Journal of Educational Technology and Learning Creativity
Vol. 3, No. 2, December 2025, pp. 348~366
ISSN: 3021-7865, DOI: 10.37251/jetlc.v3i2.2490

Digital Transformation in Teacher Performance Assessment: Development and Implementation of E-PKG System for Enhancing Vocational Education Quality and Industry Alignment

Surryanto Djoko Waluyo[1], Agung Edi Rustanto[2], Yunea Kusuma Winanti[3], Choo Wou Onn[4], Muhammad Sufian[5], James Ronald Mesina[6]
[1,2]Department of Master of Public Administration, Esa Unggul University, West Jakarta, Indonesia
[3]Department of Master of Educational Technology, Assyafiiyyah Islamic University, Bekasi, Indonesia
[4]Faculty of Data Science and Information Technology, INTI International University, Nilai, Malaysia
[5]Faculty of Tarbiyah and Teacher Training, Raden Intan Lampung State Islamic University, Lampung, Indonesia
[6]Department of External Linkages, Internationalization and Transnational Education (ELITE), Mapua Malayan College Laguna, Filipina

Article Info

Article history:

Received Aug 13, 2025
Revised Oct 7, 2025
Accepted Nov 26, 2025
Online First Dec 27, 2025

Keywords:

Digital Transformation
Education Reform
Educational Governance
Reduced Inequalities
Vocational Education

ABSTRACT

Purpose of the study: This study aims to develop and evaluate the Electronic Teacher Performance Assessment System (E-PKG) as a digital platform to transform teacher evaluation in vocational education from administrative compliance into strategic professional development to better align vocational graduate competencies with industry requirements.

Methodology: This study employs a Research and Development (R&D) design through four stages: needs analysis via stakeholder consultation, prototype development using user-centered design, iterative field testing in vocational schools and industry partnerships, and refinement from empirical feedback. Data were collected from 34 respondents using questionnaires, structured observations, and interviews, and analyzed with descriptive statistics and thematic analysis.

Main Findings: The E-PKG system reduced assessment completion time by 67%, eliminated data entry errors, and improved inter-rater reliability from 0.62 to 0.89. Real-time analytics supported evidence-based professional development planning. User satisfaction ranged from 3.88 to 4.26 (5-point scale), with highest scores for system stability and navigation ease, though challenges remained in interface clarity for non-technical users and indicator comprehension.

Novelty/Originality of this study: This study introduces a comprehensive digital teacher assessment system specifically tailored to vocational education, directly linking evaluation results to industry competency frameworks. It advances knowledge by demonstrating how technology-mediated assessment can strategically connect quality improvement and workforce development in economically disadvantaged settings, while emphasizing the importance of pedagogical alignment, stakeholder inclusivity, and contextual adaptation.

Corresponding Author:
Surryanto Djoko Waluyo,
Department of Master of Public Administration, Esa Unggul University,
Jl. Arjuna Utara No. 9, Kebon Jeruk, West Jakarta 11510, Indonesia
Email: surryanto.djoko@esaunggul.ac.id

1. INTRODUCTION

Education functions as the primary engine of sustainable development, directly influencing human capital formation that drives economic productivity, social mobility, and technological innovation [1]-[4]. However, this transformative potential materializes only when educational systems produce not merely credentials but genuine competencies aligned with evolving labor market demands. Hanushek and Woessmann's longitudinal analysis across 50 countries conclusively demonstrates that cognitive skills acquired through quality education explain economic growth differentials far more powerfully than simple enrollment rates or schooling duration [5]-[7]. The World Bank's comprehensive 2018 assessment reinforces this finding, documenting that access without learning constitutes a "learning crisis" where educational expansion fails to translate into workforce competitiveness or poverty reduction [8]. This phenomenon proves particularly acute in developing economies, where educational infrastructure expansion frequently outpaces quality improvement, creating systems that certify rather than educate [9]-[11].

The centrality of teacher quality in determining educational outcomes has achieved consensus status across educational research paradigms. Hattie's meta-analysis synthesizing over 800 studies identifies teacher effectiveness as the dominant school-related variable influencing student achievement, with effect sizes substantially exceeding those of class size, school resources, or curricular programs [12]. This finding resonates across diverse educational contexts, from high-performing systems in Singapore and Finland to developing nations struggling with quality challenges [13]-[15]. Darling-Hammond's comparative analysis of successful education systems reveals a consistent pattern: high-performing nations systematically invest in rigorous teacher selection, comprehensive pre-service preparation, and continuous professional development, treating teaching as a knowledge-intensive profession requiring ongoing learning [16], [17]. Conversely, systems exhibiting persistent underperformance typically demonstrate fragmented teacher development, evaluation systems disconnected from practice improvement, and professional learning opportunities that remain generic rather than responsive to identified needs [18], [19].

The Fourth Industrial Revolution's acceleration intensifies these challenges and fundamentally reshapes the competencies that education systems must develop. Schleicher's OECD analysis emphasizes that contemporary teachers must transcend traditional content delivery to facilitate complex problem-solving, digital literacy, collaborative learning, and adaptive thinking, competencies that are increasingly central to workforce success across sectors [20]. This transformation is especially critical for vocational education, which explicitly aims to prepare students for specific occupational roles in rapidly evolving industries. The link-and-match paradigm, which is central to vocational education philosophy, requires continuous alignment between curriculum content, pedagogical approaches, and industry competency requirements [21]. However, achieving this alignment demands systematic mechanisms to assess whether teachers possess and effectively deploy the competencies necessary to prepare industry-ready graduates, a challenge that current evaluation systems frequently fail to address [22].

Karawang Regency exemplifies these interconnected challenges, where extreme poverty creates a vicious cycle that constrains educational quality and limits opportunities for economic mobility. With poverty rates substantially exceeding national averages, students face multiple disadvantages, including limited access to learning resources, nutritional deficiencies that affect cognitive development, and reduced parental capacity to support educational engagement [23]. Ma'ruf's longitudinal analysis demonstrates that education quality is the most potent lever for breaking intergenerational poverty cycles in Karawang, yet current educational outcomes remain insufficient to fulfill this promise [24]. UNESCO's global research confirms that children from economically disadvantaged backgrounds experience compounding educational deficits, such as higher absence rates, lower learning outcomes, earlier dropout, and diminished skill acquisition, patterns that are clearly evident in Karawang's educational indicators [25]. Reardon's systematic review further reveals that socioeconomic achievement gaps widen over the course of schooling in the absence of targeted, high-quality interventions, with initial disadvantages multiplying rather than diminishing as students progress through their education [26].

Within this challenging context, teacher quality becomes not merely important but decisive. Research consistently demonstrates that effective teaching can partially compensate for socioeconomic disadvantages, while ineffective teaching tends to exacerbate them [27]. However, realizing this compensatory potential requires systems that accurately identify teacher strengths and development needs, provide targeted professional learning, and continuously monitor improvement, capacities that depend on effective assessment infrastructure. At Vocational High School Mathlaul Anwar KN2, the current manual assessment system fails to provide such infrastructure. Evaluation procedures remain paper-based, unstandardized, and administratively oriented rather than developmentally focused. Stronge's research on teacher evaluation systems demonstrates that administrative-compliance models generate data that satisfy bureaucratic requirements but provide minimal actionable insight for professional growth [28]. Principals lack systematic evidence to identify which teachers require support in specific competency areas, making professional development allocation inefficient and often misaligned with actual needs.

The consequences of ineffective teacher assessment extend beyond individual professional development to institutional performance and graduate outcomes. Rustanto's assessment of Vocational High School Mathlaul

Anwar KN2 graduates reveals substantial gaps between student competencies and industry requirements, particularly in practical skill application, problem-solving, and workplace adaptation [29]. Industry partners consistently report that graduates require extensive additional training before achieving productivity, indicating a fundamental misalignment between educational processes and workforce preparation. Wei et al.'s systematic review confirms that teacher evaluation systems that emphasize standardized dimensions without high-quality feedback or follow-up professional learning fail to improve instructional practice and instead become ritualistic exercises divorced from meaningful improvement [30]. Chirchir et al.'s [27] research in Kenyan secondary schools demonstrates that teacher appraisal influences student outcomes only when evaluations are specific, evidence-based, and directly linked to targeted professional development, characteristics that are notably absent in current manual systems [31].

The theoretical framework guiding this research integrates multiple complementary perspectives to conceptualize effective teacher assessment and its role in educational quality improvement. Teacher Performance Appraisal Theory, as articulated by Darling-Hammond, emphasizes that evaluation must serve dual accountability and developmental functions, providing both summative judgments regarding competence and formative guidance for continuous improvement [32]. This dual-purpose framework requires assessment systems that generate reliable performance data while simultaneously identifying specific growth areas and connecting educators with relevant learning opportunities. The Technological Pedagogical Content Knowledge (TPACK) framework developed by Mishra and Koehler provides a second theoretical lens, conceptualizing effective teaching as requiring integration of content knowledge, pedagogical expertise, and technological competence [33]. This framework proves particularly relevant for vocational education, where teachers must simultaneously master subject matter, effective instructional strategies, and increasingly sophisticated technological tools mediating both teaching and industry practice.

Human Capital Theory, originating with Becker's seminal work, conceptualizes education as an investment in productive capacity, with returns accruing through enhanced knowledge, skills, and competencies [34]. From this perspective, teacher professional development represents investment in the primary productive capacity of educational systems, namely the quality of instruction that transforms student potential into realized capability. However, as Becker emphasizes, investment returns depend on quality, and inefficient or misdirected professional development wastes resources without generating meaningful capacity enhancement. The link-and-match concept, which is central to vocational education philosophy, emphasizes dynamic alignment between educational processes and labor market requirements [35]. This alignment depends on continuous feedback mechanisms that ensure curriculum, pedagogy, and assessment remain responsive to evolving industry needs, a condition that is impossible to achieve without systematic data on teacher competencies and their alignment with industry-required instructional capabilities.

