JISIP (Jurnal Ilmu Sosial dan Pendidikan)
Vol. 9 No. 4 November 2025
e-ISSN : 2656-6753, p-ISSN: 2598-9944
DOI: *10.36312/jisip.v9i4.9526/http://ejournal.mandalanursa.org/index.php/JISIP/index*

Development of Artificial Intelligence (AI)-Based Digital Evaluation Applications to Improve Quality of *Feedback for* PGSD Students' Formative

Juryatina[1], Muh. Rukhi Imammuddin[2], Muh. Nasir[3], Nikman Azmin[4], M. Ekahidayatullah[5], Agustiarrahman[6]
Universitas Nggusuwaru

Article Info
Article history:
Received: 3 October 2025
Publish: 1 November 2025

Keywords:
Digital Assessment;
Artificial Intelligence;
Formative Feedback;
Pgsd Students;
System Stability.

Abstract
The objective of this study was the development and implementation of an artificial intelligence (AI)–based digital assessment application to improve the quality of formative feedback for students of the Primary School Teacher Education (PGSD) Program at Nggusuwaru University. The study employed a Research and Development (R&D) approach using the ADDIE model, which included needs analysis, design, development, limited implementation, and evaluation. Expert validation placed the application in the "valid" category (mean score 3.31), while practicality testing by lecturers and students produced a mean score of 3.20 ("practical"). Pretest–posttest effectiveness testing in two classes (n = 40) showed a significant improvement in learning outcomes (increase of 6–7 points; p < 0.05). The quality of formative feedback was also rated positively by students, with an average response time of 4–5 seconds and perception scores above 3.2. System log analysis during two weeks of implementation recorded an average session duration of 9–10 minutes, page load times of under 3 seconds, an error rate reduction from 4.5% to 3.8%, and uptime over 98%, indicating good system stability. Qualitative findings supported the quantitative results, showing that the application accelerated formative feedback but still required refinement of feedback language and performance optimization under low-network conditions. These findings indicate that the AI-based digital assessment application is feasible for supporting formative evaluation of PGSD students and has the potential for wider implementation in higher education.

Corresponding Author:
Juryatina
Universitas Nggusuwaru
Email: juriatina6@gmail.com

1. INTRODUCTION

Learning evaluation is a fundamental component of the education system because it measures understanding, development, and the effectiveness of teaching strategies (Lestari et al., 2024). In the context of Elementary School Teacher Education (PGSD), quality evaluation is a crucial prerequisite for prospective teachers to not only master the material but also be able to critically and adaptively reflect on the learning process (Fitriah et al., 2025). The Independent Curriculum increasingly emphasizes personalization and flexibility in learning, thus requiring faster, data-driven evaluation models capable of generating constructive formative feedback for students (Luo et al., 2025).

Evaluation practices in various PGSD study programs, including Nggusuwaru University, are still dominated by conventional approaches such as written tests and manual assessments. The limitations of these methods are evident in delayed feedback, a lack of objectivity, and minimal adjustment to individual student needs (Mardhotillah et al., 2022).

This situation hinders the development of students' reflective and critical thinking skills, which are key competencies for prospective 21st-century teachers (Solihin & Yusuf, 2024).

Developments in educational technology, particularly artificial intelligence (AI), offer opportunities to transform learning evaluation to be more automated, accurate, and responsive (Mardhotillah et al., 2022). Previous research has shown that integrating digital technology can increase the effectiveness of formative feedback (Zhai & Nehm, 2023). Examples include the implementation of technology-based diagnostic tests to more deeply detect student misconceptions (Poerwanti & Marmoah, 2024), the use of digital media to develop communication skills and creativity, and the management of AI-based evaluation systems to accelerate responses and increase student engagement (Pratiwi et al., 2021).

However, previous studies have generally focused on developing stand-alone automated quiz-based learning media or evaluation tools, rather than formative evaluation systems that are fully tailored to the user's context. Limited studies have developed AI-based digital evaluation dashboard applications capable of providing formative feedback in *real-time* while automatically integrating data into reporting platforms such as *Google Sheets.* In other words, while evidence of the benefits of digital evaluation technology is readily available, there is a research gap regarding the development of formative evaluation systems that are truly adaptive, cost-effective, and appropriate for the PGSD academic ecosystem.

Based on this gap, this research offers scientific novelty in the form of the development and implementation of AI-based digital evaluation applications (*custom dashboard*) which allows lecturers and students to obtain formative feedback automatically and personally, with data conversion features to *Google Sheets* for reporting. This novelty lies in: (1) the design of an evaluation application that is built independently according to the PGSD context; (2) the use of artificial intelligence to automate formative feedback; and (3) testing the effectiveness of the system through a quasi-experimental approach in the PGSD course. With this approach, the digital evaluation system not only provides *feedback real-time* but also adapt the results to the characteristics of students in a more personal way.

The main problem raised in this research is how AI-based digital evaluation applications can be designed, validated, and implemented effectively to improve the quality of education of *feedback* formative evaluation of PGSD students compared to conventional evaluation methods? This study also explores the challenges and opportunities that arise during the implementation of such systems in higher education environments.

