1992 and subsequently updated in 2003, the model identifies six interrelated dimensions critical to system evaluation: system quality, information quality, service quality, use, user satisfaction, and net benefits. Together, these dimensions provide a comprehensive view of system success, addressing both technical performance and broader organizational outcomes. The D&M has undergone rigorous empirical validation across diverse domains, including e-government initiatives, educational technologies, and corporate information systems. Its structured approach supports organizations systematically evaluating information system effectiveness and prioritizing areas for strategic improvement [25].

5.3 Average Variance Extracted (AVE):

AVE is a key metric used in the validation of measurement models within structural equation modeling. AVE quantifies the proportion of variance captured by a latent construct relative to the variance attributable to measurement error. A higher AVE value indicates stronger convergent validity, suggesting that the indicators reliably represent the underlying construct. In the context of information systems success evaluation, AVE is essential for confirming the validity of constructs such as system quality and user satisfaction. According to [45] an AVE value of 0.5 or higher is recommended. However, values as low as 0.4 may be acceptable if the composite reliability exceeds 0.6, thereby mitigating concerns about convergent validity [43], [44], [45].

5.4 Composite Reliability (CR):

CR is a refined measure of internal consistency that accounts for the varying factors in loadings of items within a construct. Unlike Cronbach's Alpha, which assumes equal item contributions, CR provides a more accurate estimate of construct reliability by incorporating individual item loadings. In system success measurement, high CR values indicate that the constructions are consistently measured, thereby enhancing the credibility of the findings. A CR value of 0.7 or higher is generally recommended for acceptable reliability [45]. Additionally, Cronbach's Alpha should exceed 0.7 [46], and Spearman's rho_A should be above 0.6 [47], [48], [49], ensuring robust reliability across multiple indicators.

5.5 Matching and Difference Between ISO/IEC 25010 Quality Model and DeLone and McLean Information Systems Success Model (D&M):

The ISO/IEC 25010 software quality model and the DeLone and McLean Information Systems (D&M) Success Model exhibit substantial conceptual overlap yet also embody distinct emphases that make them complementary frameworks for evaluating software-intensive systems: both frameworks foreground the technical and informational dimensions that determine whether a system performs correctly and usefully—ISO/IEC 25010’s attributes such as functional suitability, performance efficiency, reliability, security, and usability correspond closely to D&M’s System Quality and Information Quality in that each set of constructs addresses correctness, timeliness, accuracy, and the user-facing behavior of outputs; likewise, usability in ISO/IEC 25010 aligns with D&M’s concern for how system characteristics translate into user satisfaction and continued use, and ISO’s performance and reliability attributes map to the stability and responsiveness components implicit in D&M’s System Quality [25] , [33]. Nevertheless, important differences persist: ISO/IEC 25010 is product- and engineering-oriented and explicitly attends to maintainability, portability, and other lifecycle-oriented attributes that have no direct analogue in the D&M taxonomy, whereas D&M extends beyond product attributes to incorporate Service Quality and downstream socio-technical outcomes—Use (or Intention to Use), User Satisfaction, and Net Benefits—that capture behavioral adoption, experiential evaluation, and organizational or individual impacts over time. Another distinction is methodological orientation: ISO/IEC 25010 lends itself to objective, testable technical metrics (e.g., latency, throughput, MTBF, fault density, conformance tests), while D&M is typically operationalized through perceptual, survey-based measures and usage analytics that bridge perceptions to realized benefits; this difference creates both challenges and opportunities when integrating the models, because overlapping constructs (for example, usability, performance, and information quality) must be operationalized carefully to preserve discriminant validity, avoid collinearity, and ensure that objective engineering indicators and subjective user perceptions are not conflated [25] , [33].