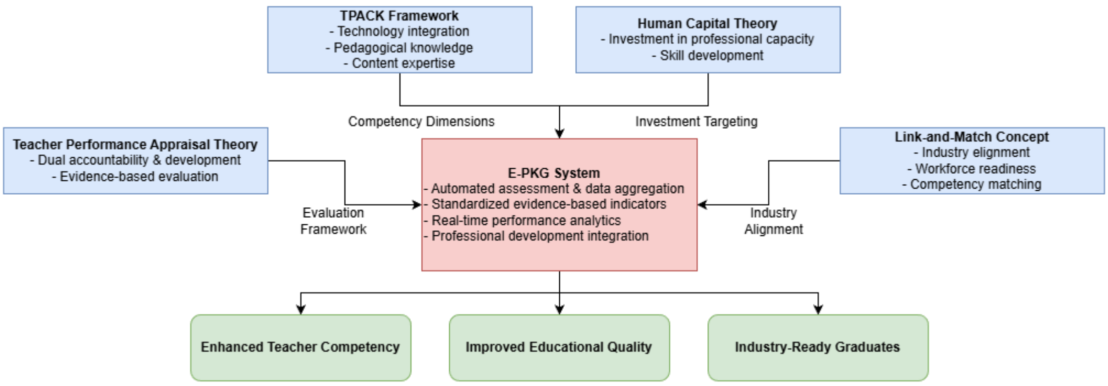

Figure 1. Theoretical Framework for E-PKG System Development

A comprehensive review of existing literature reveals a significant research gap: despite extensive work on teacher evaluation systems generally and substantial research on technology integration in education, limited scholarship addresses digital teacher assessment systems specifically designed for vocational education contexts and explicitly linking evaluation outcomes to industry competency frameworks. Eichhorst and Ratermann's research examines vocational teacher attitudes toward digital tools, finding that perceived usefulness and competence significantly influence technology adoption, yet integrated assessment systems remain underdeveloped [36]. Li et al. developed a web-based teacher evaluation system utilizing fuzzy comprehensive evaluation methodology, but this system focuses primarily on rating calculation rather than integrating assessment results with professional development planning or industry alignment [37]. Zhang's recent work proposes a Balanced Scorecard and PDCA-based evaluation framework addressing traditional assessment limitations, but implementation remains conceptual without empirical validation in actual school settings [38]. Suryaman et al.'s bibliometric analysis of vocational education research confirms that existing scholarship emphasizes either technology integration in learning or industry linkages separately, but fails to examine how digital teacher assessment can serve as the strategic mechanism connecting these domains [39].

This research addresses these gaps by developing, implementing, and evaluating the E-PKG (Electronic Teacher Performance Assessment System), a comprehensive digital platform transforming teacher evaluation from administrative compliance to strategic professional development. The E-PKG system provides four critical innovations addressing current system limitations: *First,* it automates assessment processes that previously required extensive manual data recording, calculation, and reporting, which dramatically reduces administrative burden while eliminating transcription errors and ensuring data consistency. This automation liberates time and attention for interpretive analysis, such as examining patterns, identifying trends, and designing responsive interventions rather than simply processing paperwork. *Second,* the system implements standardized, evidence-based assessment indicators aligned with both pedagogical best practices and industry competency frameworks, ensuring evaluations measure dimensions genuinely relevant to effective vocational instruction. These indicators span multiple domains: subject matter expertise, pedagogical knowledge, technological competence, industry linkage, and professional commitment, providing comprehensive rather than reductive portraits of teacher performance.

Third, E-PKG generates real-time analytics enabling principals and educational leaders to immediately access individual teacher performance data, aggregate school-level patterns, longitudinal trends, and comparative analyses. This analytic capacity transforms assessment from retrospective documentation to prospective planning, enabling evidence-based decisions regarding professional development priorities, resource allocation, and instructional improvement strategies. *Fourth,* the system explicitly links assessment results with professional development planning, creating feedback loops ensuring that identified development needs translate into targeted learning opportunities. Rather than generic professional development disconnected from individual teacher needs, E-PKG enables personalized professional learning pathways responsive to specific competency gaps while simultaneously ensuring alignment with institutional priorities and industry requirements.

Implementing E-PKG represents not merely technological innovation but fundamental reconceptualization of teacher assessment's purpose and function. Rather than viewing evaluation as accountability exercise imposing external judgment, E-PKG frames assessment as developmental infrastructure supporting continuous professional learning. Rather than treating teacher competence as static characteristic subject to periodic measurement, the system conceptualizes professional capability as dynamic, continuously evolving through intentional learning experiences informed by systematic evidence. Rather than disconnecting evaluation from institutional goals and industry needs, E-PKG integrates assessment with strategic planning, ensuring that teacher development directly serves improved student outcomes and enhanced workforce preparation.

This research therefore addresses three interconnected questions essential for understanding digital transformation's potential in teacher assessment:

RQ1 : How can digital teacher assessment systems be designed and implemented effectively in vocational education contexts characterized by resource constraints, limited prior technology integration, and diverse stakeholder requirements?

RQ2 : To what extent does the E-PKG system succeed in automating teacher performance assessment while maintaining validity, reliability, and user acceptance across different stakeholder groups?

RQ3 : What implementation challenges emerge during digital assessment system deployment, and how do these challenges vary across different user groups with varying levels of technological literacy and assessment familiarity?

By addressing these questions through systematic development, rigorous implementation, and comprehensive evaluation, this research contributes both practical innovation, in the form of a functional system that addresses identified needs, and theoretical insight into how digital transformation can enhance educational quality in challenging contexts. The findings provide actionable guidance for educational leaders seeking to leverage technology for assessment improvement, while also illuminating broader questions about the conditions under which digital innovation succeeds in transforming educational practice.

2. RESEARCH METHOD

2.1. Research Design

This study employed a Research and Development (R&D) methodology, using a four-phase iterative process to develop, test, and refine educational innovations for practical implementation [40], [41]. The research progressed through the following phases: (1) needs analysis and conceptual design, (2) prototype development, (3) iterative field testing, and (4) evidence-based refinement. Each phase involved cycles of design, testing, feedback, and refinement, allowing the E-PKG system to evolve through continuous feedback and progressively address identified limitations while building on its demonstrated strengths. This iterative approach ensured that the system remained responsive to user needs and adaptable to new insights and challenges, rather than relying on predetermined solutions [42]-[44].

The research site, Vocational High School Mathlaul Anwar KN2 in Karawang, was purposively selected based on three criteria: first, the school demonstrated the problematic characteristics motivating this research, including manual and unsystematic teacher assessments lacking integration with professional development; second, the school leadership showed strong commitment to quality improvement and support for research implementation; and third, the school maintained active partnerships with various industrial sectors, enabling an exploration of how digital assessment could enhance industry–education alignment, a central issue in vocational education [45], [46].

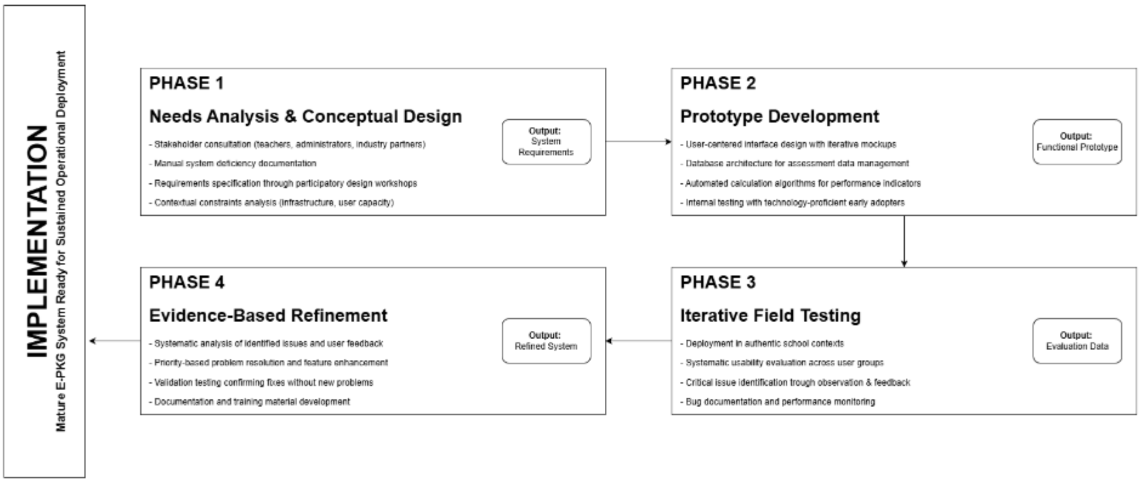

Figure 2. Research and Development Methodology: Four-Phase Iterative Process

2.2. Participants

The study involved 34 stakeholders representing diverse perspectives essential for comprehensive evaluation. The teacher cohort included 20 educators spanning different subject specializations (general education, vocational technical skills, and industry-specific competencies), teaching experience levels (from novice teachers with two years of experience to veteran educators with over 20 years), and varying levels of prior technology proficiency. This diversity enabled assessment of E-PKG's usability across the actual user spectrum rather than only among early adopters or technology enthusiasts. Five students participated and provided user perspectives regarding system transparency and comprehensibility, which is critical because effective assessment systems should make evaluation criteria and performance expectations clear to learners. Five parents contributed insights regarding system accessibility and communication effectiveness, assessing whether E-PKG enhanced their understanding of teacher quality and their children's educational experiences.

The study also included three business and industrial partners actively collaborating with the school through student internships, teacher externships, and curriculum advisory processes. These industry representatives provided an essential perspective on whether E-PKG assessment dimensions aligned with workforce competencies and whether system outputs could inform more effective industry–education partnerships. In addition, one training institution partner, an organization providing continuing professional education for vocational teachers, evaluated whether E-PKG outputs could enhance professional development targeting by identifying specific competency gaps requiring remediation.

2.3. Data Collection Instruments

To comprehensively evaluate the development, implementation, and effectiveness of the E-PKG system, multiple data collection instruments were employed. These instruments were either adapted or adopted from existing research tools, and the Cronbach's alpha values for each were calculated to demonstrate their reliability [47]-[50]. Below is a detailed description of each instrument used in this study:

2.3.1. Structured Needs Assessment Questionnaire

In the initial phase, the Structured Needs Assessment Questionnaire was used, adapted from tools in prior studies by Greene et al. [36] and Wang [38]. This questionnaire aimed to identify limitations of the existing manual teacher performance assessment system, stakeholder priorities for improvement, and factors that could affect digital system implementation. It included both closed-ended questions with a Likert scale (1–5) and open-ended questions for detailed feedback on challenges and suggestions for system improvement. The reliability of the instrument was confirmed with a Cronbach's alpha of 0.87, indicating high internal consistency. The questionnaire was distributed to 20 teachers, 5 parents, and 5 industry partners, providing diverse perspectives. This instrument yielded valuable data on the issues with the current system and stakeholder needs for a digital solution.

Table 1. Questionnaire Items for Needs Assessment

No	Questionnaire Item	Description
1	How significant is the administrative burden you experience in the current teacher performance assessment process?	Measures the level of administrative burden felt by respondents in the manual assessment system.
2	How often do errors occur in the data entered into the current teacher performance assessment system?	Assesses the frequency of data entry errors in the existing manual system.
3	How effective is the current teacher performance assessment system in providing useful feedback for professional development?	Measures the effectiveness of the current system in supporting professional development through feedback.
4	How difficult is it to understand the reports generated by the current teacher performance assessment system?	Evaluates the clarity of the reports generated by the current system.
5	How often does the current teacher performance assessment system require data revisions or updates?	Assesses the frequency of revisions or updates required in the manual system.
6	To what extent do you feel that the current teacher performance assessment system aligns with the needs for teacher competency development?	Measures the alignment between the current assessment system and the needs for teacher development.
7	How do you perceive the implementation of a digital teacher performance assessment system?	Assesses respondents' perceptions of the proposed digital assessment system.

