Volume: 28, Number: 2, September 2026, pp. 230-246
p-ISSN: 1410-8771, e-ISSN: 2580-4812
https://doi.org/10.37303/likhitaprajna.v28i2.1001

LIKHITAPRAJNA Jurnal Ilmiah
Fakultas Keguruan dan Ilmu Pendidikan

MUSSA (Mobile User Assistant for Smart School Assessment): Development of An AI-Based Learning Assessment Platform

Saddam Fathurrachman[1], Novi Eka Saputri[2]
[1]Universitas Terbuka, Tangerang Selatan, Indonesia, saddam.fathurrachman@ecampus.ut.ac.id
[2]Universitas Terbuka, Tangerang Selatan, Indonesia, novi.eka.saputri@ecampus.ut.ac.id

ARTICLE INFO

Article history:
Received 2 August 2026
Revised 7 September 2026
Accepted 15 September 2026

Keywords:
Artificial Intelligence
Learning Assessment
Indonesian Language
Elementary School
ADDIE

To cite this article:
Fathurrachman, S., & Saputri, N. (2026). MUSSA (Mobile User Assistant for Smart School Assessment): Development of An AI-Based Learning Assessment Platform. *Jurnal Likhitaprajna*, *28*(2), 230-246. https://doi.org/10.37303/likhitaprajna.v28i2.1001

ABSTRACT

Indonesian language learning assessment in elementary schools is generally still dominated by manual methods oriented towards final results, making the provision of formative feedback suboptimal. This study aims to develop and test the feasibility of the MUSSA (Mobile User Assistant for Smart School Assessment) web-based platform as an Artificial Intelligence (AI)-based digital assistant for assessing student essays. This study employs the Research and Development (R&D) method using the ADDIE model (Analysis, Design, Development, Implementation, Evaluation). The feasibility of the platform was evaluated by media and material experts, followed by a longitudinal field trial (six meetings) involving 74 5th-grade students and 3 teachers across three elementary schools in West Java. Expert validation results showed an average feasibility percentage of 86% (Highly Valid). In the implementation stage, the platform was deemed highly practical by teachers (87.2%) for reducing grading. Nevertheless, findings indicate that the AI feedback language (Natural Language Processing) needs to be refined to be more child-friendly.

Corresponding Author:
Saddam Fathurrachman
Universitas Terbuka, Malang, Indonesia; fathurrachmansaddam@gmail.com

INTRODUCTION

Learning assessment is an integral component of the educational system that serves not only to measure students' competency attainment but also to provide informative feedback for both students and teachers, thereby supporting subsequent instructional planning. Conceptually, assessment is defined as a systematic process of collecting, analyzing, and interpreting information regarding students' learning achievement and progress to facilitate informed pedagogical decision-making. Within the context of primary education, assessment plays a crucial role in monitoring students' cognitive, linguistic, and affective development during the early stages of academic competence formation. In Indonesian language learning at the primary school level, assessment possesses distinctive and complex characteristics because language is regarded as a performative competence encompassing listening, speaking, reading, and writing skills, all of which develop simultaneously and interactively.

In response to the rapid advancement of education in the digital era, assessment paradigms have shifted toward formative, continuous, and authentic approaches that enable real-time monitoring of learning progress and facilitate timely instructional interventions. Formative assessment is widely recognized as an effective strategy for enhancing learning because it allows teachers to provide continuous feedback and adapt instruction according to students' specific learning needs (Sholeh et al., 2024). Nevertheless, the consistent implementation of formative assessment remains challenging due to teachers' limited instructional time and the lack of assessment tools capable of delivering rapid and meaningful feedback. Consequently, assessment practices often remain predominantly focused on learning outcomes (assessment of learning) rather than on supporting students' continuous learning development.

The emergence of Artificial Intelligence (AI) in education has driven substantial transformations in learning assessment systems. AI offers considerable potential to enhance the efficiency and effectiveness of assessment through automated evaluation, real-time analysis of student learning data, and the provision of personalized feedback (Sabri & Wais, 2025). Theoretically, the integration of AI into assessment is grounded in a data-driven decision-making approach, in which learning data are continuously utilized to inform instructional decisions and improve learning quality. AI also enables assessment processes to become more adaptive by adjusting task difficulty and generating recommendations based on individual student learning profiles (Holmes et al., 2019).

Previous studies have demonstrated the potential of AI-assisted assessment, particularly in automated writing evaluation (AWE) and formative feedback. For example, Fleckenstein et al. (2023), through a meta-analysis of 20 studies involving 2,828 learners, found that automated feedback had a moderate positive effect on students' writing performance. However, the considerable heterogeneity across studies indicates that the effectiveness of automated assessment varies depending on the design of the system, learner characteristics, and instructional context. Similarly, Steiss et al. (2024) reported that although generative AI can provide writing feedback efficiently, human-generated feedback remains stronger in several dimensions of formative feedback quality. More recent evidence also suggests that AI-generated feedback tends to perform well in providing immediate responses related to grammar, organization, and textual coherence, but remains less consistent in addressing higher-order abilities such as reasoning, creativity, metacognition, and context-sensitive language use (Jannath & Akter, 2026). These findings suggest that the pedagogical value of AI-based assessment depends not merely on the automation of scoring or feedback, but on how closely the system is aligned with learning objectives, learner characteristics, and the specific educational context.

A further limitation concerns the population and disciplinary contexts examined in previous research. Much of the existing evidence on AI-assisted assessment has been derived from higher education and English as a Foreign Language (EFL) settings, while research focusing specifically on Indonesian language learning at the primary school level remains relatively limited. Although recent studies have begun to explore GenAI-assisted Indonesian language assessment for elementary school students, such as the development of creativity-oriented assessment instruments by Rakhmah and Anggraini (2026), these initiatives have generally focused on particular assessment competencies or on AI as an instrument-development aid rather than on an integrated assessment platform that supports continuous diagnosis, automated analysis, and personalized formative feedback. This limitation reinforces the research gap identified by Purba et al. (2025), particularly regarding the contextualized application of AI for Indonesian language assessment in primary education.

Against this background, MUSSA is proposed to extend the existing body of AI-based assessment research by developing an integrated assessment platform specifically designed for Indonesian language learning at the primary school level. Unlike general-purpose generative

AI tools or automated writing evaluation systems that primarily emphasize automated scoring and linguistic correction, MUSSA is designed to integrate assessment, analysis of students' learning performance, and personalized feedback within a single learning environment. Its contribution lies in contextualizing AI-supported assessment according to the characteristics of elementary school learners and the objectives of Indonesian language learning, rather than merely transferring generic AI assessment mechanisms into the classroom. Accordingly, the novelty of MUSSA lies not simply in the use of AI, but in the pedagogical integration of AI into a context-specific, learner-oriented, and formative assessment system intended to support teachers in identifying students' learning needs and providing more responsive follow-up instruction.