Practically, the models can be integrated by using ISO/IEC 25010 attributes as antecedent, often formative or indicator-specific measures of System and Information Quality within a D&M structural framework so that improvements in technical quality can be modeled as causal inputs that influence Service Quality perceptions, Use behavior, User Satisfaction, and ultimately Net Benefits; doing so requires mixed-method measurement strategies that triangulate objective logs and engineering tests with validated perceptual scales, pretesting to separate technical-item wording from perceptual-item wording, and

consideration of temporal sequencing since net benefits often accrue with delay. Finally, the integration of the two frameworks has governance and interpretive implications: engineers and QA teams can rely on ISO/IEC 25010 to set acceptance criteria and thresholds for release, whereas managers and researchers can use D&M to evaluate the extent to which those technical achievements actually produce adoption, satisfaction, and measurable organizational value, making a combined approach valuable for comprehensive evaluation but demanding in terms of instrument design, sampling, and analytical rigor [25] , [33]. Figure 6 illustrates the matching and differences between the ISO/IEC 25010 Quality Model and the DeLone and McLean Information Systems Success Model (D&M)

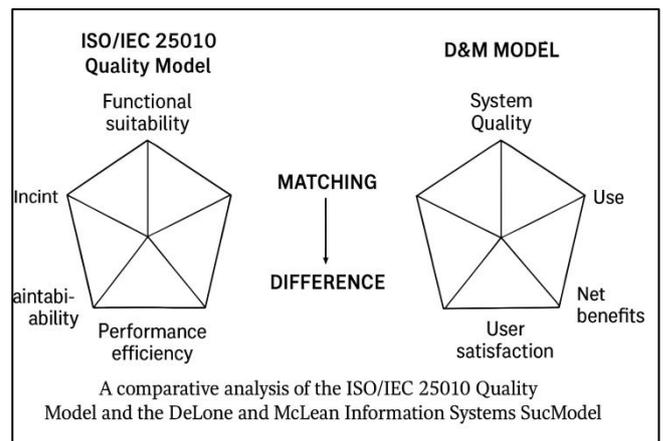


Figure 6. Matching and Difference Between ISO/IEC 25010 Quality Model and DeLone and McLean Information Systems Success Model (D&M)

6. Results and Study Findings

The research instrument underwent expert validation to ensure its rigor and suitability for measuring system success. The panel consisted of experts with over a decade of experience in both academic and professional domains, including roles as systems directors. Each expert held a doctoral degree from internationally recognized institutions. Their feedback and recommendations were thoroughly reviewed and incorporated into the final version of the instrument. The statistical validation results affirm the instrument’s reliability and construct validity. The AVE was recorded at 0.6, exceeding the recommended threshold of 0.5, thereby confirming convergent validity. The Composite Reliability (CR) value reached 0.7, indicating strong internal consistency. Cronbach’s Alpha was also 0.7, meeting the standard criterion for acceptable reliability. Additionally, Spearman’s rho_A exceeded 0.7, further supporting the robustness of the measurement model. Collectively, these metrics demonstrate that the instrument is both statistically sound and appropriate for evaluating system success.

Table. Instrument Illustration

Factor	Items Load	Items	Cronbach's Alpha	rho_A	CR	AVE
Quality of Information	0.6	The information outputs of my smart application (including on-screen and printed outputs) are complete.				
	0.8	The information outputs of my smart application (including on-screen and printed outputs) are concise and are easy to understand.				
	0.7	It is easy to find what I'm looking for when using my application.	0.6	0.6	0.7	0.8
	0.8	The information outputs of my smart application (including on-screen and printed outputs) are accurate and is free from errors.				
	0.7	My smart application provides the precise information I need.				
Quality of System	0.8	It is easy for me to become skillful by using my smart application.				
	0.7	In general, I find my application is easy to use.				
	0.7	My smart application is well integrated.	0.7	0.9	0.7	0.7
	0.9	My application has a short time lag between input and output of data as example (registration process).				
	0.8	My application has a short response time for on-line enquiry.				
Efficiency	0.6	It is possible to find in my smart application what I want in a reasonable time.				
	0.8	My smart application enables me to get on to it quickly.				
	0.7	My smart application does not use advertisements or unwanted plug-ins.	0.8	0.6	0.8	0.6
	0.7	I can access my smart application from my favorite browser.				
	0.7	It is easy to get and browse any part on my smart application.				
Functionality	0.7	It is easy to go to the home page while I'm browsing any other page in my smart application.				
	0.9	While using my smart application, I can easily navigate backwards through previously visited pages.	0.9	0.6	0.7	0.8
	0.8	My smart application provides varied search options (e.g. By faculty, courses, etc).				
	0.7	Search hints are provided when wrong search keywords are used.				