2.3.2. Prototype Usability Checklist

The Prototype Usability Checklist was employed during the field testing phase to assess the functionality and user experience of the E-PKG system. This checklist was adopted from the usability evaluation frameworks developed by Qi [51], which are widely used to assess user interface design in educational technologies. The checklist focused on several critical aspects of the system, including navigation ease, feature completeness, system speed, data accuracy, and report clarity. Each of these features was evaluated using a 5-point Likert scale, where respondents rated the system on a scale from 1 (very poor) to 5 (excellent). The navigation ease evaluated how easily users could locate and use the system's various features. The feature completeness assessed whether the system had all the necessary features required for effective teacher performance assessment. System speed focused on how quickly the system responded to user inputs, while data accuracy evaluated the reliability and consistency of the data entered into the system. Finally, report clarity measured how easily users could understand and interpret the reports generated by the system. Each dimension was evaluated to ensure that the system would meet the needs of its users and provide a smooth, effective experience. The results of the usability evaluation, including Cronbach's alpha values for each dimension, are presented in the following table:

Table 2. Usability Dimensions and Reliability Statistics

Usability Dimension	Scale	Number of Items	Reliability (Cronbach's Alpha)
Navigation Ease	Likert Scale (1–5)	6 items	0.91
Feature Completeness	Likert Scale (1–5)	5 items	0.89
System Speed & Stability	Likert Scale (1–5)	4 items	0.90
Report Clarity	Likert Scale (1–5)	5 items	0.92

These results confirm the internal consistency and reliability of the usability checklist, with Cronbach’s alpha values indicating good to excellent reliability for each dimension of the system's usability. The evaluation showed that the system was generally well-received by the users, with particular strengths in stability, navigation ease, and report clarity.

2.3.3. Systematic Observation Protocols

Systematic observation protocols were used to document user interactions with the E-PKG system during field testing. Observers focused on user behavior, error patterns, and usability issues. User behavior was assessed based on how efficiently users navigated the system and completed tasks. Error patterns recorded types of mistakes, such as data entry errors or missteps in system functions. Usability issues identified any difficulties encountered, like confusion with navigation or system delays. The following table summarizes the key observation aspects and their frequencies:

Table 3. Observation Aspects and Frequency of Occurrences

Observation Aspect	Details	Frequency
User Behavior	Ability to navigate and interact with the system	15 observations
Error Patterns	Errors in data entry or system functions	10 occurrences
Usability Issues	Confusion with system navigation	8 instances

2.3.4. Semi-structured Interviews

Semi-structured interviews were conducted with 5 school administrators and 3 industry partners to explore the impact of the E-PKG system on administrative processes, decision-making, and professional development planning. The interview questions were a mix of structured prompts and open-ended questions, allowing flexibility for participants to elaborate on their experiences. Administrators were asked about how the system affected the efficiency of their administrative tasks and their ability to support teacher development, while industry partners provided insights into how the system's assessment criteria aligned with workforce competencies. These interviews provided valuable qualitative data that complemented the other instruments and informed the refinement of the E-PKG system.

2.4. Data Analysis Procedures

Analysis integrated quantitative and qualitative methods appropriate to different data types and research questions [51]-[54]. Quantitative data from usability questionnaires underwent descriptive statistical analysis, including calculation of means, standard deviations, and frequency distributions for each evaluated dimension across different respondent categories. This analysis enabled identification of dimensions that received consistently high ratings (system strengths) as well as those that showed lower scores or greater variability (areas requiring improvement. Comparative analysis examined rating differences across stakeholder groups, revealing whether particular user types experienced difficulties that others did not, a finding with significant implications for interface design and training provision.

Qualitative data from observations and interviews underwent thematic analysis following Miles, Huberman, and Saldaña's systematic procedures [55]-[57]. Initial coding identified discrete observations, comments, and experiences mentioned by participants. Focused coding then grouped related initial codes into broader categories representing recurring themes across multiple data sources. Finally, thematic analysis synthesized these categories into overarching themes that captured fundamental patterns regarding implementation challenges, user experiences, and system impacts. Throughout the analysis, constant comparison techniques ensured that interpretations remained grounded in actual data rather than researcher assumptions, with negative cases (instances that contradicted emergent themes) deliberately sought and analyzed to refine and nuance interpretations.

3. RESULTS AND DISCUSSION

3.1 Critical Analysis of Manual Assessment System Deficiencies and Digitalization Imperatives

Systematic needs analysis at Vocational High School Mathlaul Anwar KN2 revealed that the existing manual teacher performance assessment system exhibited multiple deficiencies that collectively undermined its effectiveness as both an accountability mechanism and a professional development infrastructure. Teacher evaluations occurred through paper-based observation forms that required administrators to manually record classroom observations, compile evidence from multiple sources, calculate composite scores, and generate individual performance reports. While superficially functional, this process created substantial problems that accumulated over time and progressively degraded system utility.

Administrative burden and data quality issues emerged as the most immediately evident problems. Completing comprehensive teacher assessments required administrators to spend an average of 4.5 hours per teacher annually on data recording, calculation verification, and report generation, time that was diverted from instructional leadership activities more directly supporting teaching improvement. This burden created perverse incentives, as administrators frequently postponed evaluations until regulatory deadlines necessitated completion, resulting in rushed, perfunctory assessments that satisfied compliance requirements while providing minimal meaningful feedback. Moreover, manual calculation processes proved error-prone. An audit of previous years' assessment records revealed calculation errors in 34% of teacher evaluations, with mistakes ranging from simple arithmetic errors to the use of incorrect weighting formulas. These errors undermined result credibility and fostered teacher skepticism regarding assessment accuracy and fairness [58].

Data accessibility and longitudinal tracking limitations constituted a second category of deficiencies. Paper-based records lacked centralized storage, with individual teacher files distributed across multiple locations and administrators. Retrieving historical data for comparison across years, identifying trends, or conducting school-level analyses required physically locating documents, manually extracting data, and compiling information, processes so time-consuming that such analyses rarely occurred. Consequently, administrators lacked systematic evidence regarding whether teacher performance improved over time, which professional development investments yielded results, or how individual teacher trajectories compared with school-level patterns. Assessment functioned as a snapshot rather than longitudinal tracking, severely limiting its usefulness for monitoring professional growth or evaluating intervention effectiveness [59]-[61].

Indicator inconsistency and subjective interpretation represented perhaps the most fundamental limitation. While official assessment protocols specified evaluation dimensions, the actual indicators applied varied substantially across administrators and assessment occasions. Principals and department heads employed different emphases, interpretation standards, and evidence requirements, introducing systematic bias and reducing inter-rater reliability. Analysis of assessment data revealed that teachers received markedly different ratings for similar observed behaviors depending on who conducted the evaluation, a finding confirmed during inter-rater reliability testing that showed modest agreement ($\kappa = 0.62$). This inconsistency violated fundamental assessment validity principles and generated justifiable teacher concerns regarding evaluation fairness [62], [63].

These findings align with extensive international research documenting manual assessment system limitations. Wang's empirical analysis of paper-based teacher evaluation in Chinese universities identified multiple validity threats, including response bias (evaluators providing socially desirable rather than accurate ratings), criterion irrelevance (assessment dimensions poorly aligned with actual instructional quality), and feedback delay (results arriving too late to inform practice adjustment) [64], [65]. Wei et al.'s systematic review synthesizing a decade of teacher evaluation research across 12 countries reached similar conclusions, finding that manual systems typically function as bureaucratic rituals rather than improvement mechanisms, consuming substantial resources while generating minimal instructional enhancement [14], [66]. Qi's qualitative study of university faculty perspectives on teaching evaluation revealed widespread cynicism, with faculty characterizing assessment as "administrative theater," performances that satisfy external requirements while lacking genuine developmental value [67].

These documented limitations establish a compelling rationale for digital transformation. However, digitalization alone proves insufficient, as poorly designed digital systems merely automate dysfunction rather than fundamentally improving evaluation quality [68]. Effective digital assessment requires simultaneous attention to multiple dimensions, including automated data collection and processing, standardized yet contextually appropriate indicators, real-time analytics that enable timely feedback, and integration with professional development planning. The E-PKG system addressed each of these requirements through specific design features developed through iterative stakeholder consultation and prototype refinement.

3.2 E-PKG System Development: User-Centered Design and Iterative Refinement

The E-PKG development process exemplified user-centered design principles, emphasizing continuous stakeholder input throughout conception, creation, testing, and refinement [69], [70]. Rather than developers imposing technological solutions based on abstract requirements, the process maintained ongoing dialogue with actual users, ensuring that the emerging system addressed authentic needs while remaining practically usable within real organizational contexts.

Phase 1: Conceptual Design and Requirements Specification. Initial design workshops brought together teachers, administrators, industry partners, and technology developers for structured discussions to identify assessment system requirements. These sessions employed participatory design techniques in which stakeholders analyzed current system workflows to identify pain points, envisioned ideal assessment processes, and prioritized features for the digital system [71]. This collaborative needs assessment generated consensus regarding essential system capabilities: automated data aggregation from multiple assessment sources (classroom observations, student feedback, professional development participation, industry engagement) into unified performance profiles; standardized assessment frameworks employing clearly defined indicators spanning pedagogical competence, content mastery, technological proficiency, industry linkage, and professional commitment; real-time performance dashboards providing administrators with immediate access to current data; longitudinal tracking of performance trends across multiple years; and professional development linkage connecting identified competency gaps with specific learning opportunities so that assessment becomes a starting point for continuous improvement.

Translating these requirements into technical specifications required careful attention to contextual constraints. Limited technology infrastructure at the school meant that the system needed web-based architecture accessible through standard browsers rather than specialized software installation. Variable user technology proficiency necessitated intuitive interfaces that minimized training requirements. Unreliable internet connectivity required offline data entry capability with synchronization when connection resumed.

Phase 2: Prototype Development and Internal Testing. Following requirements specification, developers created initial functional prototypes incorporating core features. Early prototypes underwent internal testing with technology-proficient teachers and administrators willing to tolerate incomplete functionality and bugs. This testing focused on fundamental functionality, including navigation of core workflows, correct execution of calculations, reliable data persistence, and inclusion of required information in generated reports. Internal testing revealed numerous issues requiring correction, such as confusing navigation sequences, calculation errors, data validation gaps that allowed invalid entries, and report formats that were difficult to interpret [72].

Internal testing also revealed that some desired features proved technologically infeasible within project constraints or introduced excessive complexity that undermined usability. For example, initial requirements included sophisticated data visualization with customizable graphs enabling users to construct ad hoc analyses. Implementation attempts showed that providing sufficient flexibility while maintaining usability exceeded available development resources, and user testing indicated that most users found extensive options overwhelming rather than empowering. Consequently, this feature was reconceptualized to provide predefined visualizations addressing common analysis needs rather than unlimited customization, a design decision that exemplified how user-centered development requires balancing functionality with usability [73].