In response to these challenges, this study addresses the need for assessment innovation through the development of MUSSA (Mobile User Assistant for Smart School Assessment). MUSSA is designed as a web-based, AI-powered learning assessment platform that assists teachers in managing digital assignments, analyzing students' language errors, and providing contextualized automated feedback. This study aims to: (1) describe the development process of the MUSSA platform using the ADDIE instructional design model; (2) evaluate the platform's validity through expert assessments in educational pedagogy and educational technology; and (3) examine the platform's practicality in assisting teachers in monitoring students' language skill development through informative data visualization. It is expected that MUSSA will serve as an effective digital assistant to support authentic assessment practices in accordance with the objectives of the Primary School Curriculum.

METHOD

Research Approach and Design

This study employed a Research and Development (R&D) approach aimed at designing, developing, and evaluating the feasibility of an educational technology product (Borg & Gall, 2003; Sugiyono, 2017). The development process adopted the ADDIE instructional design model, consisting of five sequential phases: Analysis, Design, Development, Implementation, and Evaluation. The ADDIE model was selected because it provides a systematic, iterative, and practical framework for developing and validating educational software and technology-enhanced learning platforms (Branch, 2009; Molenda, 2015).

Research Participants and Setting

The field implementation was conducted in three public primary schools located in West Java Province, Indonesia. The participants comprised 74 fifth-grade students from three different classrooms and three classroom teachers responsible for Indonesian language instruction. In addition to the field participants, the study involved a panel of expert validators, including an Indonesian language education expert, an educational assessment expert, and an educational media/information technology expert. These experts evaluated both the research instruments and the MUSSA platform prior to its classroom implementation to ensure its content validity, technical quality, and pedagogical appropriateness (Richey & Klein, 2007).

ADDIE Development Procedure

The development of the MUSSA (Mobile User Assistant for Smart School Assessment) platform followed the five phases of the ADDIE model as described below.

1. Analysis Phase
   A preliminary needs assessment was conducted to identify the challenges encountered by teachers in implementing formative assessment for Indonesian language writing skills. The analysis focused on existing assessment practices, teachers' instructional needs, curriculum requirements, and the cognitive characteristics of primary school students. Based on these findings, the functional specifications of the AI-powered assessment system were

formulated in alignment with the Indonesian primary school curriculum and the expected cognitive competencies of upper-grade elementary students.

2. Design Phase
   During this phase, the overall architecture of the MUSSA platform was designed, including the preparation of storyboards, User Interface (UI) and User Experience (UX) designs, database structures, and AI-assisted assessment workflows. Assessment rubrics were also developed based on Indonesian language literacy indicators, enabling the AI system to evaluate students' writing in accordance with established pedagogical criteria.
3. Development Phase
   The development phase involved programming the web-based platform and integrating Artificial Intelligence functionalities into the system. Upon completion of the prototype, expert validation was conducted to evaluate its pedagogical quality, technical functionality, and usability. Revisions were subsequently made based on qualitative feedback from the expert validators to ensure system stability, eliminate software bugs, and improve the accuracy of the AI-generated language assessment and feedback.
4. Implementation Phase
   After being declared valid by the expert panel, the MUSSA platform was implemented in classroom settings through a longitudinal field trial. The implementation was carried out over six consecutive Indonesian language learning sessions, allowing teachers to continuously monitor students' writing development.
   The implementation procedure consisted of the following steps:
   a. Teachers photographed students' handwritten writing assignments.
   b. The captured images were uploaded to the MUSSA platform.
   c. The AI system extracted the handwritten text using optical character recognition (OCR) technology and conducted automated linguistic analyses, including spelling accuracy, sentence structure, and textual coherence.
   d. The platform generated immediate automated feedback, which was presented to teachers through an interactive dashboard.
   e. Teachers utilized the longitudinal assessment data to monitor students' literacy development across the six instructional sessions and employed the AI-generated feedback as evidence-based support when providing individualized guidance and recommendations for improvement.
5. Evaluation Phase
   The final phase involved evaluating the practicality and effectiveness of the MUSSA platform based on data collected during the field implementation. The evaluation examined the platform's usability, functionality, and its capacity to support teachers in conducting authentic, continuous, and technology-assisted assessments of students' Indonesian language literacy skills.
   To provide a clearer overview of the development process, the workflow of the MUSSA platform is illustrated in the following figure.