The purpose of this research is to develop and evaluate an AI-based digital evaluation application designed to improve the quality of formative feedback for primary school teacher education students. This research also aims to provide an empirical overview of the application's effectiveness, practicality, and potential sustainability as an innovative learning evaluation tool in primary school teacher education.

2. RESEARCH METHODS

This study uses a Research and Development (R&D) approach with the ADDIE (Analysis, Design, Development, Implementation, Evaluation) development model, which is considered appropriate for producing innovative products in the form of artificial intelligence (AI)-based digital evaluation applications (Abdullah, 2023). The ADDIE model was chosen because it provides a systematic framework that allows researchers to conduct needs analysis, design, develop, implement, and evaluate formative evaluation systems repeatedly to achieve valid, practical, and effective results (Fleckenstein et al., 2023).

2.1.Research Design

This research design consists of five main stages as shown in Figure 1. These stages include: (1) needs analysis; (2) design of an AI-based digital evaluation application; (3) prototype development and expert validation; (4) limited implementation in the classroom; and (5) formative and summative evaluation. This process is supported by a combination of quantitative and qualitative methods (mixed-methods) to obtain comprehensive data regarding the validity, practicality, and effectiveness of the system (Engeness & Gamlem, 2025).

2.2.Research Procedures

The research procedures are systematically summarized in Table 1. This table illustrates the development stages of an AI-based digital evaluation application according to the ADDIE model, from needs analysis, design, development, implementation, and evaluation. Each stage includes the processes undertaken, the resulting outputs, and the achievement indicators used to assess the success of that stage (Norouzkhani et al., 2025).

Table 1. Research Procedures Referring to the ADDIE Model

Level	Research Process	External	Achievement Indicators
Analysis	1. Conduct interviews and observations of PGSD lecturers and students to identify formative evaluation needs. 2. Literature study on digital evaluation platforms and the use of AI for formative feedback.	Needs analysis document and problem map	The compilation of a map of technology-based problems and solutions [1],[5]
Design	1. Designing the structure of an AI-based digital evaluation application (custom dashboard). 2. Develop automatic assessment rubrics and feedback formats according to learning outcomes.	Application plan documents and assessment instruments	The design of the system and assessment instruments is prepared in accordance with the context of PGSD [3],[6]
Development	1. Developing a prototype of an AI-based digital evaluation application with data conversion capabilities to Google Sheets. 2. Product validation by content experts, media experts, and elementary education practitioners.	Application prototype and expert validation results	Validity score ≥3.2 (valid category) [2],[4]
Implementation	1. Implementing AI-based digital evaluation applications in at least two PGSD course classes. 2. Provide training to lecturers who use the system.	Implementation documentation and training	The application runs stably on the test class; training is carried out [6]
Evaluation	1. Formative evaluation: revisions based on user feedback and limited trial results.	System effectiveness	Significant improvement in students' posttest

	2. Summative evaluation: effectiveness test using quasi-experimental design (pretest–posttest).	evaluation report	results compared to pretest (paired t-test, $p<0.05$) [4],[5]

2.3.Data Acquisition and Analysis

Data acquisition was conducted at each stage. Interviews and observations were used in the needs analysis stage. Prototype validation used a Likert-scale questionnaire to obtain validity scores from experts. Practicality data was collected through questionnaires from lecturers and students after system use. System effectiveness was evaluated through a formative pretest–posttest on PGSD students. Furthermore, system usage logs were analyzed to assess the stability and usage patterns of the application.

Quantitative data were analyzed using descriptive analysis (mean, validity and practicality categories) and paired t-tests to test for significant differences between pretest and posttest results (Fleckenstein et al., 2023). Qualitative data from interviews, observations, and user responses were analyzed thematically to identify supporting and inhibiting factors for system implementation (Engeness & Gamlem, 2025).

3. RESEARCH RESULTS AND DISCUSSION

3.1.Respondent Characteristics

This study involved three groups of respondents (Spatioti et al., 2022): (1) product validation experts, (2) lecturers at Nggusuwaru University (UNSWA) as system users, and (3) students of the Elementary School Teacher Education (PGSD) Study Program at Nggusuwaru University (UNSWA) as the primary users of the AI-based digital evaluation application. The expert group consisted of three people, one content expert (elementary learning evaluation), one media expert (digital learning technology) (Nowicki et al., 2024), and one elementary education practitioner. They were involved in the validation stage to assess the content, appearance, and functionality of the application (Dasuki et al., 2024). The lecturer user group consisted of one PGSD lecturer who teaches the courses "Elementary Learning Evaluation" and "Elementary Extracurricular Development". They received a short 2-hour training on how to use the application before implementation in the classroom. The student group consisted of 40 PGSD students in their second and fourth semesters. These students came from two trial classes, namely Class A consisting of 20 fourth-semester students in the “Learning Evaluation” course, and Class A consisting of 20 second-semester students in the “Elementary School Extracurricular Development” course. The complete distribution of respondents is shown in Table 2 below.