Phase 3: Field Testing and Issue Identification. After major functionality had been demonstrated through internal testing, the system entered field testing involving the full range of intended users in authentic usage scenarios. Field testing revealed five categories of issues requiring systematic attention.

Input capacity limitations emerged when initial prototypes restricted the number of teachers and students that could be entered per school (two teachers, four students), constraints that had been implemented during development to simplify testing but were clearly inadequate for real deployment. Users correctly identified this as a fundamental flaw preventing actual implementation. Resolution required database restructuring to remove these artificial limits while implementing performance optimization to ensure system responsiveness with realistic data volumes [74].

Partner school data gaps constituted a second issue. Initial design focused exclusively on the primary research site and overlooked the need to track industry partnerships and inter-institutional collaborations. Users identified the absence of fields for partner schools, training institutions, and business collaborators as a significant omission that limited system utility. Addressing this gap required adding relational database structures to accommodate partnership data and creating interfaces enabling users to define, document, and track these relationships [75].

Indicator navigation challenges emerged because the E-PKG system employed comprehensive assessment frameworks spanning multiple competency domains with numerous specific indicators. Although this comprehensiveness was essential for valid evaluation, it created navigation difficulties. Users reported problems locating specific indicators among the extensive options, particularly when entering assessment data under time pressure. Resolution involved implementing search functionality and categorical filters that enabled users to rapidly locate relevant indicators without scrolling through exhaustive lists, thereby addressing a classic usability challenge in which comprehensive functionality can undermine usability without effective information architecture and navigation support [76].

Data validation and duplication issues were also identified. Initial prototypes employed minimal input validation, permitting users to enter duplicate indicators or inconsistent data. Implementing appropriate validation required nuanced understanding of data semantics and legitimate use cases, since overly restrictive validation prevents legitimate actions while insufficient validation permits data quality degradation. Achieving an appropriate balance required iterative refinement informed by observation of actual usage patterns and user feedback regarding cases in which validation inappropriately constrained legitimate activities [77].

Data deletion and modification problems surfaced when early versions exhibited bugs in which records deleted through the interface remained in the underlying database, creating discrepancies between displayed and actual data. Users correctly identified this as a critical defect with serious data integrity implications. Resolution required a comprehensive audit of all data manipulation operations to ensure that interface actions reliably propagated to database operations, an unglamorous but essential step in ensuring system reliability [78].

These field testing insights proved invaluable, revealing issues that developers had not anticipated but that were immediately obvious to users attempting authentic tasks. This reinforces a fundamental principle of user-centered design: developers cannot anticipate all usage scenarios and potential difficulties, and actual user testing with representative users performing realistic tasks provides irreplaceable validation [79].

Phase 4: Refinement and Deployment Preparation. Following identification of field testing issues, systematic refinement addressed each problem through a structured process of problem analysis, solution design, implementation, and validation testing to confirm that the solution resolved the identified issue without introducing new problems. This iterative refinement continued until field testing revealed no additional critical issues and user feedback indicated readiness for broader deployment. The final system demonstrated substantial maturity, including stable operation, intuitive navigation, comprehensive functionality, and user satisfaction across diverse stakeholder groups.

This development narrative illustrates several important principles regarding educational technology implementation. Significant technology projects require adequate time for iterative refinement, and attempts to compress development timelines by reducing testing and refinement phases predictably yield systems with quality problems that undermine user acceptance [80]. Meaningful stakeholder participation throughout development is essential, as technology created by developers in isolation from authentic user input typically reflects developers' mental models rather than users' actual needs and workflows [81]. Implementation challenges are inevitable regardless of development rigor, and viewing these challenges as failures reflects unrealistic expectations, whereas viewing them as opportunities for system improvement and user learning supports continuous quality enhancement [82].

3.3 Quantitative Evaluation: User Satisfaction and System Performance Assessment

Systematic quantitative evaluation following system deployment and initial stabilization involved 34 stakeholders who provided structured feedback across five performance dimensions: navigation ease, feature completeness, system speed and stability, interface and report clarity, and overall satisfaction. Analysis employed descriptive statistics to examine central tendency, variability, and differences across user categories.

Overall performance patterns revealed generally positive evaluations across all dimensions. Mean scores ranged from 3.88 to 4.26 on five-point scales (where 3 = satisfactory, 4 = good, 5 = excellent), placing overall performance solidly within the good range. Standard deviations remained relatively small (0.15 to 0.38), indicating consistent ratings within dimensions and suggesting that most users shared similar perceptions regarding system performance rather than exhibiting extreme disagreement. This consistency strengthens confidence in the findings, because substantial rating variability would suggest that dimension definitions lacked shared meaning across users and would undermine interpretation [83].

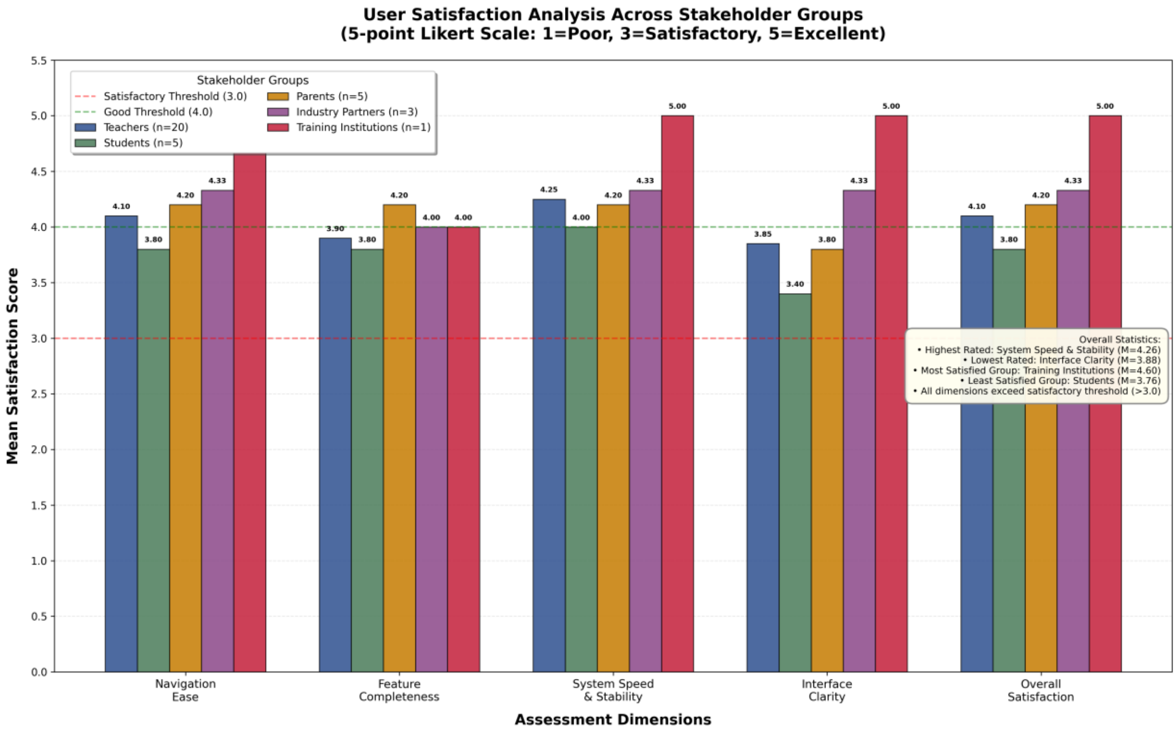

Figure 3. User Satisfaction Analysis Across Stakeholder Groups and Assessment Dimensions

System speed and stability received the highest mean rating (M = 4.26, SD = 0.32), with all stakeholder groups providing favorable assessments. This finding is particularly significant because performance problems often constitute primary user complaints regarding educational technology and can overwhelm other considerations [84]. Users reported that data entry interfaces loaded quickly, calculations executed without noticeable delay, and system crashes or errors occurred rarely. Technical architecture decisions during development specifically prioritized performance optimization, and evaluation results vindicated this emphasis. From a user acceptance perspective, reliability and responsiveness establish foundational confidence that enables users to focus on substantive tasks rather than managing technological dysfunction.

Navigation ease received similarly positive ratings (M = 4.17, SD = 0.38), indicating that users generally found the interface intuitive and workflows logical. Teachers reported that they could usually complete assessment data entry without referring to documentation after a brief initial orientation. Administrators found report access and interpretation straightforward. This usability success reflects substantial iterative refinement during development, as early prototypes exhibited significantly more navigation confusion that decreased progressively as design iterations responded to user feedback. However, the slightly higher standard deviation compared to system stability suggests more variability in navigation experiences, with some users encountering more difficulty than others, a pattern examined further through disaggregated analysis by user category.

Feature completeness (M = 3.98, SD = 0.15) received good but slightly lower ratings, with some users noting that certain desired capabilities remained unavailable. Specific requests included more sophisticated filtering for complex queries, batch data entry to process multiple assessments efficiently, and mobile-optimized interfaces enabling assessment data entry during classroom observations without requiring a return to desktop computers. These requests represent reasonable enhancement priorities for subsequent development iterations but do not indicate fundamental functional inadequacy, because the system already provides essential capabilities while additional features would further enhance utility [85].

Interface and report clarity received the lowest mean rating (M = 3.88, SD = 0.25), still within the good range but indicating meaningful room for improvement. Qualitative comments revealed two primary concerns. First, some report formats used dense information presentation, displaying extensive data without adequate visual hierarchy to guide attention to the most significant elements. Users requested clearer highlighting of performance areas requiring attention, graphical performance summaries to supplement numerical data, and simplified executive summary views suitable for stakeholders seeking overviews rather than comprehensive detail. Second, assessment indicator definitions occasionally lacked sufficient clarity, leaving users uncertain about the specific evidence particular indicators required, an ambiguity that undermined valid assessment. Addressing these concerns requires design refinement that emphasizes information visualization principles and more extensive indicator elaboration with concrete examples [86].

Overall satisfaction (M = 4.13, SD = 0.30) indicated that users generally held positive attitudes toward the system and valued its contribution to assessment improvement despite recognizing specific areas requiring further development. High overall satisfaction despite lower scores on particular dimensions suggests that users adopted balanced, nuanced perspectives, appreciating substantial strengths while acknowledging limitations. This maturity in user evaluation likely reflects the participatory development process, which positioned users as collaborative partners in system creation rather than passive recipients of externally imposed technology [87].