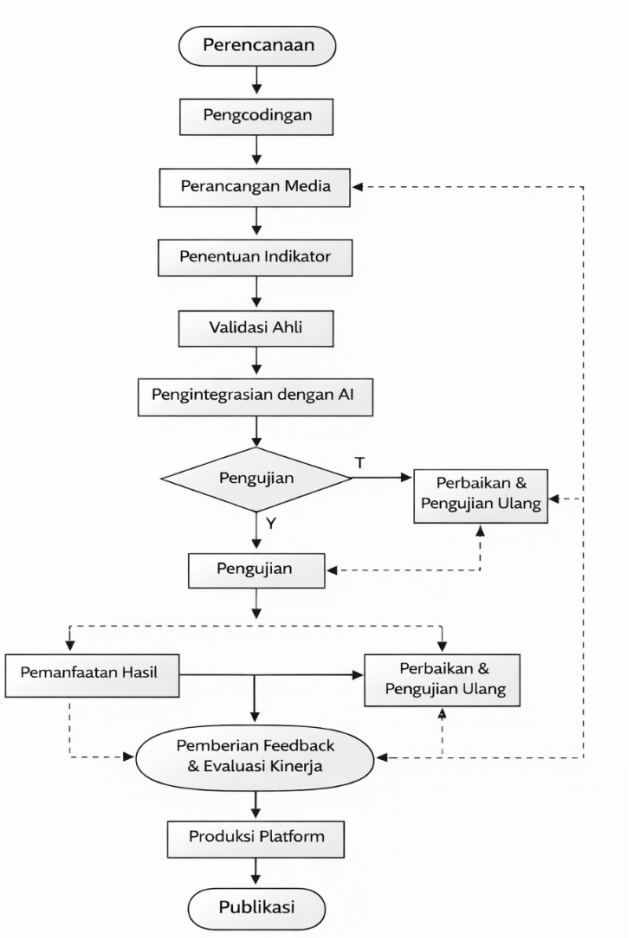

Figure 1. Flowchart of the MUSSA Platform Development

Data Collection and Data Analysis

Data were collected using three primary techniques: expert validation questionnaires, classroom implementation observations conducted throughout the six instructional sessions, and practicality questionnaires completed by both teachers and students. The questionnaire instruments employed a five-point Likert scale to measure perceptions of the platform's quality, usability, and practicality (Arikunto, 2020). Data obtained from expert validation and practicality assessments were analyzed using descriptive quantitative statistics. The score obtained for each component was divided by the maximum possible score and multiplied by 100 to produce a percentage score. The interpretation criteria were adapted from Riduwan (2012) and applied consistently to both validity and practicality assessments. Percentage scores of 81–100% were categorized as Highly Valid/Highly Practical, 61–80% as Valid/Practical, 41–60% as Moderately Valid/Moderately Practical, 21–40% as Less Valid/Less Practical, and 0–20% as Invalid/Impractical. Thus, the labels Highly Valid and Highly Practical were not assigned arbitrarily but were determined according to predetermined percentage intervals derived from the adopted assessment criteria.

To complement the descriptive feasibility and practicality analysis, the evaluation of MUSSA should also consider evidence of its effectiveness during classroom implementation. Where pretest and posttest learning-outcome data are available, effectiveness can be examined by comparing students' scores before and after the use of MUSSA. The magnitude of improvement may be calculated using normalized gain (N-gain), which represents the proportion of the actual gain relative to the maximum possible gain (Hake, 1998). In addition, a paired-samples t-test may be employed when the difference scores meet the assumption of normality; otherwise, the Wilcoxon signed-rank test may be used as a non-parametric

alternative. Effect size should also be reported to indicate the magnitude of the observed improvement rather than relying solely on statistical significance (Cohen, 1988).

When learning-outcome data are not available, evidence of implementation effectiveness can be strengthened using data collected across the six instructional sessions. In this case, classroom observation data can be analyzed not only as overall percentages but also as session-by-session trends, including students' task completion, participation, response to feedback, ability to complete assessment activities, and teachers' ability to use assessment information for instructional follow-up. Such longitudinal descriptive analysis provides additional evidence of whether the platform contributes to increasingly effective and manageable assessment practices during actual classroom use.

Accordingly, the evaluation of MUSSA encompasses three complementary dimensions: validity, which reflects the appropriateness of the platform's content and design as judged by experts; practicality, which represents the perceived usability of the platform from the perspectives of teachers and students; and effectiveness, which reflects observable changes in students' learning performance or in the quality of the assessment process during classroom implementation. This multidimensional evaluation provides a stronger basis for determining the overall quality and potential applicability of the MUSSA platform than relying on percentage-based feasibility scores alone.

RESULTS AND DISCUSSION

RESULT

The development of the MUSSA (Mobile User Assistant for Smart School Assessment) platform was carried out through a systematic process in which the features of the final product were directly derived from the needs identified during the Analysis phase. The preliminary needs assessment revealed several key challenges in Indonesian language writing assessment, including the time-consuming nature of manual grading, the limited provision of timely formative feedback, the need for assessment criteria aligned with curriculum standards, and the difficulty of continuously monitoring individual students' writing development. These identified needs were subsequently translated into specific functional requirements for the MUSSA platform.

Accordingly, each major feature of MUSSA was designed to respond to a specific instructional and assessment need. The essay upload and Optical Character Recognition (OCR) functions were developed to reduce the burden of manually transferring and reviewing handwritten student work, while the AI-assisted assessment feature was designed to support faster analysis of spelling, sentence structure, and textual coherence. The configurable assessment rubric was incorporated to ensure that automated scoring remained aligned with Indonesian language writing competencies and teachers' instructional objectives. In addition, the student management system and centralized database were developed to organize assessment records systematically, whereas the Learning Progress dashboard was designed to enable longitudinal monitoring of individual student performance across successive instructional sessions. These features allow teachers not only to obtain assessment results more efficiently but also to use accumulated assessment data for timely instructional follow-up and individualized guidance.

Therefore, the findings of this study are presented by emphasizing the relationship between the needs identified during the Analysis phase and the resulting product features. The results are organized into three main sections: (1) the description of the final MUSSA product and the extent to which its interface and core features address the identified assessment needs; (2) the results of expert validation examining the technical and pedagogical appropriateness of those features; and (3) the findings from field implementation and practicality evaluation, which assess how the developed features function in authentic classroom settings.

Platform Interface and Features of MUSSA

MUSSA was developed as a responsive web-based platform that can be accessed through various digital devices. The User Interface (UI) was designed with a strong emphasis on simplicity, accessibility, and ease of navigation, enabling primary school teachers to operate the system with minimal technical expertise. The platform integrates AI-assisted assessment tools into an intuitive workflow that supports teachers throughout the assessment process.

1. Landing Page

The landing page introduces the core value proposition of MUSSA as an AI-powered platform for facilitating the assessment of students' essay writing. Upon accessing the platform, teachers are presented with an intuitive navigation interface that enables them to upload handwritten student essays, securely store assessment data in a centralized database, and monitor students' learning progress through an integrated dashboard.

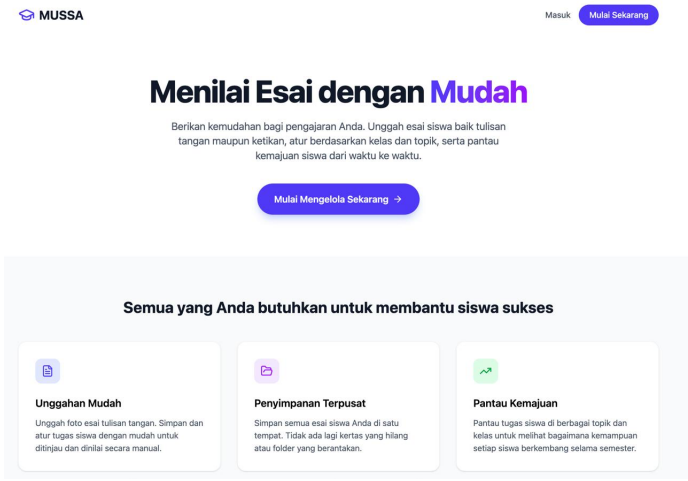

Figure 2. Interactive Landing Page of The MUSSA Platform

2. Main Dashboard and Student Management

After logging in, teachers are directed to the Main Dashboard, which provides a comprehensive overview of key assessment metrics, including the total number of students, the number of classes, the number of essays that have been assessed, and the average student score. To improve workflow efficiency, the dashboard also incorporates a Quick Actions feature that enables teachers to perform essential tasks, such as managing student records and uploading new essay submissions, with minimal navigation. Within the Student List module, the platform systematically organizes student information according to their respective classrooms, allowing teachers to efficiently access, manage, and monitor student data throughout the assessment process.