Reliability	0.9	My smart application never stops unexpectedly.				
	0.9	When there is a problem in some part or parts in my smart application I still can browse and perform some of process.				
	0.8	In case of interruption of fault, my smart application recovers properly.	0.9	0.9	0.9	0.9
	0.8	In general, my smart application is available 24/7.				
	0.8	I believe that my smart application is reliable.				
Usability	0.8	The interface design of my smart application is attractive.				
	0.7	All interface elements are well combined and harmonious in my application.				
	0.8	My smart application protects me from making errors when interrering with data.	0.7	0.8	0.9	0.9
	0.8	My smart application errors messages clearly indicate to me how to correct the problem.				
	0.7	In my smart application, it is easy to recover from the error quickly.				
Security	0.8	I believe my smart application is secure.				
	0.6	Overall, I trust my smart application.				
	0.8	My smart application has adequate security features that make you feel secure while using.	0.6	0.7	0.8	0.7
	0.7	I believe that the information offered on the smart application is sincere and honest.				
	0.7	The output information of my smart application is secure.				
Ease of Use	0.8	I find my smart application flexible to interact with.				
	0.7	My interactions with my smart application during doing online process were clear and understandable.				
	0.6	My smart application is convenient for me.	0.6	0.7	0.9	0.8
	0.8	My smart application is laid out in a modern and fashionable.				
	0.6	My smart application is of high quality.				



Figure 7. Quality Metrics Representation

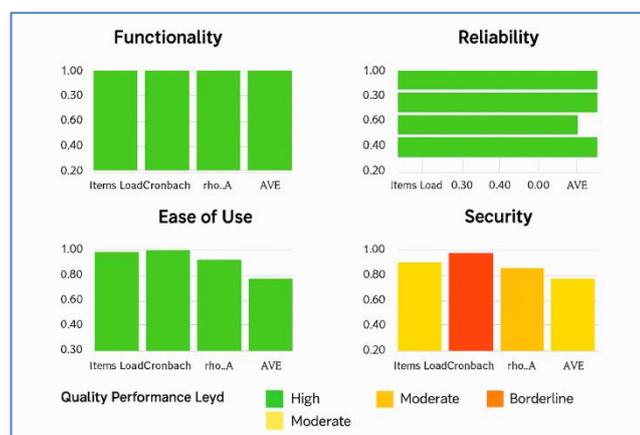


Figure 8. Quality Metrics Representation

Figures 7 and 8 present the quality metrics representation, clearly illustrating the evaluation results of the psychometric properties of survey instruments developed to measure perceived quality dimensions of a smart application, based on the ISO/IEC 25010 standard. Confirmatory Factor Analysis and reliability testing reveal a nuanced pattern of construct validity and reliability, with several factors demonstrating strong performance while others require substantial revision. The measurement model was assessed using factor loadings, CR, AVE, and Cronbach's Alpha. The construct Reliability emerged as the most robust, with all factor loadings exceeding 0.80 and demonstrating high internal consistency and convergent validity. This indicates that items measuring system stability and availability are both conceptually coherent and reliably captured. Functionality also performed well, with strong factor loadings and satisfactory convergent validity, suggesting that items related to navigation and search functionality effectively represent the construct.