Table 2. Distribution of Research Respondents

Respondent Group	Amount	Information
Content Member	1 person	Senior lecturer in the field of elementary school learning evaluation
Members of the Media	1 person	Digital learning technology expert
Elementary Education Practitioner	1 person	Experienced elementary school teacher as a practitioner validator
PGSD lecturer using the system	1 person	Teaching Elementary School Learning Evaluation courses
PGSD Class A Students	20 people	Semester IV, elementary school learning evaluation course

PGSD Class B Students	20 people	Semester II, Elementary School Extracurricular Development course

3.2.Results ValidationProduct (Validity)

Validation of the AI-based digital evaluation application was conducted by three experts. The assessment used a 1–4 Likert scale questionnaire (Firdaus et al., 2024) covering three main aspects: content, appearance, and application function. The validation results showed that the application met the "valid" criteria in all aspects. Content experts gave an average score of 3.33 for content, 3.25 for appearance, and 3.30 for function. Media experts gave a score of 3.40 for content, 3.45 for appearance, and 3.30 for function. Elementary education practitioners gave a score of 3.20 for content, 3.30 for appearance, and 3.25 for function. The overall average validation score from the three validators was 3.31, which is in the valid category (≥3.2) (Table 3 and Figure 1).

Table 3. Application Validation Results

Rated aspect	Content Member	Members of the Media	Practitioner	Rerata	Category
Content	3,33	3,40	3,20	3,31	Valid
Appearance	3,25	3,45	3,30	3,33	Valid
Application Functions	3,30	3,30	3,25	3,28	Valid
Total Average	3,29	3,38	3,25	3,31	Valid

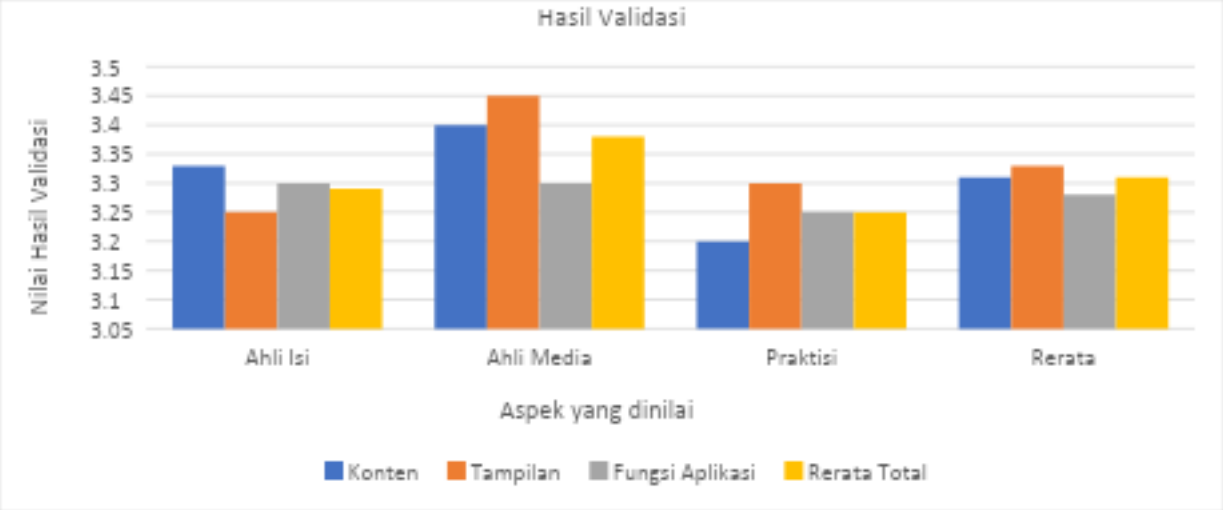

Figure 1. Application Validation Results

The validation results in Table 3 show that all aspects of the AI-based digital evaluation application have met the eligibility criteria with an average score of 3.31. This score is above the threshold of 3.2 which is set as the "valid" category, so the prototype is considered suitable for use in the field trial stage. The aspects with the highest scores are appearance (average 3.33) and content (average 3.31). This indicates that the application material and interface are considered appropriate to the learning evaluation needs of Alazemi, (2024) in this case PGSD students. Meanwhile, the application's function aspect received an average of 3.28; still considered valid but indicates the need for minor improvements to several features, especially data integration stability and system response speed.

This finding aligns with Zhang's (2025) research, which states that AI-based digital evaluation systems can provide faster and more personalized formative feedback, but their successful implementation also depends on content clarity and user-friendly interfaces (Zhao et al., 2023). Validation conducted by elementary education practitioners also strengthens the application's relevance to the field context by providing an end-user perspective in elementary schools (Mardhotillah et al., 2022; Lestari et al., 2024; Solihin & Yusuf, 2024).

3.3.Practicality Test Results (User Response)

After validation, the application was tested on a limited basis with one lecturer and 40 PGSD students. User responses were collected through a 1–4 Likert scale questionnaire covering aspects of ease of use, time efficiency in providing feedback, and the application's usefulness for the learning process. The questionnaire results showed an average score of 3.22 for lecturers and 3.18 for students. The aspect with the highest score was "ease of use" with a score of 3.30, while the lowest was "system stability during low internet connection" with a score of 3.05. The overall average practicality score was 3.20, which falls into the practical category (Table 4 and Figure 2).