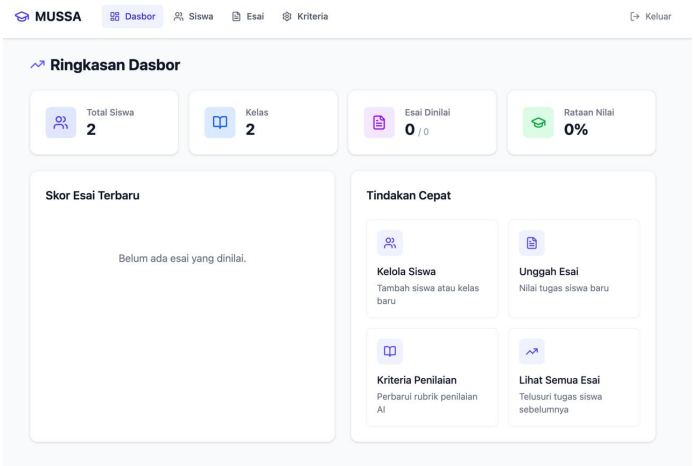

Figure 3. Main Dashboard (left) and Student Management Interface (right)

3. Essay Upload Management and Assessment Criteria Configuration (AI Rubric)

One of the core functionalities of the MUSSA platform is its ability to automatically extract and assess students' handwritten essays using Artificial Intelligence. Teachers simply capture a photograph of a student's handwritten essay and upload the image through the Essay List module. The platform then processes the uploaded document for automated language assessment.

To ensure that AI-generated evaluations remain pedagogically sound and aligned with curriculum standards, MUSSA provides an Assessment Criteria module that enables teachers to configure the scoring rubric before the assessment process begins. Teachers can assign customized weights to each assessment criterion, including Content and Ideas (30 points), Organization and Structure (20 points), Grammar (20 points), Vocabulary and Word Choice (20 points), and Spelling and Punctuation in accordance with the Indonesian Spelling System (PUEBI) (10 points). This configurable rubric allows the AI engine to generate assessments that are transparent, consistent, and aligned with teachers' instructional objectives.

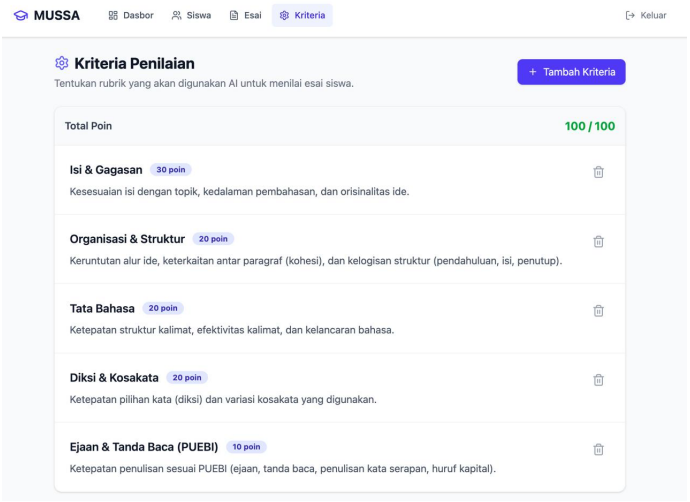

Figure 4. Essay Upload Interface and AI-Based Assessment Rubric Configuration

4. Student Learning Progress Monitoring

MUSSA transforms AI-generated assessment results into interactive visualizations that enable teachers to monitor students' learning progress over time. Through each student's individual profile (e.g., Budi Santoso), teachers can access a complete record of previously submitted essays together with graphical representations of performance trends across successive learning sessions. This longitudinal monitoring feature enables teachers to identify students' strengths and areas requiring improvement in real time, thereby supporting timely instructional interventions and evidence-based decision-making.

The learning progress dashboard also provides comprehensive performance analytics, allowing teachers to evaluate the effectiveness of instructional strategies while tracking each student's literacy development throughout the learning process.

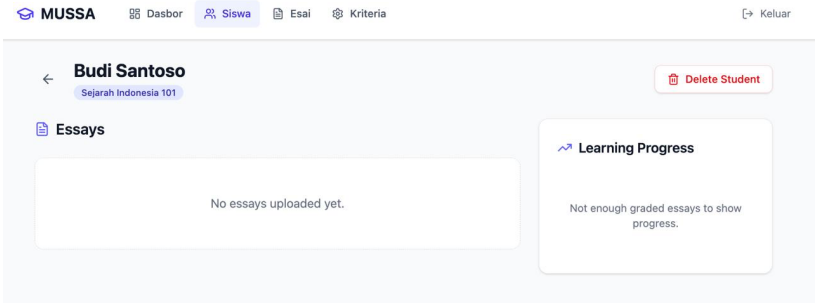

Figure 5. Student Essay History Dashboard and Learning Progress Monitoring

Expert Validation Results

Prior to field implementation, the feasibility of the MUSSA platform was evaluated through an expert validation process involving three specialists: an Indonesian language education expert, an educational assessment expert, and an educational media and information technology expert. The evaluation instrument employed a five-point Likert scale, with scores ranging from 1 (strongly disagree) to 5 (strongly agree). The validation results were analyzed by calculating the feasibility percentage using the following formula (Riduwan, 2012):

$$P = \frac{\sum X}{\sum X_{maks}} \times 100\%$$

Keterangan:

P = Persentase kelayakan

$\sum X$ = Jumlah skor aktual yang diperoleh dari validator

$\sum X_{maks}$ = Jumlah skor maksimal ideal

The level of validity was interpreted based on percentage conversion criteria, in which scores of 81–100% were classified as "Highly Valid/Highly Feasible," 61–80% as "Valid/Feasible," 41–60% as "Moderately Valid," 21–40% as "Less Valid," and below 20% as "Not Valid" (Arikunto, 2020). A summary of the validation results obtained from the three expert validators is presented in Table 1.