Usability and Ease of Use displayed acceptable reliability, though factor loadings were moderately strong, implying that while these constructs are measurable, some items may benefit from refinement to improve their representativeness. Quality of Information and Efficiency

exhibited borderline acceptability, with one item in each scale falling below the preferred loading threshold. Quality of Information also showed lower internal consistency, indicating potential ambiguity in items intended to measure the completeness and conciseness of information.

Security was the weakest construct, with very low convergent validity and subpar reliability, suggesting that the items fail to collectively capture a unified perception of security among users. These findings partially align with the ISO 25010 framework but highlight significant challenges in operationalizing certain quality attributes, particularly Security and Quality of Information. The strong performance of Reliability and Functionality underscores their importance in users' evaluations of application quality, consistent with prior studies where system dependability and functional clarity are critical to user satisfaction. Conversely, the poor performance of Security indicates a potential misalignment between theoretical definitions and user interpretations. Users may conceptualize security in broader terms than anticipated, possibly encompassing privacy, trust, or system integrity, or the items may have been phrased in a way that introduced measurement error. The moderate reliability of Quality of Information suggests that items such as completeness and conciseness may be too abstract or overlapping in meaning from the user's perspective. The acceptable but mixed results for Efficiency and Usability further imply that while users can evaluate interactivity and performance, the items may require simplification or contextual examples to improve clarity and discrimination.

From a methodological perspective, this study indicates that not all quality dimensions in the ISO 25010 model are equally easy to measure through self-report surveys. Dimensions such as Reliability and Functionality can be assessed with high precision, whereas Security and Information Quality require more nuanced item development. The results advocate for iterative pretesting and validation of survey instruments, especially when applied in new cultural or technological contexts, to ensure items are interpreted as intended. For practitioners, these findings suggest that improvements in system reliability and functionality are likely to be most perceptible and valued by users. Effectively communicating and measuring security may require more than traditional Likert-scale items; incorporating behavioral or scenario-based questions could yield more valid assessments.

A key limitation of this study is the uneven performance of the constructs, which may reflect item wording, cultural interpretation, or sampling characteristics. Future research should qualitatively explore how users conceptualize problematic constructs such as Security and Information Quality, refine and revalidate the weak items possibly using structured interviews or cognitive pretesting, and examine the instrument's performance in a larger and more diverse

sample to enhance generalizability. While the instrument demonstrates promise for evaluating several key quality dimensions of smart applications, further refinement is necessary, particularly for the Security and Information Quality constructs, before it can be reliably deployed in academic or practical settings.

7. Conclusion

The study employed the term “Smart application” throughout its instrumentation and reporting; however, for conceptual clarity and to align with prevailing literature and industry practice, vendors should adopt a standardized label such as “AI-based mobile application” or “intelligent mobile app” when describing product capabilities, specifications, and marketing claims. This terminological precision matters because it reduces ambiguity for users, facilitates valid measurement of perceived intelligence, and enables clearer mapping between claimed technical features and observable system behaviors in empirical evaluations. Empirically, the present analysis identified substantial measurement weaknesses in the Security and Quality of Information constructs: both exhibited low internal consistency and poor convergent validity, indicating that the current items do not coherently capture the intended latent domains. Practically, companies developing mobile applications must therefore prioritize strengthening technical security measures (e.g., robust authentication, secure data storage, explicit consent flows) and ensure that these measures are translated into user-facing, behaviorally specific features that users can perceive and evaluate. Simultaneously, information quality should be operationalized through clear, disaggregated facets (accuracy, completeness, timeliness, conciseness) with concrete examples or anchors in the user interface (e.g., provenance labels, update timestamps, summarized vs. full-report views).

From an assessment standpoint, developers and researchers should (1) revise ambiguous or jargon-laden survey items into behaviorally anchored statements, (2) pilot test items with representative users to confirm interpretability, (3) combine objective technical metrics (logs, security audit results, latency/accuracy measures) with perceptual survey items to triangulate evidence, and (4) treat security and information-quality improvements as both engineering priorities and measurement tasks—documenting changes, revalidating instruments after revision, and reporting item-level loadings, reliability indices, and AVE so that subsequent findings reliably reflect real improvements in app trustworthiness and information value.