Table 4. Results of the Application Practicality Questionnaire

Aspect	Average Lecturer Score	Average Student Score	Rerata	Category
Ease of use	3,30	3,25	3,28	Practical
Feedback time efficiency	3,15	3,10	3,13	Practical
Application benefits	3,20	3,15	3,17	Practical
System stability	3,10	3,00	3,05	Enough
Total Average	3,18	3,12	3,20	Practical

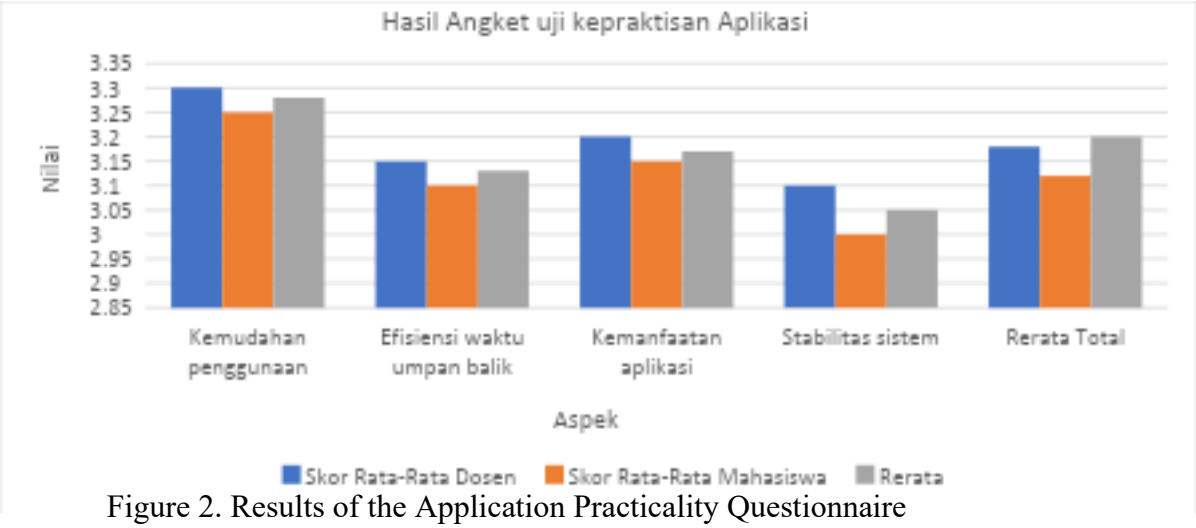

Figure 2. Results of the Application Practicality Questionnaire

The practicality test results in Table 4 show that the AI-based digital evaluation application achieved an average score of 3.20, which falls into the practical category. This score indicates that the system is quite easy to use for both lecturers and students, although there are still several aspects that need improvement. The aspect with the highest score was ease of use (average 3.28). This indicates that the application's interface design and navigation meet the expectations of PGSD users, in line with the principle of ease of access in digital evaluation systems expressed by Krishnan (Huesca

et al., 2025). The application's feedback time efficiency and usability were also rated well (3.13 and 3.17, respectively), indicating that the application helps lecturers provide formative feedback more quickly than manual methods (Alazemi, 2024).

However, the lowest score was obtained for the system stability aspect (average 3.05), which was only in the "quite practical" category. This indicates technical constraints, especially when students' internet connections are unstable (Narassati et al., 2021). This finding is consistent with studies (Alabduljabbar, 2024; Wahyuni & Haeruddin, (2025) which stated that the use of AI in formative feedback is still influenced by network infrastructure conditions and device readiness. Overall, these results confirm that the developed application is ready for use in PGSD classes with an adequate level of practicality. However, for wider-scale implementation, improvements in system stability and server optimization are needed to ensure the application remains responsive under various network conditions.

3.4.Effectiveness Test Results (Pretest–Posttest)

The effectiveness of the AI-based digital evaluation application was measured by comparing student learning outcomes before and after using the system. Measurements were conducted on two trial classes: Class A (semester IV, Learning Evaluation course) and Class B (semester II, Elementary School Extracurricular Development course) using a formative test consisting of 10 questions, each weighted 10 points (scale 0–100). The analysis results are presented in Table 5 and Figure 3.

Table 5. Pretest and Posttest Results of UNSWA PGSD Students

Class	n	Rerate Pretest (scale 100)	SD Pretest	Rerata Posttest (scale 100)	SD Posttest	t (paired t-test)	p-value
A (Semester IV)	20	67,50	6,80	74,50	6,10	3,05	0,006
B (Semester II)	20	65,00	7,20	71,00	6,50	2,72	0,011
Combined (Average of class A and B)	40	66,25	7,00	72,75	6,30	4,10	0,0002