Table 1. Expert Validation Results of the MUSSA Platform

Validator	Assessment Indicator	Percentage (%)	Category
Media Expert	User Interface (UI/UX) design and user-friendly navigation.	85%	Highly Valid
	System functionality, including essay upload features, dashboard integration, and Learning Progress metrics visualization.	84%	Highly Valid
	Stability and reliability of the Optical Character Recognition (OCR) process for recognizing and extracting handwritten text from uploaded images.	84%	Highly Valid
	Average Score – Media Expert	85%	Highly Valid
Content Expert	Alignment of the AI assessment rubric with the writing competency standards, including Content, Organization, Grammar, Vocabulary and Word Choice, and Indonesian Spelling and Punctuation (PUEBI).	87%	Highly Valid
	Accuracy of the AI algorithm in identifying language error patterns and generating relevant automated feedback.	85%	Highly Valid
	Appropriateness of the assessment complexity with the cognitive developmental characteristics of Grade 6 primary school students.	85%	Highly Valid
	Average Score – Content Expert	86%	Highly Valid
Overall Result	Overall Feasibility Score of the MUSSA Platform	86%	Highly Valid

As presented in Table 1, the expert validation results provide a detailed evaluation of the MUSSA platform from both technical and pedagogical perspectives. The media expert assigned the highest score (85%) to the User Interface (UI/UX) design and ease of navigation, indicating that the platform offers an intuitive and user-friendly interface that can be operated efficiently by primary school teachers. Likewise, the platform's system functionality and the

reliability of its Optical Character Recognition (OCR) technology were both rated highly, resulting in an overall media validation score of 85%, which falls within the Highly Valid category.

From the pedagogical perspective, the content expert evaluated the quality and appropriateness of the AI-assisted assessment framework. The highest score (87%) was awarded to the alignment of the AI assessment rubric with the established writing competency standards, confirming that the embedded assessment criteria accurately reflect the linguistic and pedagogical requirements of Indonesian language instruction. In addition, the AI's capability to identify language errors and the appropriateness of its assessment complexity for primary school students received highly favorable evaluations. Consequently, the overall content validation score reached 86%, also categorized as Highly Valid.

Overall, the combined validation results from the expert reviewers yielded an average feasibility score of 86%, demonstrating that the MUSSA prototype possesses a high level of technical reliability and pedagogical validity. These findings indicate that the platform satisfies the required quality standards for educational implementation and is suitable for proceeding to the field-testing phase without requiring substantial revisions.

Platform Development Framework

The MUSSA platform was developed by integrating multiple layers of modern technologies to ensure system stability, scalability, and the accuracy of AI-assisted assessment. The technical framework underlying the platform is described as follows.

1. System Architecture

   MUSSA adopts a responsive web-based client–server architecture, which separates the frontend (user interface) from the backend (business logic and data processing). This modular architecture enhances system maintainability, scalability, and overall performance while enabling independent development and optimization of each component. Through this architecture, teachers can access the student management dashboard and upload handwritten essays from a variety of devices, including laptops, tablets, and smartphones. All user interactions are synchronized with the central server in real time, ensuring seamless data consistency and continuous access to assessment records across different devices. This design also facilitates secure data management and supports efficient communication between the AI processing engine and the web application.

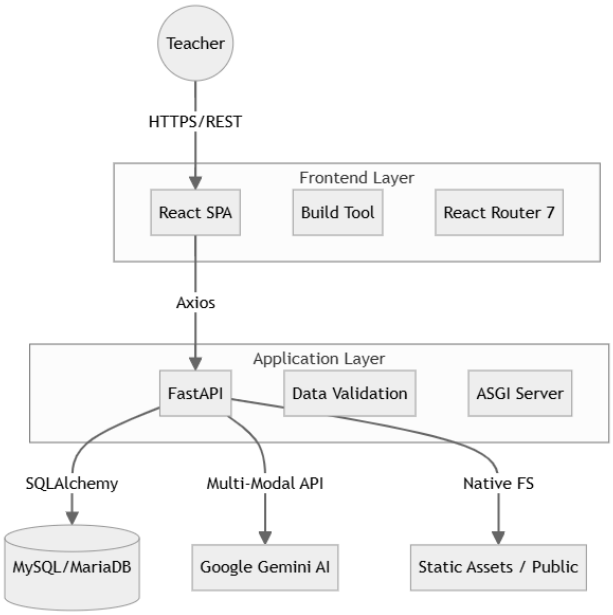

Figure 6. Client–Server Architecture of the MUSSA Platform

The MUSSA platform is built upon a web-based N-tier architecture that integrates a React Single-Page Application (SPA) as the Frontend Layer and the FastAPI framework as the Application Layer. Communication between the frontend and backend is established through secure HTTPS/REST protocols using the Axios library for efficient API interactions. Relational data are managed using MySQL/MariaDB, with SQLAlchemy serving as the Object-Relational Mapping (ORM) framework to facilitate robust and scalable database operations.

The platform's Artificial Intelligence capabilities are powered by a multimodal AI architecture integrated with Google Gemini AI, enabling automated Optical Character Recognition (OCR) for handwritten text extraction and Natural Language Processing (NLP) for linguistic analysis and assessment. The overall software development process followed an iterative Software Development Life Cycle (SDLC), beginning with the design of assessment indicators and continuing through expert validation to ensure the reliability, accuracy, and pedagogical appropriateness of the AI assessment engine.

2. Database Structure

The MUSSA platform employs a relational database management system to handle complex educational data and assessment metadata efficiently. The database schema was designed to support the following core data entities:

- User Profiles : Stores information related to teachers and students, including names, grade levels, and classroom assignments.
- Essay Repository : Maintains uploaded handwritten essay images, OCR-extracted text, and submission history for each student.
- Assessment Metadata : Records AI-generated scores for each assessment dimension—Content, Organization, Grammar, Vocabulary and Word Choice, and Indonesian Spelling and Punctuation (PUEBI)—which are subsequently linked to the Learning Progress dashboard for longitudinal visualization of students' literacy development.