This study has established a validated instrument for measuring system success by integrating the ISO IEC 25010 standard and D&M. The psychometric evaluation demonstrated strong results with high item loadings, Cronbachs Alpha values above 0.7, satisfactory AVE scores, and rho A values exceeding 0.6. These findings confirm the instruments’ reliability and convergent validity across measured constructs. The research contributes to the field by providing a robust tool that captures both technical performance and user centric outcomes essential for

comprehensive system quality assessment. This instrument enables organizations to evaluate system success with greater precision while aligning with strategic objectives and user needs.

Future studies should focus on further validation of the instrument in diverse contexts and populations particularly examining its applicability to evolving technologies and cross-cultural environments. Additional refinement may strengthen constructs with lower discriminant validity while maintaining the frameworks overall robustness. This instrument serves as a foundation for developing more nuanced system quality evaluation frameworks that can adapt to increasingly complex digital ecosystems.

References

- [1] M. Jami Pour, J. Mesrabadi, and M. Asarian, “Meta-analysis of the DeLone and McLean models in e-learning success: the moderating role of user type,” *Online Inf. Rev.*, vol. 46, no. 3, pp. 590–615, 2022.
- [2] I. E. I. Fadhel, S. Z. Syed Idrus, A. A. E. Ali Ibrahim, M. B. Omar, S. K. Albzeirat, and F. Baheshwan, “Measuring system success in new context by adapting DM 2003 framework with the external factor management support,” *Journal of Physics: Conference Series*, vol. 12003, IOP Publishing, 2018.
- [3] I. E. Ibrahim Fadhel, S. Z. Bin Syed Idrus, M. S. Y. Abdullah, A. A. E. A. Ibrahim, M. Omar, and S. Saad, “Nias-mukalla web-based systems success measurement and students satisfaction evaluation based on security factor of systems quality engineering theory (ISO 25010) and other factors,” 2019.
- [4] I. E. I. Fadhel, *The Development of Web-based Systems Quality Engineering Framework in the Universities Domain*. School of Human Development and Technocommunication, 2019. [Online]. Available: <https://books.google.com/books?id=6wuuzQEACAAJ>
- [5] I. E. I. Fadhel, *An evaluation of information system success based on students’ perspective: The case of Hadhramout University*, M.S. thesis, Universiti Utara Malaysia, 2015.
- [6] A. A. I. Almetwaly and I. E. I. Fadhel, “Original Research Article Integrate between information systems engineering and software engineering theories for successful quality engineering measurement of software: Valid instrument pre-results,” 2024.
- [7] V. M. A. Edgar-Nevill, “Evaluation of the SEI software capability model within an information systems context; in pursuit of software quality,” *WIT Trans. Inf. Commun. Technol.*, vol. 8, 2025.
- [8] I. Pysmennyi, R. Kyslyi, and K. Kleshch, “AI-driven tools in modern software quality assurance: an assessment of benefits, challenges, and future directions,” *arXiv Prepr. arXiv2506.16586*, 2025.
- [9] F. Z. F. Oktavia, “Factors affecting user satisfaction on e-filing system in Indonesia,” *Kompartemen J. Ilm. Akunt.*, pp. 239–253, 2022.
- [10] A. Gormantara and E. Elisabeth, “Evaluation of the Success of the Academic Information System (SIAMIK) with the DeLone and McLean Model,” *J. Teknol. Inf. dan Pendidik.*, vol. 15, no. 2, pp. 99–109, 2022.
- [11] S. A. Melgis, R. Aryani, D. Lestari, and M. N. A.