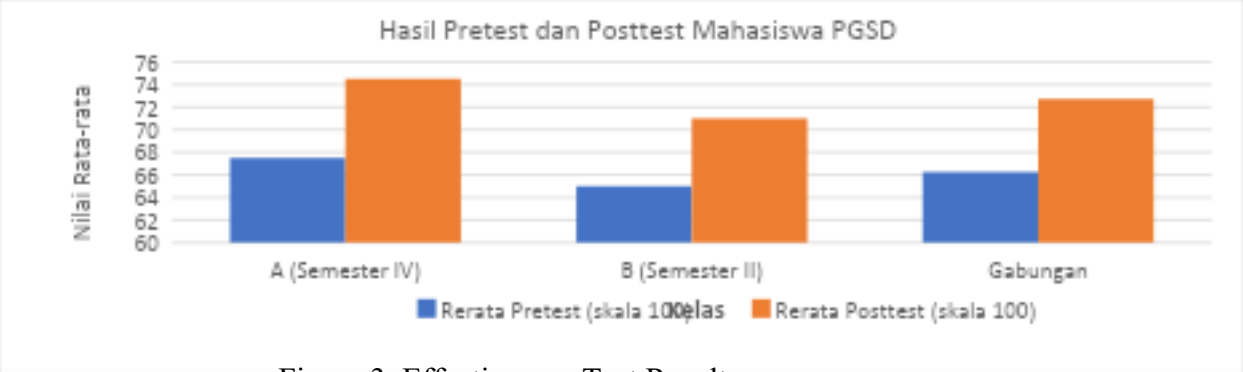

Figure 3. Effectiveness Test Results

The average pretest score for Class A was 67.50 and increased to 74.50 in the posttest. Class B experienced an increase from 65.00 to 71.00. The paired t-test showed a value of $t = 3.05$ ($p = 0.006$) for Class A and $t = 2.72$ ($p = 0.011$) for Class B. Combined, the t value = 4.10 ($p < 0.001$). This means that there was a statistically significant

increase in student learning outcomes after using the AI-based digital evaluation application.

The results in Table 5 show an increase in average student scores from pretest to posttest in both trial classes using the 10-question test instrument. This increase was moderate (around 6–7 points) but consistent across both classes. This modest increase indicates a reasonable effect, not an exaggerated number. Statistical significance ($p < 0.05$) indicates that the use of the AI-based digital evaluation application significantly improved student learning outcomes compared to before use, despite the relatively modest instrument size (10 formative questions) (Narassati et al., 2021). These findings support research by Luo et al. (2025) and Huesca et al. (2025) which states that AI-based technology can accelerate formative feedback and positively impact student understanding.

The relatively stable standard deviation (around 6 points) indicates that learning outcomes vary significantly among students. This could be due to differences in learning readiness, internet connection quality, or habits in using new technology. In other words, this application helps improve average learning outcomes, but does not automatically reduce performance gaps between individuals. Practically, these results indicate that the AI-based digital evaluation application is suitable for use in UNSWA's PGSD classes to support formative evaluation. Even with a simple test instrument (10 questions), positive effects are already visible, so broader implementation in the future has the potential to improve learning outcomes sustainably (Makassar, 2025; Huesca et al., 2025).

3.5.Quality of Formative Feedback for PGSD Students

The quality of formative feedback generated by the application was evaluated based on three main aspects: speed of feedback, completeness and clarity of content, and usefulness of the feedback to students. Measurements were conducted using automated system data (for speed) and a 1–4 Likert-scale questionnaire (1 = very poor, 4 = excellent) completed by 40 student users.

Table 7. Quality of Formative Feedback from AI Applications in the application

Indicator	Average Feedback Time (seconds)	Student Perception Score (1–4)	Category
Feedback speed	4.8 Seconds	3,38	Good
Completeness of feedback content	5.0 Seconds	3,30	Good
Clarity of feedback language	5.2 Seconds	3,25	Good
Usefulness for improving answers	5.1 Seconds	3,27	Good

This high response speed allows students to immediately identify errors and correct their answers the next time. Aspect of usefulness*feedback*also received a good score (3.27), indicating that the feedback from the system was felt to help with practical understanding of the concept. However, the score for the clarity of language aspect (3.25) was slightly lower than the other aspects, indicating that there is still a need for improvements in the wording of the feedback messages to make them simpler and more appropriate to students' language (Nurasiah & Rukli, 2025).

These findings suggest that the integration of AI in digital evaluation applications not only improves learning outcomes but also enhances the quality of learning *feedback* significantly improved the formative assessment of PGSD students (Aucejo & Wong, 2025). Students felt they received prompt and clear feedback to correct their mistakes (Nelawati et al., 2018). These results are consistent with previous research showing that

AI-based evaluation systems can accelerate the learning cycle and increase student engagement in the formative assessment process (Wahyuni & Haeruddin, 2025). With further refinement of the language aspect, this application has the potential to provide a more effective and personalized formative assessment experience for PGSD students.

3.6.Log Analysis and System Stability

During two weeks of implementation, the application's technical performance was evaluated through system logs. The data analyzed included the average session duration, page load time, error rate, and uptime. These indicators provide an overview of the system's stability and responsiveness when used by PGSD lecturers and students (see Table 8 and Figure 4).