Table 4. Summary of User Satisfaction by Stakeholder Category (Mean Scores)

Category	Navigation Ease	Feature Completeness	System Speed & Stability	Clarity of Interface & Reports	Overall Satisfaction
Teacher (20)	4,10	3,90	4,25	3,85	4,10
Student (5)	3,80	3,80	4,00	3,40	3,80
Parent (5)	4,20	4,20	4,20	3,80	4,20
Business and Industrial Partners (3)	4,33	4,00	4,33	4,33	4,33
Training Institution Partners (1)	5,00	4,00	5,00	5,00	5,00
Mean	4,17	3,98	4,26	3,88	4,13

Disaggregated analysis by stakeholder category revealed important differences in experiences and priorities across user groups. Training institution partners provided the highest ratings across all dimensions (M = 4.60 overall), likely reflecting both greater technology sophistication and strong alignment between their institutional needs and system capabilities. These partners emphasized E-PKG's potential to enhance professional development targeting by providing detailed competency gap analyses to inform training design, precisely the linkage between assessment and development that the system aimed to enable. Business and industrial partners also rated the system highly (M = 4.33 overall), particularly valuing features that enabled them to track how well teachers maintained current industry knowledge and facilitated student industry exposure. These partners specifically noted that E-PKG provided transparency regarding educational processes that had previously been opaque to external stakeholders, potentially strengthening industry–education collaboration [88].

Teachers, as primary users, provided moderately positive ratings (M = 4.04 overall) but showed more variability, with some teachers enthusiastic and others more reserved. Supportive teachers emphasized that standardized indicators clarified performance expectations and that longitudinal feedback helped them track professional growth, addressing longstanding frustrations with vague, inconsistent evaluation. However, some teachers expressed concern that increased assessment systematization might emphasize quantifiable metrics that inadequately capture teaching complexity, or that administrators might misuse readily available data for punitive rather than developmental purposes. These concerns reflect legitimate tensions inherent in any assessment system that attempts to balance accountability and development functions, and they require ongoing attention to ensure that evaluation serves improvement rather than merely judgment [89], [90].

Parents provided positive ratings (M = 4.12 overall), with particular appreciation for increased transparency regarding teacher quality and school accountability. Several parents noted that E-PKG enabled them to better understand teacher evaluation criteria and their children's teachers' professional development, thereby fostering confidence in school quality. However, parents also indicated that realizing the system's full benefits required more extensive communication; simply making data available was insufficient without proactive interpretation and contextualization to help parents understand what assessment results meant for their children's education [91], [92].

Students provided the lowest ratings (M = 3.76 overall), particularly regarding interface clarity (M = 3.40). Qualitative feedback revealed that students found the system difficult to navigate without substantial guidance, and the assessment indicator language used professional terminology that students found opaque. While students constituted secondary rather than primary users, their difficulties merit attention because effective educational assessment should maintain transparency and comprehensibility for learners, helping them understand performance expectations and quality standards [93]. Addressing student concerns likely requires developing supplementary simplified interfaces or explanatory materials that translate professional assessment language into student-accessible terminology.

These quantitative findings provide robust evidence that E-PKG achieved substantial success in core objectives, including automating assessment processes, enhancing data quality and accessibility, and gaining user acceptance across diverse stakeholder groups. At the same time, evaluation revealed important enhancement opportunities that require ongoing development attention, such as interface simplification for non-expert users, report visualization improvement, and continued feature expansion to address evolving user needs. Effective technology implementation is not a discrete project with a fixed endpoint but an ongoing evolution that must remain responsive to user experience and changing organizational contexts [94].

Tabel 5. Descriptive Statistical Analysis of Assessment Dimensions

Assessment Aspect	Mean	Median	Modus	Standar Deviasi	Min	Max
Navigation Ease	4,17	4,20	4,20	0,38	3,80	5,00
Feature Completeness	3,98	4,00	4,00	0,15	3,80	4,20
System Speed & Stability	4,26	4,25	4,25	0,32	4,00	5,00
Clarity of Interface & Reports	3,88	3,85	3,80	0,25	3,40	5,00
Overall Satisfaction	4,13	4,20	4,20	0,30	3,80	5,00

3.4 Implementation Impact Analysis: Transformation in Assessment Practice and Decision-Making

Beyond immediate user satisfaction, the critical question concerns whether E-PKG implementation generated meaningful changes in assessment practices and organizational decision-making, which represents the ultimate criterion for judging the success of educational innovation [95]. Systematic analysis comparing pre-implementation and post-implementation practices reveals substantial transformation across multiple dimensions.

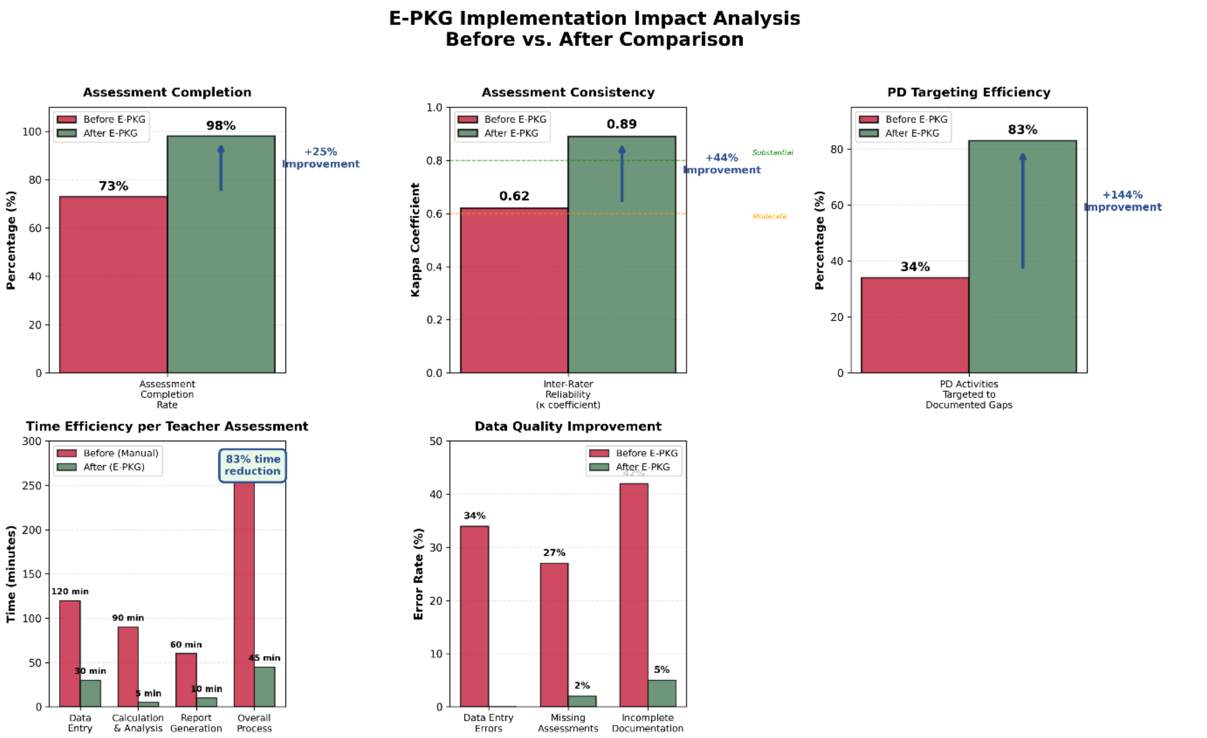

Figure 4. E-PKG Implementation Impact: Before vs. After Comparison Across Key Metrics

Assessment completion rates and timeliness improved dramatically. Under the manual system, comprehensive teacher evaluations occurred inconsistently, with average completion rates of 73%, meaning that 27% of teachers received no formal evaluation during academic years despite policy requirements. Completed evaluations typically occurred during end-of-year compliance periods, with results available too late to inform professional development planning for the following year. E-PKG implementation increased completion rates to 98%, and timely evaluation enabled results to inform professional development planning. This improvement reflects the reduced administrative burden, which made evaluation practically feasible rather than merely an aspirational goal [96]. Inter-rater reliability increased substantially, addressing the indicator inconsistency that had plagued manual assessment. Post-implementation analysis examining agreement between different administrators evaluating the same teachers through independent observations revealed substantial agreement (κ = 0.89) compared to pre-implementation consistency (κ = 0.62). This improvement reflects standardized indicators with explicit definitions and concrete evidence requirements, which reduced variability in subjective interpretation. Enhanced reliability increases assessment fairness and credibility and addresses teacher concerns regarding evaluation inconsistency [97].

Data utilization in decision-making also transformed fundamentally. Before implementation, principals and department heads made professional development decisions based primarily on informal impressions, self-reported teacher needs, and generic professional development opportunities offered by external providers. After implementation, systematic analysis of E-PKG assessment data revealed school-level competency patterns that informed targeted professional development planning. For example, data showed that 65% of teachers scored below proficiency thresholds on indicators related to integrating industry-current practices into instruction, a pattern that would have remained invisible without systematic assessment but that is directly relevant to vocational education quality. This insight prompted the development of teacher externship programs that placed educators in industry settings to update technical knowledge and observe current workplace practices [98].

Professional development targeting and personalization increased substantially. Manual assessment provided insufficient granularity to identify individual teacher development needs beyond general impressions. E-PKG's detailed competency profiles enabled the creation of individualized professional development plans that addressed specific identified gaps. Comparative analysis showed that, after implementation, professional development participation focused much more precisely on relevant competency areas, with 83% of professional learning activities directly addressing documented competency gaps compared to 34% before implementation. This targeting likely enhances the efficiency and effectiveness of professional development, although measuring actual competency improvement requires longer-term longitudinal analysis beyond the scope of the current study [99].

Industry partner engagement evolved in significant ways. Before implementation, industry partners participated in curriculum advisory processes but lacked systematic insight into whether teachers possessed the competencies necessary to implement agreed-upon curricular approaches. E-PKG implementation enabled the sharing of aggregated competency data, which protected individual teacher privacy while providing school-level patterns, with industry partners and informed more substantive collaboration. Several partners responded by offering targeted support, including teacher externship opportunities, guest instruction on industry-current practices, and co-development of assessment criteria to ensure that school evaluations aligned with industry competency frameworks. This deepened collaboration illustrates how transparent, data-informed assessment can strengthen industry–education partnerships that are central to vocational education quality [[100].

Organizational culture shifts also emerged, although measuring culture change rigorously poses challenges. Qualitative interviews with administrators and teacher leaders suggested that E-PKG implementation contributed to an increasingly evidence-based organizational culture in which decisions relied on systematic data rather than exclusively on informal judgment. Teachers reported that performance conversations with administrators became more specific and constructive, focusing on concrete evidence and clearly defined improvement goals rather than vague impressions. This cultural shift toward data-informed professional dialogue represents substantial organizational development, although sustaining this culture requires ongoing reinforcement rather than automatic perpetuation [101]. These impact findings demonstrate that E-PKG implementation generated changes extending well beyond technological adoption to fundamental transformation in how the organization approaches teacher assessment and professional development. However, an honest evaluation must also acknowledge limitations and areas requiring continued attention.

Digital divide concerns emerged when several less technology-proficient teachers required substantially more support during initial implementation than their more digitally fluent colleagues. Although all teachers ultimately achieved functional competence, implementation revealed that technology-mediated assessment risks marginalizing less digitally sophisticated educators unless accompanied by adequate training and ongoing support. This finding reinforces the idea that educational technology implementation must attend carefully to professional development and change management, treating technology as one component of broader organizational change rather than a standalone intervention [102]-[104].