- Abdulnazar, “Analyzing the Quality of Academic Information Systems on System Success,” *INTENSIF J. Ilm. Penelit. dan Penerapan Teknol. Sist. Inf.*, vol. 8, no. 1, pp. 144–165, 2024.
- [12] J. N. Amanda, W. A. Winarno, and A. T. Agustini, “Assessing the Success of Village Asset Management Systems: An Employee Perspective,” *Electron. J. Knowl. Manag.*, vol. 21, no. 3, pp. P174–190, 2023.
- [13] K. F. Brahmantyo, B. Paguna, and L. D. Prawati, “Measuring the success of corporate annual tax online reporting: Applying the DeLone & McLean Information System Success Model,” in *E3S Web of Conferences*, EDP Sciences, 2023, p. 1094.
- [14] T. Vuckovic, D. Stefanovic, D. Ciric Lalic, R. Dionisio, Â. Oliveira, and D. Przulj, “The extended information systems success measurement model: e-learning perspective,” *Appl. Sci.*, vol. 13, no. 5, p. 3258, 2023.
- [15] Z. Bao and Y. Zhu, “Understanding customers’ stickiness of live streaming commerce platforms: an empirical study based on modified e-commerce system success model,” *Asia Pacific J. Mark. Logist.*, vol. 35, no. 3, pp. 775–793, 2023.
- [16] J. Zhang, S. Zhong, T. Wang, H.-C. Chao, and J. Wang, “Blockchain-based systems and applications: a survey,” *J. Internet Technol.*, vol. 21, no. 1, pp. 1–14, 2020.
- [17] P. Khairunnisa and M. S. Setyowati, “Analysis of Factors Influencing User Satisfaction with The Benefits of The Information System In The SIPD-SIKD Application of The Special Capital Region of Jakarta,” *Asian J. Soc. Humanit.*, vol. 3, no. 1, pp. 98–116, 2024.
- [18] V. Jayakody and J. Wijayanayake, “Critical success factors for DevOps adoption in information systems development,” *Int. J. Inf. Syst. Proj. Manag.*, vol. 11, no. 3, pp. 60–82, 2023.
- [19] L. Pinedo, E. Panadero, J. Fernández-Ruiz, and C. Rodríguez-Hernández, “Students’ experiences in self-assessment: training, processes and feedback use in secondary and higher education,” *Assess. Educ. Princ. Policy Pract.*, vol. 30, no. 5–6, pp. 448–480, 2023.
- [20] V. Sarasi, I. Chaerudin, and I. A. Sundoro, “The DeLone and McLean model for measuring success in online learning systems: Indonesian evidence,” *J. Educ. Learn.*, vol. 17, no. 4, pp. 566–574, 2023.
- [21] S. B. Nauli, P. D. Nusantara, and A. Priambodo, “Academic Information System Success Model and Maturity Level Comparison for Improvement Recommendation,” *J. Ilm. FIFO*, vol. 14, no. 2, p. 179, 2023.
- [22] P. D. König, S. Wurster, and M. B. Siewert, “Sustainability challenges of artificial intelligence and Citizens’ regulatory preferences,” *Gov. Inf. Q.*, vol. 40, no. 4, p. 101863, 2023.
- [23] J. König *et al.*, “Teacher education effectiveness as an emerging research paradigm: a synthesis of reviews of empirical studies published over three decades (1993–2023),” *J. Curric. Stud.*, vol. 56, no. 4, pp. 371–391, 2024.
- [24] L. Hsu, “What makes good LMOOCs for EFL learners? Learners’ personal characteristics and Information System Success Model,” *Computer. Assist. Lang. Learn.*, vol. 36, no. 1–2, pp. 1–25, 2023.
- [25] W. H. DeLone and E. R. McLean, “Information Systems Success Measurement,” *Found. Trends® Inf. Syst.*, vol. 2, no. 1, pp. 1–116, 2016.
- [26] I. E. I. Fadhel, S. Z. B. S. Idrus, M. S. Y. Abdullah, A. A. E. A. Ibrahim, and M. B. Omar, “Systems success measurement: instrument & framework a new perspective,” *Indep. J. Manag. Prod.*, vol. 10, no. 5, pp. 1572–1606, 2019.
- [27] I. E. I. Fadhel, S. Z. B. S. Idrus, M. S. Y. Abdullah, A. A. E. A. Ibrahim, M. Omar, and A. Khred, “Towards Development a Novel Framework of Web-Based Systems Quality Engineering by the Integration between Information Systems and Software Engineering Theories: Context of Higher Education,” in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 22005.
- [28] C. J. Gelderman and R. J. Kusters, “Measuring information systems success: A comment on the use of perceptions,” in *Measuring organizational information systems success: new technologies and practices*, IGI Global, 2012, pp. 23–38.
- [29] A. D. Dzhalilova, R. A. Mannapova, I. G. Ivanova, S. E. Akopov, and K. Y. Tikhomirov, “Development of export Potential of Business in the AI Economy through Product Quality Management with The Help of Corporate information Systems,” *Proc. Eng.*, vol. 7, no. 1, pp. 137–146, 2025.
- [30] M. M. Hasan *et al.*, “Human-AI Collaboration in Software Design: A Framework for Efficient Co-Creation,” *AIJMR-Advanced Int. J. Multidiscipline. Res.*, vol. 3, no. 1, 2025.
- [31] G. Voria, S. Lambiase, M. C. Schiavone, G. Catolino, and F. Palomba, “From expectation to habit: Why do software practitioners adopt fairness toolkits?,” in *2025 IEEE/ACM 47th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, IEEE, 2025, pp. 94–105.
- [32] C. Gürkut and M. Nat, “Important factors affecting student information system quality and satisfaction,” *EURASIA J. Math. Sci. Technol. Educ.*, vol. 14, no. 3, pp. 923–932, 2017.
- [33] I. E. I. Fadhel, S. Z. B. S. Idrus, M. S. Y. Abdullah, A. A. E. A. Ibrahim, M. Omar, and A. Khred, “A New Perspective of Web-Based Systems Quality Engineering Measure by Using Software Engineering Theory (ISO 25010): An Initial Study,” in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 22004.
- [34] I. E. I. Fadhel, S. Z. B. S. Idrus, A. A. E. A. Ibrahim, and M. Omar, “Ibrahim’s triangle model (satisfaction–benefit–loyalty) for systems success measurement,” *Indep. J. Manag. Prod.*, vol. 10, no. 2, pp. 499–511, 2019.
- [35] I. E. I. Fadhel, S. Z. S. Idrus, A. A. E. A. Ibrahim, M. Omar, A. A. A. Bahashwan, and A. A. M. Al-Ansi, “An integration between information systems engineering and software engineering theories towards engineering a novel framework of web-based systems success for institutions based on students’ perceptions,” *J. Phys. Conf. Ser.*, vol. 1019, no. 1, p. 12081, 2018, [Online]. Available: <http://stacks.iop.org/1742-6596/1019/i=1/a=012081>