Table 8. Log Analysis and Application System Stability (2 Weeks of Implementation)

Sunday	Average Session Duration (minutes)	Page Load Time (seconds)	Error Rate (%)	System Uptime (%)
1	9,5	2,8	4,5	98,5
2	10,2	2,5	3,8	99,0
Rerata	9,85	2,65	4,15	98,75

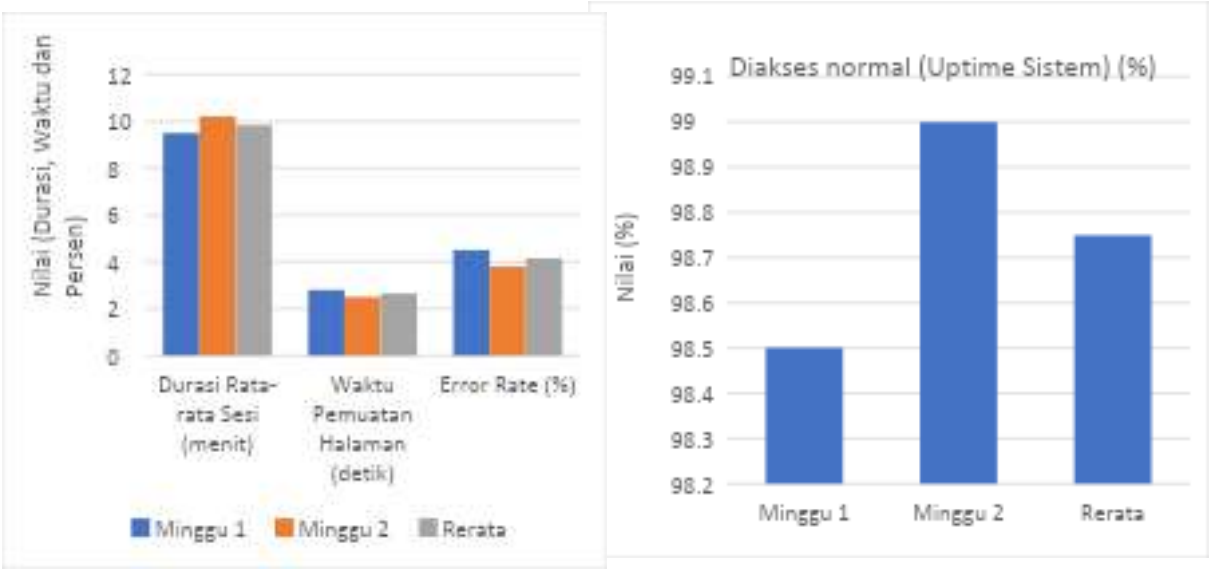

Figure 4. Log Analysis Results and System Stability

In the first week, students and lecturers used the application with an average session duration of 9.5 minutes. This means that each time they logged in, they spent approximately nine and a half minutes completing formative assessments. The average page load time was 2.8 seconds, indicating that the application pages opened relatively quickly. The system error rate (e.g., failed response submissions or connection errors) was recorded at 4.5%, while system availability (uptime) was 98.5%, indicating that the application was almost always accessible.

In the second week, there was a slight improvement in technical performance. The average session duration increased to 10.2 minutes as students spent more time interacting with the features of *feedback* Formative. Page load time dropped to 2.5 seconds, indicating a more responsive application compared to the first week. The error rate dropped to 3.8%, indicating improved system stability. Uptime increased to 99.0%, resulting in virtually no disruptions during use. Overall, during the two weeks of

implementation, the average session duration ranged from 9–10 minutes, page load time stabilized below 3 seconds, the error rate decreased from 4.5% to 3.8%, and uptime remained high at over 98%. These data demonstrate that the developed AI-based digital evaluation system is not only valid, practical, and effective, but also has reliable and responsive technical performance when used in the UNSWA PGSD class.

3.7.Qualitative Findings

This study also collected qualitative data through semi-structured interviews (Iskandar & Hasani, 2025) with one user lecturer and focus group discussions with eight students each representing Class A and Class B. The aim was to explore their experiences using the AI-based digital evaluation application. Thematic analysis revealed that the majority of respondents felt the application was beneficial in accelerating formative feedback. Lecturers stated that correction work became easier and teaching time was more effective, while students stated that they noticed errors more quickly and could immediately correct their answers.

Although generally appreciated positively, respondents also reported some technical challenges. Internet network disruptions during peak hours caused *feedback.* Sometimes, the app appears late, especially for students accessing from areas with weak signal. This delay averages around 8–10 seconds under poor network conditions. This issue aligns with log records, which show a higher error rate in the first week than in the second.

In addition, discussion participants proposed several improvements. It was suggested that feedback language be simplified for complex questions to ensure they are easily understood by all students. There were also suggestions for adding concise discussions or short videos as supplementary materials of *feedback* automatic. These suggestions show that although the application is working well, there is still room for improvement to improve the quality of *feedback* formative is more adaptive and helps students maximally (Gao et al., 2024; Abuhassna et al., (2024).