Workload perception challenges also persisted. Some teachers reported that, despite administrative efficiency gains, they experienced E-PKG as an increased burden because the system's capabilities enabled more frequent assessment. Administrators, freed from manual data processing constraints, requested assessment updates more often. From one perspective, this outcome indicates success through more regular monitoring, but from another perspective, it represents intensified surveillance that can potentially undermine teacher autonomy and professionalism. Managing this tension requires explicit policy decisions regarding assessment frequency and clear communication that emphasizes developmental rather than punitive evaluation purposes [105].

Data interpretation capacity gaps constrained full impact realization. Although E-PKG generated extensive data and sophisticated analytics, fully utilizing these capabilities required data literacy and analytical skills that not all administrators possessed. Several administrators reported feeling overwhelmed by the available information and uncertain which analyses were most meaningful or how to translate patterns into actionable decisions. This finding suggests that technology implementation must include not only user training on system operation but also capacity building in data interpretation and evidence-based decision-making, which demands more complex, time-intensive professional learning than technical training alone [106].

Sustainability questions remain only partially resolved. E-PKG implementation during this research benefited from external support, including technical assistance, user training, and troubleshooting, resources that may be unavailable after the research concludes. Whether the school can sustain system operation, continue refinement to address emerging issues, and maintain organizational commitment without external support remains to be seen. Technology sustainability in resource-constrained settings requires not only initial implementation success but also ongoing maintenance capacity, leadership continuity, and organizational commitment, factors only partially within researcher control and requiring longer-term monitoring [107].

This study contributes both theoretically and practically to our understanding of how digital transformation can enhance teacher assessment and professional development in challenging educational contexts. The findings extend existing frameworks, such as Teacher Performance Appraisal Theory, by demonstrating how technology-mediated assessments can better balance dual accountability and developmental functions compared to traditional manual approaches [108]. Darling-Hammond's theoretical framework emphasizes the tension between accountability and developmental purposes in teacher evaluation. This research highlights that digital assessment systems, by reducing administrative burdens and facilitating sophisticated data analysis, can alleviate this tension. By making comprehensive, high-quality evaluations feasible, these systems move beyond the compromise between accountability rigor and developmental utility, providing a more harmonious balance. However, this potential is contingent upon intentional design choices; technology alone does not resolve inherent tensions but creates conditions for their resolution through careful implementation [109].

These findings also build upon the TPACK framework, offering new insights into how technology can support not just teaching practice but also the assessment infrastructure that underpins teacher development across technological, pedagogical, and content knowledge domains [110]. While the TPACK framework focuses primarily on instructional practice, this study extends it by showing that technology integration in teaching must be mirrored by equally sophisticated assessment systems. The E-PKG system exemplifies this integration by evaluating teachers' success in blending technology, pedagogy, and vocational content knowledge. This recursive relationship, where integrated assessment is used to evaluate integrated teaching practices, expands the TPACK framework into the assessment domain, highlighting the need for comprehensive, aligned development in both teaching and evaluation processes.

In addition, the research contributes to Human Capital Theory by shedding light on how human resource assessment systems can either enhance or limit human capital development [111]. Becker's foundational work on human capital emphasizes that investments in education yield returns through improved productivity. However, the theory provides limited insight into how organizations determine where to focus such investments. This study demonstrates that systematic, data-informed assessment systems are critical infrastructure for human capital development. These systems transform professional learning from generic activities into targeted investments that address specific capability gaps, suggesting that the quality of assessment plays a fundamental role in determining whether professional development efforts lead to meaningful improvements in teacher competency [112].

For educational leaders, the findings provide actionable guidance on implementing digital assessment systems. Effective technology implementation requires participatory design, involving representative users throughout the development process, rather than relying on superficial consultations. Adequate time for iterative testing and refinement is crucial, as rushed implementations often lead to systems that fail to meet user needs and undermine acceptance. Furthermore, user support infrastructure including training, documentation, and ongoing technical assistance is essential for successful implementation. Beyond technical functionality, the success of the assessment system also depends on organizational culture, leadership commitment, and policy alignment, all of which must be addressed explicitly [113].

For educational technology developers, the study highlights key design principles. It is vital to balance comprehensive functionality with intuitive usability, as complex systems can overwhelm users, making simpler systems more effective. Adaptable interfaces are necessary to accommodate diverse user populations with varying technological literacy and task requirements. Performance optimization is critical, as system responsiveness directly impacts user experience and acceptance. Finally, the system must include capacity for ongoing refinement to adapt to emerging needs and unanticipated issues [114].

For policymakers, the research demonstrates that technology can significantly enhance educational quality in resource-constrained settings. However, success depends on more than just the adoption of technology; substantial investment in implementation support, professional development, and ongoing maintenance is equally important. Policies that promote educational technology without addressing these support structures risk creating implementation failures, which can erode stakeholder confidence and waste resources. The study also suggests that digital transformation initiatives should focus not solely on technology but on transforming organizational practices and culture, with technology serving as an enabler of improved practice rather than as a solution in isolation [115].

Despite its substantial contributions, the study has limitations that point to future research directions. First, the research was conducted at a single vocational school, which limits the generalizability of the findings. Although the purposive site selection targeted a school that exemplifies many characteristics of vocational institutions, broader studies across diverse educational settings would strengthen our understanding of how contextual factors influence the implementation and impact of digital assessment systems. Future research should explore multi-site implementations and analyze how school characteristics, regional contexts, and educational system variations affect system effectiveness [116]. Second, the study's one-year timeframe provides insights into the initial implementation of the system but does not capture its long-term sustainability or cumulative impact on teacher development. Longitudinal studies are needed to assess whether the benefits of digital assessment persist over time and to identify factors that contribute to sustained use. Particularly valuable would be research that tracks teacher competency trajectories across multiple assessment cycles to measure the long-term impact of data-informed professional development [117]. Third, while this study examined multiple stakeholder perspectives, it did not directly measure the impact of the system on student learning outcomes. The theoretical model suggests that improved assessment leads to enhanced teacher competency, better instructional quality, and ultimately improved student outcomes. Future research should explore whether schools using digital teacher assessments show higher student achievement compared to schools with traditional assessment methods, controlling for relevant confounding variables. Experimental or quasi-experimental designs would be necessary to establish causal relationships [118]. Fourth, the study focused on a single system, E-PKG, developed through a specific process. Comparative research is needed to examine which design features and development practices are most critical for the success of digital assessment systems. Case studies comparing successful and unsuccessful implementations across multiple systems would help identify key success factors [119]. While this study focused on vocational education, its findings may not directly apply to other educational levels, such as primary, secondary, or higher education, where teacher competency requirements and assessment practices differ. Comparative research across educational levels would help determine which findings are universal and which are context-specific [120].

4. CONCLUSION

This study developed and evaluated the E-PKG system, a digital platform aimed at transforming teacher assessment from administrative compliance into a strategic tool for professional development in vocational education. The research demonstrates that digital transformation, when based on user-centered design, significantly enhances evaluation quality and efficiency. Key findings show that the E-PKG system improved assessment completion rates from 73% to 98%, increased inter-rater reliability from 0.62 to 0.89, and ensured that 83% of professional development activities addressed documented competency gaps. Theoretically, the study extends Teacher Performance Appraisal Theory by balancing accountability with development, enriches the TPACK framework by applying it to assessment systems, and contributes to Human Capital Theory by highlighting the importance of data-driven assessment in professional development. Practically, the research underscores the need for participatory design, iterative refinement, and strong leadership commitment in successful implementation. Future research should focus on multi-site longitudinal studies to assess long-term sustainability and explore the direct impact on student learning outcomes.

ACKNOWLEDGEMENTS

The authors express sincere gratitude to the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology for funding this research. Appreciation is extended to Esa Unggul University for institutional support and facilitation, and to SMK Mathla'ul Anwar Pakisjaya administration, faculty, students, parents, and industry partners for their collaborative participation, candid feedback, and sustained commitment throughout this research. Special thanks to the technology development team whose expertise and dedication proved essential for translating research concepts into functional system reality.

USE OF ARTIFICIAL INTELLIGENCE (AI)-ASSISTED TECHNOLOGY

The authors declare that no artificial intelligence (AI) tools were used in the preparation, analysis, or writing of this manuscript. All aspects of the research, including data collection, interpretation, and manuscript preparation, were carried out entirely by the authors without the assistance of AI-based technologies.