- [36] S. Mardiana, J. H. Tjakraatmadja, and A. Aprianingsih, "Validating the Conceptual Model for Predicting Intention to Use as Part of Information System Success Model: The Case of an Indonesian Government Agency," *Procedia Computer. Sci.*, vol. 72, pp. 353–360, 2015.
- [37] A. L. McNab and D. A. Ladd, "Information quality: the importance of context and trade-offs," in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, IEEE, 2014, pp. 3525–3532.
- [38] E. W. Mwangi, "Information quality assessment framework: case of the national safety net program single registry system," 2016, *Strathmore University*.
- [39] R. C. MacCallum, K. F. Widaman, K. J. Preacher, and S. Hong, "Sample size in factor analysis: The role of model error," *Multivariate Behav. Res.*, vol. 36, no. 4, pp. 611–637, 2001.
- [40] C. M. König, C. Karrenbauer, and M. H. Breitner, "Critical success factors and challenges for individual digital study assistants in higher education: A mixed methods analysis," *Educ. Inf. Technol.*, vol. 28, no. 4, pp. 4475–4503, 2023.
- [41] A. A. I. Almetwaly and I. E. I. Fadhel, "Integrate between information systems engineering and software engineering theories for successful quality engineering measurement of software: Valid instrument pre-results," 2024.
- [42] J. De Leeuw and M. Lovric, "Statistical Software: Overview," *Int. Encycl. Stat. Sci.*, pp. 2582–2588, 2025.
- [43] C.-C. Huang, Y.-M. Wang, T.-W. Wu, and P.-A. Wang, "An empirical analysis of the antecedents and performance consequences of using the moodle platform," *Int. J. Inf. Educ. Technol.*, vol. 3, no. 2, p. 217, 2013.
- [44] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *J. Mark. Res.*, pp. 39–50, 1981.
- [45] J. F. Hair, *A primer on partial least squares structural equation modeling (PLS-SEM)*. sage, 2014.
- [46] A. A. I. Almetwaly and I. E. I. Fadhel, "Integrate between information systems engineering and software engineering theories for successful quality engineering measurement of software: Valid instrument pre-results," *Computer. Software. Media Appl.*, vol. 6, no. 1, p. 3382, 2024.
- [47] G. D. Garson, "Computerized simulation in the social sciences: A survey and evaluation," *Simul. Gaming*, vol. 40, no. 2, pp. 267–279, 2009.
- [48] G. D. Garson, *Hierarchical linear modeling: Guide and applications*. Sage, 2013.
- [49] S. A. Bustin et al., "The MIQE Guidelines: M minimum Information for Publication of Quantitative Real-Time PCR Experiments," 2009, *Oxford University Press*.
- [50] J. Pallant, *SPSS survival manual*. McGraw-Hill Education (UK), 2013.