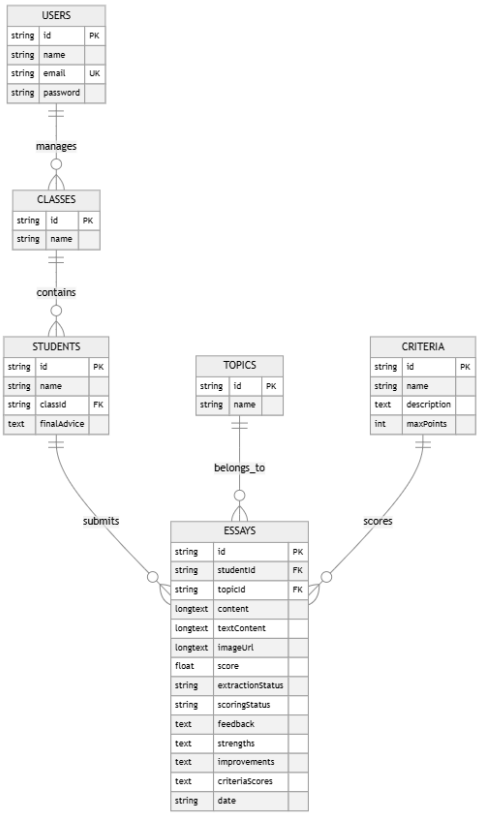

Figure 7. Relational Database Schema for Managing Educational Metadata in the MUSSA Platform

The MUSSA platform employs a relational database model designed to organize educational data systematically in support of AI-assisted assessment. This data architecture establishes a structured management hierarchy in which the Users entity (teachers) manages the Classes entity, which in turn is linked to the Students entity through one-to-many relationships. Such an integrated database structure enables the centralized management of user profiles, classroom information, and individual student learning records.

The core of the assessment process is represented by the Essays entity, which connects student records with the Topics and Assessment Criteria entities through foreign key relationships. This entity stores essential assessment information, including uploaded handwritten essay images (imageUrl), OCR-extracted text (textContent), cumulative assessment scores, and AI-generated automated feedback. The relational schema ensures data consistency, minimizes redundancy, and supports efficient retrieval of longitudinal assessment records for monitoring students' literacy development over time.

3. Software Development Life Cycle (SDLC)

The development of the MUSSA platform followed the ADDIE instructional design model, consisting of the Analysis, Design, Development, Implementation, and Evaluation phases. The development process was conducted iteratively, with each phase undergoing expert validation before progressing to the subsequent stage. This iterative approach ensured that every functional component of the platform—including features such as Quick Actions and the Summary Dashboard—was systematically refined to address teachers' practical needs while maintaining technical reliability and pedagogical effectiveness.

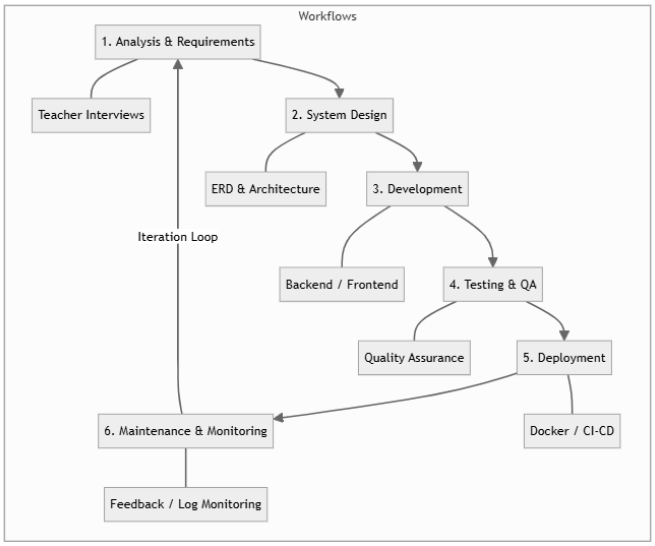

Figure 8. Software Development and Maintenance Cycle of the MUSSA Platform

The development of the MUSSA platform followed a systematic and iterative Software Development Life Cycle (SDLC) to ensure the quality, reliability, and sustainability of the AI-assisted assessment system. The process began with the Analysis and Requirements phase, during which in-depth interviews were conducted with primary school teachers to identify the challenges associated with implementing formative assessment in Indonesian language learning. The findings from this stage informed the functional and technical requirements of the platform. During the System Design phase, the overall system architecture and the Entity Relationship Diagram (ERD) were developed to ensure a well-structured integration of student profiles, assessment data, and AI-based evaluation rubrics. This design phase established the foundation for efficient data management and seamless communication among the platform's functional components. The technical implementation proceeded through the Development phase, which involved the construction of both the frontend and backend components of the

application. This was followed by Testing and Quality Assurance (QA) to verify the accuracy of handwritten text extraction, the reliability of automated language assessment, and the overall stability of the system. Comprehensive testing ensured that the AI-powered assessment engine performed consistently under classroom conditions.

Following successful validation, the platform was deployed using modern software engineering practices, including Docker-based containerization and Continuous Integration/Continuous Deployment (CI/CD) pipelines, to facilitate scalable, reliable, and maintainable deployment. The platform then entered the Maintenance and Monitoring phase, during which system performance, operational logs, and user interactions were continuously monitored. Finally, the development cycle incorporated an Iteration Loop, in which monitoring data and user feedback were systematically analyzed to support continuous improvement of the platform. This iterative refinement process was particularly directed toward enhancing the performance of the Natural Language Processing (NLP) engine, enabling the AI to generate more accurate, context-sensitive, and age-appropriate feedback for primary school students while continuously improving the overall user experience.

DISCUSSION

The MUSSA platform was implemented through a limited-scale field trial involving 74 fifth-grade primary school students from three primary schools in West Java Province, Indonesia, together with three Indonesian language teachers. The field evaluation was conducted through four writing-based learning assignments integrated into regular classroom instruction. The primary objective of this implementation was to examine the practicality of the platform when used in authentic teaching and learning environments.

Practicality data were collected using a four-point Likert-scale questionnaire administered to both teachers and students at the conclusion of the fourth instructional session. The questionnaire consisted of four response categories: Highly Practical (4), Practical (3), Moderately Practical (2), and Less Practical (1). The collected responses were analyzed using descriptive quantitative statistics by calculating the percentage score for each indicator, thereby providing an overall measure of the platform's practicality.

The results of the practicality evaluation are presented in Table 2.