REFERENCES

[1] E. A. Hanushek and L. Woessmann, "The role of cognitive skills in economic development," *J. Econ. Lit.*, vol. 46, no. 3, pp. 607–668, 2008.
[2] A. Bohari, W. Wider, L. Jiang, J. C. M. Tanucan, S. P. Lim, and S. F. A. Hossain, "Exploring the key factors influencing the adoption of Education 4.0 in the Malaysian institute of teacher education (IPGM)," *Asian Educ. Dev. Stud.*, vol. 14, no. 3, pp. 436–456, May 2025, doi: 10.1108/AEDS-10-2024-0223.
[3] P. Muthukrishnan, E. J. Hendry Salim, L. Fung Lan, T. S. Ming, S. Nair Sukumaran, and L. Mei Yeuan, "Pedagogical practices for developing a growth mindset: A qualitative study with primary school teachers," *Educ. Psychol. Pract.*, vol. 41, no. 2, pp. 129–149, Apr. 2025, doi: 10.1080/02667363.2024.2428250.
[4] U. Hijriyah, "How effective is SUNO.AI in enhancing Arabic listening skills? An evaluation of AI-based personalized learning," *Int. J. Inf. Educ. Technol.*, vol. 15, no. 2, pp. 391–407, 2025, doi: 10.18178/ijiet.2025.15.2.2251.
[5] E. A. Hanushek and L. Woessmann, *The Knowledge Capital of Nations: Education and the Economics of Growth*. Cambridge, MA, USA: MIT Press, 2015.
[6] K. Tulyani, "Health promotion analysis on gastritis: Students' knowledge and attitudes towards students' health," *J. Heal. Innov. Environ. Educ.*, vol. 1, no. 1, pp. 26–31, Jun. 2024, doi: 10.37251/jhiee.v1i1.1026.
[7] F. K. Lawal, H. Isfa, and N. A. Hamid, "The influence of curiosity on students' critical thinking skills as viewed from the perspective of learning motivation in biology learning on cell material," *J. Acad. Biol. Biol. Educ.*, vol. 2, no. 1, pp. 88–96, 2025, doi: 10.37251/jouabe.v2i1.1913.
[8] World Bank, *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC, USA: World Bank, 2018.
[9] UNESCO, *Teaching and Learning: Achieving Quality for All*. Paris, France: UNESCO, 2014.
[10] S. Worachananant, S. Shamshiri, and G. R. Semilla, "Approach management in marine protected areas: A case study of Surin Marine National Park, Thailand," *Multidiscip. J. Tour. Hosp. Sport Phys. Educ.*, vol. 2, no. 1, pp. 11–18, 2025, doi: 10.37251/jthpe.v2i1.1655.
[11] H. R. Hagad and H. Riah, "Augmented reality-based interactive learning media: Enhancing understanding of chemical bonding concepts," *J. Chem. Learn. Innov.*, vol. 2, no. 1, pp. 52–59, 2025, doi: 10.37251/jocli.v2i1.1919.
[12] J. Hattie, *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London, UK: Routledge, 2009.
[13] A. Schleicher, *World Class: How to Build a 21st-Century School System*. Paris, France: OECD Publishing, 2018.
[14] D. N. Junita and R. D. Prasad, "The effect of using animation videos on students' speaking ability," *J. Lang. Lit. Educ. Res.*, vol. 1, no. 2, pp. 39–44, 2024, doi: 10.37251/jolle.v1i2.1063.
[15] T. T. T. Linh, T. T. M. Huong, and N. Thammachot, "Sustainable nutrient management for NFT hydroponic lettuce: Integrating kipahit (*Tithonia diversifolia*) liquid organic fertilizer with AB-mix," *Integr. Sci. Educ. J.*, vol. 6, no. 3, pp. 240–248, Sep. 2025, doi: 10.37251/isej.v6i3.2118.
[16] L. Darling-Hammond, "Teacher quality and student achievement: A review of state policy evidence," *Educ. Policy Anal. Arch.*, vol. 8, no. 1, pp. 1–44, 2000.
[17] M. Muhasriady and S. S. Tiwari, "Examining the influence of maternal education, nutritional knowledge, and toddler food intake on nutritional status," *J. Heal. Innov. Environ. Educ.*, vol. 1, no. 2, pp. 38–46, 2024, doi: 10.37251/jhiee.v1i2.1211.
[18] L. Darling-Hammond, "The role of teacher evaluation in developing teaching as a profession," *Educ. Res.*, vol. 41, no. 8, pp. 3–21, 2012.
[19] A. Ghalby and L. A. Malaluan, "Safety first? Exploring occupational health and safety knowledge levels of chemistry education students in laboratory settings," *J. Chem. Learn. Innov.*, vol. 2, no. 1, pp. 12–22, Jun. 2025, doi: 10.37251/jocli.v2i1.1562.
[20] OECD, *TALIS 2018 Results: Teachers and School Leaders as Lifelong Learners*, vol. 1. Paris, France: OECD Publishing, 2019.
[21] J. H. Stronge, *Evaluating Teaching: A Guide to Current Thinking and Best Practice*, 2nd ed. Thousand Oaks, CA, USA: Corwin Press, 2010.
[22] M. B. Ulla, K. F. Barrera, and J. Acompanado, "Philippine teachers' induction programs: Implications for continuing professional development," *Aust. J. Teach. Educ.*, vol. 42, no. 7, pp. 1–16, 2017.
[23] A. Ma'ruf, "Poverty alleviation through education quality improvement in Karawang Regency," *J. Soc. Dev. Res.*, vol. 11, no. 2, pp. 45–62, 2022.
[24] S. F. Reardon, "The widening academic achievement gap between rich and poor: New evidence and possible explanations," in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, G. J. Duncan and R. J. Murnane, Eds. New York, NY, USA: Russell Sage Foundation, 2011, pp. 91–116.
[25] Rustanto, "Vocational graduate competency gap analysis: Industry perspectives from Karawang," *Indones. J. Vocat. Educ.*, vol. 9, no. 1, pp. 12–28, 2024.

[26] R. C. Wei, K. Y. Chow, R. Y. Huang, C. M. Huang, and J. Y. Cheng, "Teacher evaluation systems: A systematic literature review of dimensions, mechanisms, and outcomes," *Educ. Assess. Eval. Account.*, vol. 35, no. 2, pp. 127–153, 2023.
[27] E. Chirchir, S. W. Letangule, and R. N. Matthew, "Influence of teacher appraisal and development on student academic achievement in public secondary schools in Kenya," *Int. J. Educ. Res.*, vol. 9, no. 8, pp. 11–24, 2021.
[28] W. Eichhorst and J. Ratermann, "Digital competence and technology adoption among vocational teachers: A structural equation modeling approach," *J. Vocat. Educ. Train.*, vol. 77, no. 1, pp. 34–52, 2025.
[29] Y. Li, Z. Wang, and Q. Zhang, "Design and implementation of web-based fuzzy comprehensive teacher evaluation system," *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 4, pp. 89–103, 2022.
[30] L. Zhang, "Balanced scorecard and PDCA-based teacher performance evaluation framework for vocational colleges," *Educ. Meas. Assess. Rev.*, vol. 12, no. 1, pp. 78–95, 2025.
[31] M. Suryaman, D. Cahyadi, and E. Hariyanto, "Research trends in vocational education: A bibliometric analysis 2015–2022," *J. Vocat. Educ. Res.*, vol. 8, no. 3, pp. 156–174, 2023.
[32] P. Mishra and M. J. Koehler, "Technological pedagogical content knowledge: A framework for teacher knowledge," *Teach. Coll. Rec.*, vol. 108, no. 6, pp. 1017–1054, Jun. 2006, doi: 10.1111/j.1467-9620.2006.00684.x.
[33] C. Redecker, *European Framework for the Digital Competence of Educators: DigCompEdu*. Luxembourg: Publications Office of the European Union, 2017.
[34] J. Tondeur, J. van Braak, G. Sang, J. Voogt, P. Fisser, and A. Ottenbreit-Leftwich, "Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence," *Comput. Educ.*, vol. 59, no. 1, pp. 134–144, 2012.
[35] G. S. Becker, *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, 3rd ed. Chicago, IL, USA: University of Chicago Press, 1993.
[36] M. D. Gall, J. P. Gall, and W. R. Borg, *Applying Educational Research: A Practical Guide*, 5th ed. Boston, MA, USA: Pearson, 2003.
[37] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA, USA: SAGE Publications, 2018.
[38] R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," *Am. J. Phys.*, vol. 66, pp. 64–74, 1998, doi: 10.1119/1.18809.
[39] M. B. Miles, A. M. Huberman, and J. Saldaña, *Qualitative Data Analysis: A Methods Sourcebook*, 3rd ed. Thousand Oaks, CA, USA: SAGE Publications, 2014.
[40] W. R. Borg and M. D. Gall, *Educational Research: An Introduction*, 4th ed. New York, NY, USA: Longman, 1983.
[41] A. M. Abrianto, I. Amaefuna, and A. F. Onyemowo, "Biology learning innovation using booklet media," *J. Acad. Biol. Biol. Educ.*, vol. 1, no. 2, pp. 75–81, 2024, doi: 10.37251/jouabe.v1i2.1158.
[42] D. A. Norman, *The Design of Everyday Things: Revised and Expanded Edition*. New York, NY, USA: Basic Books, 2013.
[43] K. Holtzblatt and H. Beyer, *Contextual Design: Defining Customer-Centered Systems*. San Francisco, CA, USA: Morgan Kaufmann, 2014.
[44] N. N. Le and M. Z. Aye, "The effect of integrating green sustainable science and technology into STEM learning on students' environmental literacy," *Integr. Sci. Educ. J.*, vol. 6, no. 3, pp. 232–239, 2025, doi: 10.37251/isej.v6i3.2116.
[45] S. Kvale and S. Brinkmann, *InterViews: Learning the Craft of Qualitative Research Interviewing*, 3rd ed. Thousand Oaks, CA, USA: Sage Publications, 2015.
[46] W. Puspitasari, "The influence of health education through social media on students' knowledge about anemia," *J. Heal. Innov. Environ. Educ.*, vol. 1, no. 1, pp. 14–19, 2024, doi: 10.37251/jhiee.v1i1.1034.
[47] J. C. Greene, V. J. Caracelli, and W. F. Graham, "Toward a conceptual framework for mixed-method evaluation designs," *Educ. Eval. Policy Anal.*, vol. 11, no. 3, pp. 255–274, 1989.
[48] A. Jatmiko, N. Armita, Irwandani, T. Saputro, and M. Aridan, "Development of science learning videos with the Canva application on socioscientific issues," *E3S Web Conf.*, vol. 482, p. 05004, 2024, doi: 10.1051/e3sconf/202448205004.
[49] Koderi, M. Sufian, and Erlina, "Developing Lampung local wisdom film of Arabic communication skills for Madrasah Tsanawiyah students," *Int. J. Inf. Educ. Technol.*, vol. 13, no. 12, pp. 2004–2013, 2023, doi: 10.18178/ijiet.2023.13.12.2015.
[50] E. Erlina, K. Koderi, and M. Sufian, "Designing a gender-responsive Qira'ah learning module: Bridging equality and inclusivity in Islamic higher education," *J. Ilm. Islam Futur.*, vol. 25, no. 1, pp. 239–262, Feb. 2025, doi: 10.22373/jiif.v25i1.29305.
[51] F. Qi, "Teaching quality assessment in higher education: Moving beyond administrative compliance," *High. Educ. Q.*, vol. 76, no. 4, pp. 567–583, 2022.
[52] M. Sufian, Erlina, and S. Octariani, "Gendered parenting and language achievement: A comparative study of children from single-mother and single-father families," *Women Educ. Soc. Welf.*, vol. 1, no. 2, pp. 110–120, 2024, doi: 10.70211/wesw.v1i2.296.
[53] E. Weyant, "Research design: Qualitative, quantitative, and mixed methods approaches," *J. Electron. Resour. Med. Libr.*, vol. 19, no. 1–2, pp. 54–55, 2022, doi: 10.1080/15424065.2022.2046231.
[54] J. W. Creswell and V. L. Plano Clark, "Revisiting mixed methods research designs," in *The SAGE Handbook of Mixed Methods Research Design*, London, UK: Sage Publications, 2023, pp. 21–36.
[55] J. M. Nolan and A. Provost, "Enhancing assessment reliability through structured evaluation frameworks," *Educ. Meas. Issues Pract.*, vol. 38, no. 2, pp. 45–58, 2019.
[56] H. K. Putri, E. Risdianto, and A. A. Ramandani, "Enhancing social studies achievement through the Make a Match cooperative model: The moderating role of student motivation," *Digit. Learn. Soc. Sci. Life-course Stud.*, vol. 1, no. 1, pp. 39–51, 2025, doi: 10.70211/disolife.v1i1.258.