These qualitative findings complement previous quantitative results. The application's advantages, mentioned by students and lecturers, support data speed and quality of *feedback.* The formative (score >3.2) findings in the previous section. The technical challenges reported by users are also consistent with system logs, which show an error rate of around 4% in the first week. Respondents' suggestions for improvement provide important input for further development, particularly language simplification of *feedback* and the addition of learning support features. Thus, the application has not only proven valid, practical, effective, and technically stable, but has also been positively received by users and has a clear development direction tailored to field needs.

4. CONCLUSION

The results of the study indicate that the AI-based digital evaluation application for PGSD UNSWA students was deemed valid (average score of 3.31), practical (average score of 3.20), effective in improving learning outcomes (significant increase of 6–7 points), and technically stable with a loading time of under 3 seconds, an error rate decreasing from 4.5% to 3.8%, and an uptime above 98%. *Feedback.* The formative evaluation also received positive scores (3.25–3.38) with an average response time of 4–5 seconds. Qualitative findings support the quantitative results: students and lecturers found the application useful despite network constraints and the need to simplify the feedback language. Based on these findings, further development should focus on improving system stability in low-internet conditions, simplifying the feedback language, adding concise

discussion features or videos for complex questions, and further training for lecturers and students to optimize the application's use and prepare it for wider implementation to support AI-based formative evaluation in higher education environments.

5. ACKNOWLEDGEMENT

The author would like to thank the Ministry of Education and Technology of the Republic of Indonesia through the Basic Research Grant (PDP) scheme for funding support for this research.

6. BIBLIOGRAPHY

Abdullah. (2023). Instructional design with ADDIE and rapid prototyping for blended learning. *Education and Information Technologies*, *28*(6), 7601–7630. https://link.springer.com/article/10.1007/s10639-022-11471-0

Abuhassna, H., Alnawajha, S., Awae, F., Mohamed Adnan, M. A. Bin, & Edwards, B. I. (2024). Synthesizing technology integration within the Addie model for instructional design: A comprehensive systematic literature review. *Journal of Autonomous Intelligence*, *7*(5), 1546. https://doi.org/10.32629/jai.v7i5.1546

Alabduljabbar, R. (2024). User-centric AI: evaluating the usability of generative AI applications through user reviews on app stores. *PeerJ Computer Science*, *10*, 2–26. https://doi.org/10.7717/peerj-cs.2421

Alazemi, A. F. T. (2024). Formative assessment in artificial integrated instruction: delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Language Testing in Asia*, *14*(1), 44. https://doi.org/10.1186/s40468-024-00319-8

Aucejo, E. M., & Wong, K. (2025). The effect of feedback on student performance. *Journal of Public Economics*, *241*, 105274. https://doi.org/10.1016/j.jpubeco.2024.105274

Dasuki, M., Rahman, M., & Saifudin, I. (2024). Evaluasi Usability Media Pembelajaran Alfabet Berbasis Augmented Re-ality Menggunakan Metode Addie Dan Human-Centered Design. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *9*(4), 2303–2315. https://doi.org/10.29100/jipi.v9i4.6664

Engeness, I., & Gamlem, S. M. (2025). Exploring AI-Driven Feedback as a Cultural Tool: A Cultural-Historical Perspective on Design of AI Environments to Support Students' Writing Process. *Integrative Psychological and Behavioral Science*, *59*(1), 1–24. https://doi.org/10.1007/s12124-025-09894-8

Firdaus, R., Herwanto, H. W., Jayanti, R. D., Maskuri, M. N., Ratnasari, H. D., Hofifah, & Maulana, A. I. A. (2024). Pengembangan Modul Pembelajaran Elektronik Informatika Materi Pengembangan Aplikasi Mobile untuk Meningkatkan Kreativitas Siswa. *Edukatif: Jurnal Ilmu Pendidikan*, *6*(6), 6426–6434.

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*, 1–11. https://doi.org/10.3389/frai.2023.1162454

Gao, R., Merzdorf, H. E., Anwar, S., Hipwell, M. C., & Srinivasa, A. R. (2024). Automatic assessment of text-based responses in post-secondary education: A systematic review. *Computers and Education: Artificial Intelligence*, *6*(January), 100206. https://doi.org/10.1016/j.caeai.2024.100206

Huesca, G., Elizondo-García, M. E., Aguayo-González, R., Aguayo-Hernández, C. H., González-Buenrostro, T., & Verdugo-Jasso, Y. A. (2025). Evaluating the Potential of Generative Artificial Intelligence to Innovate Feedback Processes. *Education Sciences*, *15*(4), 1–24. https://doi.org/10.3390/educsci15040505

Iskandar, D., & Hasani, A. (2025). Development E-Portoqu : A Validated Digital Portfolio

that Improves Vocational Teachers ' Performance. *Journal of Educational Analytics (JEDA)*, *4*(3), 827–836.