Table 2. Practicality Evaluation Results of the MUSSA Platform

Respondent	Practicality Indicator	Percentage (%)	Category
Teachers (N = 3)	Time efficiency in grading student assignments and managing assessment administration.	88.5%	Highly Practical
	Ease of system operation, including handwritten essay image upload and interpretation of assessment metrics.	91.0%	Highly Practical
	Relevance and accuracy of AI-generated feedback to support instructional guidance and student improvement.	82.3%	Practical*
	Average Teacher Response	**87.26%**	**Highly Practical**

Respondent	Practicality Indicator	Percentage (%)	Category
Teachers (N = 3)	Time efficiency in grading student assignments and managing assessment administration.	88.5%	Highly Practical
	Ease of system operation, including handwritten essay image upload and interpretation of assessment metrics.	91.0%	Highly Practical
	Relevance and accuracy of AI-generated feedback to support instructional guidance and student improvement.	82.3%	Practical*
	Average Teacher Response	**87.26%**	**Highly Practical**

As shown in Table 2, teachers' responses indicated a high level of practicality, with an overall practicality score of 87.2%, categorized as Highly Practical. Teachers reported substantial benefits in terms of time efficiency (88.5%) and ease of system operation (91.0%). However, the relevance of AI-generated feedback received a comparatively lower score (82.3%). Several teachers observed that the feedback generated by the AI was occasionally too rigid, formal, or linguistically sophisticated to be readily understood by primary school students. Consequently, teachers were still required to mediate and verbally explain the AI-generated recommendations before students could effectively benefit from them.

The development and field implementation of the MUSSA (Mobile User Assistant for Smart School Assessment) platform provide empirical evidence that Artificial Intelligence (AI) can support innovation in formative language assessment at the primary education level. This finding is consistent with the broader view that AI in education should function as an intelligent support system that enhances teachers' capacity to analyze learning evidence, provide timely feedback, and make informed instructional decisions rather than replace professional pedagogical judgment (Luckin et al., 2016; Holmes et al., 2019). MUSSA offers a distinctive contribution by integrating several assessment functions within a single platform, including handwritten essay digitization through Optical Character Recognition (OCR), AI-assisted linguistic analysis, configurable assessment rubrics aligned with Indonesian language competencies, centralized student assessment records, and longitudinal learning-progress visualization. Through this integrated workflow, MUSSA extends the function of AI beyond automated scoring by connecting assessment results with continuous monitoring and instructional follow-up. This design directly responds to the practical challenges identified during the Analysis phase, particularly teachers' limited time for manual assessment, the need for timely formative feedback, and the difficulty of monitoring individual students' writing development across successive learning sessions.

Theoretically, these findings reinforce the shift from assessment of learning toward assessment for learning, in which assessment information is not treated merely as a final judgment of achievement but is continuously used to guide subsequent teaching and learning decisions (Redecker & Johannessen, 2013). In the MUSSA platform, AI-generated assessment results are transformed into longitudinal learning information that enables teachers to identify students' strengths, weaknesses, and patterns of development over time. This mechanism illustrates how AI-based assessment can strengthen the formative feedback loop by connecting evidence of student performance with instructional action. Such an approach is also aligned with Holmes et al. (2019), who emphasize the potential of AI to support more adaptive and data-informed educational processes through the analysis of individual learning profiles. Thus, the theoretical contribution of MUSSA lies not simply in automating assessment, but in operationalizing AI as part of an integrated formative assessment ecosystem in which data collection, analysis, feedback, monitoring, and pedagogical decision-making are interconnected.

At the same time, the field implementation demonstrates that AI-supported formative assessment at the primary school level requires strong pedagogical mediation. Although MUSSA was considered highly practical by teachers, the AI-generated feedback was occasionally perceived as too formal, rigid, or linguistically complex for primary school students, requiring teachers to reinterpret the feedback in more accessible language. This finding supports previous concerns that AI systems not specifically calibrated for children may generate responses that are insufficiently sensitive to learners' developmental and linguistic characteristics (Chassignol et al., 2018; Chuang & Yan, 2025). Therefore, the effectiveness of AI-based formative assessment should not be evaluated solely in terms of technological accuracy or automation efficiency, but also in terms of developmental appropriateness, comprehensibility, and pedagogical usability.

From this perspective, MUSSA contributes a more contextualized model of AI-supported assessment for primary education. Its distinctive value lies in combining automated analysis with curriculum-referenced assessment criteria, longitudinal learner data, and teacher-mediated interpretation. This integration reinforces the principle that AI should operate as a pedagogical partner that augments, rather than substitutes, teachers' professional expertise (Luckin et al., 2016; Holmes et al., 2019). The findings therefore imply that future AI-based formative assessment systems for elementary education should be designed around three interrelated principles: technological functionality, pedagogical alignment, and developmental appropriateness. In practical terms, AI should assist teachers in processing complex assessment information rapidly, while teachers remain responsible for interpreting that information, adapting feedback to students' cognitive and linguistic levels, and determining appropriate instructional follow-up. This human–AI complementarity represents an important conceptual direction for the development of more responsive, authentic, and data-informed formative assessment systems in primary education.

The expert validation results, which yielded an average feasibility score of 86% (Highly Valid), demonstrate that MUSSA satisfies both technological and pedagogical standards for assessment instruments. Furthermore, the platform's Optical Character Recognition (OCR) capability successfully transforms handwritten essays into machine-readable digital text, effectively bridging the gap between conventional handwriting practices and contemporary data-driven learning analytics. This capability is particularly significant in primary education, where handwriting remains an essential component of fine motor skill development while schools simultaneously strive to strengthen teachers' and students' digital literacy competencies (Gilster, 1997).

From the perspective of practicality, the highly positive responses from teachers (87.2%) confirm that AI can substantially reduce the administrative burden that has traditionally constrained the implementation of formative assessment. Conventionally, assessing Indonesian language essays requires teachers to manually evaluate multiple linguistic dimensions, ranging from spelling accuracy to paragraph coherence, a process that is both time-consuming and susceptible to cognitive overload and subjective bias (Hanis & Wahyudin, 2024; Sholeh et al., 2024). By contrast, MUSSA provides teachers with an integrated dashboard that automatically detects linguistic errors and summarizes assessment results. This transformation represents a pedagogical shift from assessment of learning, which primarily emphasizes summative outcomes, toward assessment for learning, in which assessment data are continuously utilized to inform instruction and improve students' learning processes (Redecker & Johannessen, 2013).

Despite these advantages, classroom observations and teachers' reports also revealed several limitations. Teachers indicated that the AI-generated recommendations frequently employed overly formal diction, complex sentence structures, and advanced linguistic terminology that exceeded the comprehension level of children aged 11–12 years. As a result, teachers often assumed the role of interpreters, translating the AI-generated feedback into language that was more accessible and meaningful for their students. This finding is consistent with the studies of Chassignol et al. (2018) and Chuang and Yan (2025), which caution that generative AI systems not specifically calibrated for children's education may produce language outputs characterized by excessive linguistic complexity and limited developmental appropriateness.