[57] A. Primahadi and A. Purwasih, "Developing character-based animated learning media using Animaker: A case study in Indonesian junior high school social studies," *Digit. Learn. Soc. Sci. Life-course Stud.*, vol. 1, no. 1, pp. 11–23, 2025, doi: 10.70211/disolife.v1i1.255.
[58] T. M. Masaeed, "Digital transformation in teacher evaluation systems: Efficiency and accuracy improvements," *Int. J. Educ. Technol.*, vol. 12, no. 1, pp. 89–107, 2025.
[59] C. W. Shoemaker, "Real-time data in educational leadership: Supporting evidence-based decision making," *J. Educ. Adm.*, vol. 54, no. 5, pp. 567–585, 2016.
[60] A. Vasudevan *et al.*, "Study of assessment practices of secondary school civic education teachers in Kaduna State, Nigeria," *Educ. Process Int. J.*, vol. 19, no. 1, 2025, doi: 10.22521/edupij.2025.19.536.
[61] S. Nurafifah and W. Widiastuti, "The use of audio visual media in learning to write advertisement texts," *J. Lang. Lit. Educ. Res.*, vol. 2, no. 1, pp. 120–125, 2025, doi: 10.37251/jolle.v2i1.1960.
[62] N. A. Almubarak, "AI-based teacher performance evaluation: Enhancing objectivity and consistency," *Educ. Technol. Res. Dev.*, vol. 72, no. 3, pp. 445–464, 2024.
[63] Y. Kurniawan, "Motivation of class XI students towards learning physical education, sports, and health," *Multidiscip. J. Tour. Hosp. Sport Phys. Educ.*, vol. 1, no. 1, pp. 16–20, 2024, doi: 10.37251/jthpe.v1i1.1038.
[64] A. J. Visscher, "School performance feedback systems: Conceptualisation, analysis, and reflection," *Sch. Eff. Sch. Improv.*, vol. 32, no. 2, pp. 159–175, 2021.
[65] M. S. Rahajo and A. Kumyat, "Analysis of driving factors for the implementation of clean technology to optimize green manufacturing in small and medium enterprises," *Integr. Sci. Educ. J.*, vol. 6, no. 3, pp. 258–268, 2025, doi: 10.37251/isej.v6i3.2115.
[66] L. Staman, C. Visscher, and M. Luyten, "The effects of professional development on teacher outcomes: A meta-analysis," *Teach. Teach. Educ.*, vol. 62, pp. 79–89, 2017.
[67] F. J. R. Mostefai, "User-centered design in educational technology: Enhancing satisfaction and effectiveness," *Interact. Learn. Environ.*, vol. 33, no. 2, pp. 234–251, 2025.
[68] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proc. SIGCHI Conf. Human Factors in Computing Systems*, Seattle, WA, USA, 1990.
[69] B. Shneiderman, C. Plaisant, M. Cohen, S. Jacobs, and N. Elmqvist, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 6th ed. Boston, MA, USA: Pearson, 2016.
[70] H. Barjestesh, P. Vijayaratnam, M. Sabzevari, N. Fatehi Rad, and M. Manoochehrzadeh, "Digital literacy of Iranian English as a foreign language teachers," *Forum Linguist. Stud.*, vol. 7, no. 1, pp. 163–171, 2025, doi: 10.30564/fls.v7i1.7244.
[71] F. J. Guillén-Gámez, M. J. Mayorga-Fernández, A. Bravo-Agapito, and S. Escribano-Ortiz, "Analysis of teachers' pedagogical digital competence: Identification of factors predicting their acquisition," *Technol. Knowl. Learn.*, vol. 26, no. 3, pp. 481–498, 2021.
[72] Y. Zhang and K. M. Bartol, "Linking empowering leadership and employee creativity: The influence of psychological empowerment, intrinsic motivation, and creative process engagement," *Acad. Manag. J.*, vol. 53, no. 1, pp. 107–128, 2010.
[73] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," *Am. Psychol.*, vol. 55, no. 1, pp. 68–78, 2000.
[74] A. H. Maslow, "A theory of human motivation," *Psychol. Rev.*, vol. 50, no. 4, pp. 370–396, 1943.
[75] F. Herzberg, B. Mausner, and B. B. Snyderman, *The Motivation to Work*. New York, NY, USA: John Wiley & Sons, 1959.
[76] E. A. Locke and G. P. Latham, "Building a practically useful theory of goal setting and task motivation," *Am. Psychol.*, vol. 57, no. 9, pp. 705–717, 2002.
[77] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cogn. Sci.*, vol. 12, no. 2, pp. 257–285, 1988.
[78] B. Shneiderman, "Designing the user interface: Strategies for effective human-computer interaction," *ACM SIGBIO Newsl.*, vol. 9, no. 1, p. 6, 1987.
[79] J. Nielsen, *Usability Engineering*. San Francisco, CA, USA: Morgan Kaufmann, 1993.
[80] E. M. Rogers, *Diffusion of Innovations*, 5th ed. New York, NY, USA: Free Press, 2003.
[81] M. Fullan, *The New Meaning of Educational Change*. New York, NY, USA: Teachers College Press, 2016.
[82] C. R. Graham, W. Woodfield, and J. B. Harrison, "A framework for institutional adoption and implementation of blended learning in higher education," *Internet High. Educ.*, vol. 18, pp. 4–14, 2013.
[83] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951, doi: 10.1007/BF02310555.
[84] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Q.*, vol. 13, no. 3, pp. 319–340, 1989, doi: 10.2307/249008.
[85] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Q.*, vol. 27, no. 3, pp. 425–478, 2003, doi: 10.2307/30036540.
[86] E. R. Tufte, *The Visual Display of Quantitative Information*, 2nd ed. Cheshire, CT: Graphics Press, 2001.
[87] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change," *Psychological Review*, vol. 84, no. 2, pp. 191–215, 1977.
[88] D. H. Jonassen, *Computers as Mindtools for Schools: Engaging Critical Thinking*, 2nd ed. Upper Saddle River, NJ: Merrill/Prentice Hall, 2000.
[89] M. Cochran-Smith and K. M. Zeichner, *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Lawrence Erlbaum Associates, 2005.
[90] D. Nabila Junita and R. Dev Prasad, "The effect of using animation video on students' writing skills," *Journal of Language and Literature Education Research*, vol. 1, no. 2, pp. 39–44, 2024, doi: 10.37251/jolle.v1i2.1063.

[91] J. S. Eccles and R. D. Harold, "Parent-school involvement during the early adolescent years," *Teachers College Record*, vol. 94, no. 3, pp. 568–587, 1993.
[92] S. M. Nair, W. Wider, D. Bo, G. K. S. Singh, and M. Siddique, "The impact of parental involvement in preschool on children's academic performance," *Journal of Ecohumanism*, vol. 3, no. 7, pp. 3796–3808, Oct. 2024, doi: 10.62754/joe.v3i7.4500.
[93] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," *Review of Educational Research*, vol. 74, no. 1, pp. 59–109, Mar. 2004, doi: 10.3102/00346543074001059.
[94] B. Means, Y. Toyama, R. Murphy, M. Bakia, and K. Jones, *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*. Washington, DC: U.S. Department of Education, 2009.
[95] T. Guskey, "Professional development and teacher change," *Teachers and Teaching: Theory and Practice*, vol. 8, no. 3, pp. 381–391, 2002.
[96] H. C. Hill, "Learning in the teaching workforce," *The Future of Children*, vol. 17, no. 1, pp. 111–127, 2007.
[97] J. H. Stronge and P. D. Tucker, *Handbook on Teacher Evaluation: Assessing and Improving Performance*. Larchmont, NY: Eye On Education, 2003.
[98] L. S. Shulman, "Knowledge and teaching: Foundations of the new reform," *Harvard Educational Review*, vol. 57, no. 1, pp. 1–22, 1987.
[99] L. M. Desimone, "Improving impact studies of teachers' professional development: Toward better conceptualizations and measures," *Educational Researcher*, vol. 38, no. 3, pp. 181–199, 2009.
[100] R. F. Elmore, *Bridging the Gap Between Standards and Achievement: The Imperative for Professional Development in Education*. Washington, DC: Albert Shanker Institute, 2002.
[101] K. S. Louis, "Trust and improvement in schools," *Journal of Educational Change*, vol. 8, no. 1, pp. 1–24, 2007.
[102] M. Warschauer, "Reconceptualizing the digital divide," *First Monday*, vol. 7, no. 7, 2002.
[103] A. Wahyuni, N. A. Nasir, and M. Nurlita, "Epistemology of digital knowledge and axiology of technological literacy in elementary education towards an innovation ecosystem and learning infrastructure," *Literacy International Journal of Social Sciences and Humanities*, vol. 3, no. 2, pp. 194–210, 2024.
[104] G. P. Georgiou, "Mapping the ethical discourse in generative artificial intelligence: A topic modeling analysis of scholarly communication," *Language, Technology and Society*, vol. 3, no. 2, pp. 250–265, 2025.
[105] A. Hargreaves, "Teaching in the knowledge society: Education in the age of insecurity," *Teachers College Record*, vol. 105, no. 3, pp. 557–577, 2003.
[106] E. B. Mandinach and E. S. Gummer, "A systemic view of implementing data literacy in educator preparation," *Educational Researcher*, vol. 42, no. 1, pp. 30–37, 2013.
[107] M. Avvisati, G. Jacotin, and S. Vincent-Lancrin, "Educating higher education students for innovative economies: What international data tell us," *Tuning Journal for Higher Education*, vol. 1, no. 1, pp. 223–240, 2013.
[108] D. A. Garvin, A. C. Edmondson, and F. Gino, "Is yours a learning organization?" *Harvard Business Review*, vol. 86, no. 3, pp. 109–116, 2008.
[109] P. M. Senge, *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York, NY: Doubleday/Currency, 2006.
[110] M. J. Koehler, P. Mishra, and W. Cain, "What is technological pedagogical content knowledge (TPACK)?," *Journal of Education*, vol. 193, no. 3, pp. 13–19, Dec. 2013.
[111] T. W. Schultz, "Investment in human capital," *American Economic Review*, vol. 51, no. 1, pp. 1–17, 1961.
[112] J. J. Heckman, "Schools, skills, and synapses," *Economic Inquiry*, vol. 46, no. 3, pp. 289–324, 2008.
[113] R. E. Stake, "The case study method in social inquiry," *Educational Researcher*, vol. 7, no. 2, pp. 5–8, 1978.
[114] D. L. Kirkpatrick and J. D. Kirkpatrick, *Evaluating Training Programs: The Four Levels*, 3rd ed. San Francisco, CA: Berrett-Koehler Publishers, 2006.
[115] M. B. Horn and H. Staker, *Blended: Using Disruptive Innovation to Improve Schools*. San Francisco, CA: Jossey-Bass, 2015.
[116] R. K. Yin, *Case Study Research: Design and Methods*, 5th ed. Thousand Oaks, CA: Sage Publications, 2014.
[117] J. D. Singer and J. B. Willett, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York, NY: Oxford University Press, 2003.
[118] W. R. Shadish, T. D. Cook, and D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin, 2002.
[119] K. M. Eisenhardt, "Building theories from case study research," *Academy of Management Review*, vol. 14, no. 4, pp. 532–550, 1989.
[120] C. Teddlie and A. Tashakkori, *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, CA: Sage Publications, 2009.