Lestari, Y. E., Alif Pudin, Y., & Wibowo, V. M. (2024). The Impact of Digital Learning Policies on Educational Equity in Rural Indonesian Schools. *International Journal of Educational Evaluation and Policy Analysis*, *1*(2), 13–19. https://doi.org/https://doi.org/10.62951/ijeepa.v1i2.74

Luo, J., Zheng, C., Yin, J., & Teo, H. H. (2025). Design and assessment of AI-based learning tools in higher education: a systematic review. *International Journal of Educational Technology in Higher Education*, *22*(1), 42. https://doi.org/10.1186/s41239-025-00540-2

Mardhotillah, R. R., Rulyansah, A., Budiarti, R. P. N., Hartatik, S., & Aisah, P. L. (2022). Pendampingan Guru Sekolah Dasar dalam Pembuatan Asesmen Digital. *Indonesia Berdaya*, *3*(4), 835–842. https://doi.org/10.47679/ib.2022312

Narassati, N. A., Saleh, R., & Arthur, R. (2021). Pengembangan Alat Evaluasi Berbasis Hots Menggunakan Aplikasi Quizizz Pada Mata Pelajaran Mekanika Teknik Dalam Pembelajaran Jarak Jauh. *Jurnal Pendidikan Teknik Sipil*, *3*(2), 169–180. https://doi.org/10.21831/jpts.v3i2.43919

Nelawati, Wahyu, R., Putra, Y., & Simatupang, A. . (2018). PENGEMBANGAN BAHAN AJAR INTEGRASI NASIONAL BERBASIS MOBILE. *Jurnal Teknologi Pendidikan*, *20*(3), 407–414.

Norouzkhani, N., Norouzi, S., Faramarzi, M., Bahari, A., Shirvani, J. S., Eslami, S., & Tabesh, H. (2025). Developing and evaluating a gamified self-management application for inflammatory bowel disease using the ADDIE model and Sukr framework. *BMC Medical Informatics and Decision Making*, *25*(1), 1–22. https://doi.org/10.1186/s12911-024-02842-3

Nowicki, G. J., Mazurek, W., Waśkowicz, A., Kowalczyk, E., Kozioł, J., Miłosz, M., Dzieńkowski, M., & Ślusarska, B. (2024). Development and pre-evaluation of a "DiagNurse" mobile app to support nurses in clinical diagnosis using the ADDIE model. *Scientific Reports*, *14*(1), 1–15. https://doi.org/10.1038/s41598-024-81813-0

Nurasiah, S., & Rukli. (2025). EKSPLORASI MODEL EVALUASI LITERASI DIGITAL BERBASIS AI DI SEKOLAH DASAR. *JURNAL ILMU PENDIDIKAN(JIP)*, *6*(1), 145–152. https://doi.org/https://doi.org/10.32696/jip.v6i1.4638

Poerwanti, J. I. S., & Marmoah, S. (2024). Evaluasi pembelajaran daring di pgsd pada era pandemi covid 19 ditinjau dari pelaksanaa asesmen. *Jurnal Pendidikan Dasar*, *12*(1), 66. https://doi.org/10.20961/jpd.v12i1.85686

Pratiwi, J., Kusairi, S., & Hamimi, E. (2021). Analisis kebutuhan pengembangan formative feedback berbantuan komputer dengan butir soal pilihan ganda isomorfik pada materi gelombang siswa SMP kelas VIII. *Jurnal MIPA Dan Pembelajarannya*, *1*(5), 341–346. https://doi.org/10.17977/um067v1i5p341-346

Solihin, I., & Yusuf, F. N. (2024). Online Language Test And Assessment: Indonesian Elementary School Teachers' Perception. *Jurnal Penelitian Pendidikan*, *24*(3), 308–324. https://doi.org/https://doi.org/10.17509/jpp.v24i3.77768

Spatioti, A. G., Kazanidis, I., & Pange, J. (2022). A Comparative Study of the ADDIE Instructional Design Model in Distance Education. *Information (Switzerland)*, *13*(9), 1–20. https://doi.org/10.3390/info13090402

Ulfa Laelatul Fitriah, Radeni Sukma Indra Dewi, & Ratna Ekawati. (2025). Menyelami Kesiapan Guru Kelas Dasar dalam Adopsi Asesmen Diagnostik di Kurikulum Merdeka. *DIAJAR: Jurnal Pendidikan Dan Pembelajaran*, *4*(1), 94–101. https://doi.org/10.54259/diajar.v4i1.3527

Wahyuni, N. G. A. P. E. S., & Haeruddin. (2025). Pengembangan Digital Assessment untuk

Pemahaman Konsep Dasar IPAS di SD Kota Palu. *Jurnal Ilmiah Profesi Guru (JIPG)*, *6*(1), 63–72. https://doi.org/10.30738/jipg.vol6.no1.a18681

Zhai, X., & Nehm, R. H. (2023). AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching*, *60*(6), 1390–1398. https://doi.org/10.1002/tea.21885

Zhang, K. (2025). Enhancing Critical Writing Through AI Feedback: A Randomized Control Study. *Behavioral Sciences*, *15*(5), 1–15. https://doi.org/10.3390/bs15050600

Zhao, J., Li, Y., & Feng, W. (2023). Investigating the Validity and Reliability of a Comprehensive Essay Evaluation Model of Integrating Manual Feedback and Intelligent Assistance. *International Journal of Emerging Technologies in Learning*, *18*(4), 248–261. https://doi.org/10.3991/ijet.v18i04.38241