These limitations provide important directions for the future refinement of the MUSSA platform. In particular, the platform requires improvements in backend prompt engineering to enable its Natural Language Processing (NLP) component to generate child-directed language that is developmentally appropriate for primary school learners. Rather than merely identifying writing errors, future versions of the system should deliver feedback that is communicative,

child-friendly, and motivational, incorporating affirmative language that encourages students to engage in independent reflection and self-assessment.

Overall, the implementation of MUSSA reinforces the theoretical perspective that Artificial Intelligence in primary education is intended to complement rather than replace teachers (Holmes et al., 2019; Luckin et al., 2016). Teachers continue to play the central role in pedagogical interpretation, instructional decision-making, and socio-emotional support, whereas AI contributes by automating complex analytical processes and providing timely, evidence-based assessment information. The integration of these complementary roles demonstrates that authentic, formative, and responsive assessment of Indonesian language skills can be effectively realized while meeting contemporary curricular demands and responding to the ongoing digital transformation of education.

CONCLUSION

This study successfully developed and evaluated the feasibility of MUSSA (Mobile User Assistant for Smart School Assessment), a web-based AI-assisted assessment platform, using the ADDIE instructional design model. The expert validation results demonstrated that MUSSA achieved a Highly Valid rating, with an overall feasibility score of 86%, indicating its suitability as an assessment platform for Indonesian language learning in primary education. The longitudinal field implementation involving 74 fifth-grade primary school students further demonstrated that the platform substantially reduced teachers' administrative workload, achieving an overall practicality score of 87.2% (Highly Practical). These findings demonstrate that MUSSA has a high level of feasibility and practicality for supporting teachers in conducting authentic and continuous formative assessment and in improving the efficiency of classroom assessment practices. Nevertheless, the present study did not examine the platform's effectiveness in improving students' learning outcomes or writing performance. Accordingly, the conclusions of this study should be interpreted within the scope of feasibility and practicality rather than learning effectiveness.

Nevertheless, the study also revealed that the language used in the AI-generated feedback tended to be overly formal and linguistically complex for primary school students. Consequently, future research should focus on refining the platform's Natural Language Processing (NLP) capabilities and backend prompt engineering to generate feedback that is more child-friendly, developmentally appropriate, and supportive of students' independent self-assessment.

Overall, the findings reinforce the view that Artificial Intelligence should function as a pedagogical partner rather than a substitute for teachers. By combining teachers' pedagogical expertise with the analytical capabilities of AI, the MUSSA platform demonstrates the potential to establish a more effective, authentic, and data-informed formative assessment ecosystem for Indonesian language learning in primary schools.

REFERENCES

Al-Imamah, A., et al. (2023). Harnessing the Use of Artificial Intelligence in Language Assessment: A Systematic Comprehensive Review. *TELL-US Journal*, 9(3), 7366. https://doi.org/10.22202/tus.2023.v9i3.7366

Arikunto, S. (2020). *Evaluasi Program Pendidikan: Pedoman Teoritis Praktis Bagi Mahasiswa dan Praktisi Pendidikan*. Bumi Aksara.

Borg, W. R., & Gall, M. D. (2003). *Educational research: An introduction*. Allyn & Bacon.

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. https://doi.org/10.1007/978-0-387-09506-6

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. *Procedia Computer Science*, 127, 16-27. https://doi.org/10.1016/j.procs.2018.01.003

Chuang, P.-L., & Yan, X. (2025). Language assessment in the era of generative artificial intelligence: Opportunities, challenges, and future directions. *System*, 134, 103846. https://doi.org/10.1016/j.system.2025.103846

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Fathurrachman, S., & Saputri, N. E. (2025). Teacher's Perceptions and Institutional Preparedness for Implementing AI in Learning Assessment at The Elementary School Level. *IQRO: Journal of Islamic Education*, *8*(1), 49-60. https://doi.org/10.24256/iqro.v8i1.6484

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. Frontiers in Artificial Intelligence, 6, 1162454. https://doi.org/10.3389/frai.2023.1162454

Gilster, P. (1997). *Digital literacy*. Wiley Computer Publishing.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66(1), 64–74. https://doi.org/10.1119/1.18809

Hanis, M., & Wahyudin, D. (2024). Pemanfaatan Artificial Intelligence (AI) Dalam Penyusunan Asesmen Pembelajaran Bagi Guru Sekolah Dasar. *Jurnal Ilmiah Profesi Pendidikan*, 9(2), 1199–1207. https://doi.org/10.29303/jipp.v9i2.2252

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign. https://doi.org/10.1007/978-3-030-23707-9

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.

Molenda, M. (2015). In search of the elusive ADDIE model. *Performance Improvement*, 54(2), 40–42. https://doi.org/10.1002/pfi.21461

Purba, N., Pujiati, D., et al. (2025). The Use of AI in Elementary School Learning: A Systematic Literature Review. *QALAMUNA Jurnal Pendidikan*, 17(1), 6761. https://doi.org/10.37680/qalamuna.v17i1.6761

Redecker, C., & Johannessen, Ø. (2013). Changing assessment — Towards a new assessment paradigm using ICT. *European Journal of Education*, 48(1), 79–96. https://doi.org/10.1111/ejed.12018

Richey, R. C., & Klein, J. D. (2007). *Design and development research*. Lawrence Erlbaum Associates.

Riduwan. (2012). *Dasar-dasar Statistika*. Alfabeta.

Sabri, M., & Wais, A. (2025). Artificial Intelligence dalam Pengukuran dan Penilaian Pendidikan: Kajian Literatur dari Perspektif Inovasi Evaluatif. *Jurnal Evaluasi Pendidikan*, 16(2), 95–107. https://doi.org/10.21009/jep.v16i2.60738

Sabri, M., & Wais, A. (2025). Artificial intelligence dalam pengukuran dan penilaian pendidikan: Kajian literatur dari perspektif inovasi evaluatif. Jurnal Evaluasi Pendidikan, 16(2), 95–107. https://doi.org/10.21009/jep.v16i2.60738

Sholeh, M. I., Sokip, Syafi'i, A., & Sulistyorini. (2024). Artificial Intelligence in Education: Transforming Assessment and Personalized Learning. *Journal of Education and Learning Sciences*, 4(1), 118. https://doi.org/10.56404/jels.v4i1.118

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. Learning and Instruction, 91, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Sugiyono. (2017). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators?. *International Journal of Educational Technology in Higher Education*, 16(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhai, X., He, P., & Li, Y. (2020). Examining the effects of artificial intelligence–based educational tools on students' literacy learning: A systematic review. *Computers & Education*, 151, 103857. https://doi.org/10.1016/j.compedu.2020.103857