الذكاء الاصطناعي في تطبيقات الهواتف المحمولة لشركات الصرافة: قياس النجاح من خلال هندسة الجودة - أداة تحقق صالحة من منظور المستخدمين في مدينتي الشحر والمكلا، اليمن

المخلص: تعد أدوات التقييم الحالية غير كافية في كثير من الأحيان لقياس فعالية هندسة الجودة في الأنظمة، لا سيما تلك التي تعتمد على تقنيات الذكاء الاصطناعي. ومن هذا المنطلق، يُصبح من الضروري امتلاك فهم شامل لمفهوم نجاح النظام. تستهدف هذه الدراسة تطوير أداة جديدة ومُعتمدة تجريبياً لقياس نجاح جودة النظام، وذلك من خلال دمج الأطر النظرية المستمدة من مجال نظم المعلومات وتحديداً نموذج ديون وماكلين ومجال هندسة البرمجيات كما ورد في معيار ISO 25010 ولضمان الصرامة المنهجية والمصادقية العلمية للأداة، تم تشكيل لجنة أولية من عشرة خبراء أكاديميين، ثم تم توسيعها لاحقاً لتضم ثلاثة عشر خبيراً، بهدف تعزيز دقة بناء الأداة وتقييمها. وتُعد هذه الأداة أول محاولة منهجية لتقييم نجاح الأنظمة من خلال منظور تكاملي يجمع بين نظم المعلومات وهندسة البرمجيات، مع تركيز خاص على تطبيقات الهواتف المحمولة المعتمدة على الذكاء الاصطناعي. وبعد الانتهاء من مرحلة التحقق من صلاحية الأداة عبر الخبراء، تم اختبارها تجريبياً على عينة مكونة من 105 مستخدمين من اليمن. وقد أظهرت نتائج التحليلات الإحصائية مستوى عالياً من الموثوقية والصلاحية، مما يؤكد فعالية الأداة في قياس نجاح جودة أنظمة الجوال.