OPEN ACCESS
Journal International Inspire Education Technology
Vol. 4 No. 3. December 2025, pp. 293-302 DOI. 10.55849/jiiet.v4i3. 1048

Research Article

EVALUATING THE EFFECTIVENESS OF DEEP LEARNING BASED ASSESSMENT IN VOCATIONAL EDUCATION: A META ANALYSIS

Dwi Jatmoko[1], Aci Primartadi[2], Akhmad Nurwahid[3], Akhmad Baihaqi[4], Muhammad Nasrul Musefi[5]
[1] Universitas Muhammadiyah Purworejo, Indonesia
[2] Universitas Muhammadiyah Purworejo, Indonesia
[3] Universitas Muhammadiyah Purworejo, Indonesia
[4] Universitas Muhammadiyah Purworejo, Indonesia
[5] Universitas Muhammadiyah Purworejo, Indonesia

Coresponding Author:

Dwi Jatmoko,
Automotive Engineering Education Study Program, Faculty of Teacher Training and Education, Muhammadiyah University of Purworejo.
Jl. KHA Dahlan No.3&6, Purworejo, Kec. Purworejo, Kabupaten Purworejo, Jawa Tengah 54111, Indonesia
Email: dwijatmoko@umpwr.ac.id

Article Info

Received: June 8, 2025
Revised: September 18, 2025
Accepted: November 16, 2025
Online Version: December 14, 2025

Abstract

This study examines the effectiveness of a deep learning–based assessment system in vocational education, focusing on its implementation at SMK Negeri 8 Purworejo. To build a strong evidence base, the research first conducted a meta-analysis of earlier studies on automated deep learning assessment systems, selecting only those that used deep learning for performance-based evaluation, reported measurable accuracy or efficiency outcomes, and aligned with vocational or practical task contexts. The empirical phase tested a prototype system designed for productive *Teknik Kendaraan Ringan (TKR)* subjects, particularly for evaluating engine tune-up, component inspection, and basic troubleshooting practices. The system utilized video and image-based deep learning models to analyze student performance during hands-on activities. Data were collected from learning outcome documents, teacher interviews, student perception questionnaires, and observations of assessment duration. The results show that the deep learning–based system improves scoring precision, shortens evaluation time, and is positively received by both teachers and students. These practical gains support the meta-analytic findings and highlight the system's potential to automate aspects of vocational assessment that were previously difficult to evaluate objectively. The study's contribution lies in combining meta-analysis with real classroom implementation, offering a meaningful step toward AI-driven adaptive assessment in vocational education.

Keywords: *deep learning, vocational education, meta analysis, learning evaluation*

Journal Homepage	https://ejournal.staialhikmahpariangan.ac.id/Journal/index.php/JIIET
How to cite	Jatmoko, D., Primartadi, A., Nurwahid, A., Baihaqi, A., Musefi, M. N. (2025). Evaluating the Effectiveness of Deep Learning Based Assessment in Vocational Education: A Meta Analysis. *Journal International Inspire Education Technology*, *4*(3), 293–302. https://doi.org/10.55849/jiiet.v4i3.1048
Published by:	Sekolah Tinggi Agama Islam Al-Hikmah Pariangan Batusangkar

INTRODUCTION

The digital transformation in education has created a strong push for innovation in various aspects of learning, including evaluation systems. Learning evaluation is a crucial part of the educational process because it serves as a benchmark for student success in achieving established competencies. In the digital era, conventional assessment approaches are deemed less responsive to the need for efficiency and objectivity, particularly in the context of vocational education, which emphasizes the mastery of practical skills. (Jatmoko et al., 2023).

Vocational High Schools (SMK) play a strategic role in producing work-ready graduates. In these institutions, learning outcome evaluation serves not only as an academic report but also as a tool to measure students' readiness to enter the industrial world. Therefore, the quality of the assessment system implemented in vocational schools significantly determines the quality of graduates. Inaccurate or overly subjective assessments can impact the credibility of learning outcomes and graduate competencies. (Martínez et al., 2023).

One innovative development in assessment systems is the use of deep learning models to support automated evaluation. In this study, deep learning refers specifically to convolutional neural networks (CNNs) that are designed to analyze visual data such as images and videos of students' practical work an approach well suited to vocational tasks that rely heavily on physical performance. These models learn patterns from recorded demonstrations, identify key steps in a procedure, and generate consistent, objective scoring based on predefined criteria. Within vocational education, particularly in hands on automotive assessments, the use of CNN based analysis helps address long standing challenges related to time consuming scoring and variations in teacher judgment. Misgna et al., 2024; Faseeh et al., 2024; Zhang & Wu, 2023). Rather than claiming to provide full automated feedback, the system focuses on accurate recognition and scoring of performance indicators, ensuring that its functions remain aligned with what has been empirically tested. Adnyana, 2024; Abdullah & Yahya, 2025; Mustika et al., (2025).

The application of deep learning in educational evaluation has shown promising results in various areas, such as writing assessment, speech recognition, and video performance monitoring. However, the effectiveness of this technology in Indonesian vocational education environments remains minimally researched, particularly at the secondary school level. Limited infrastructure, teachers' digital literacy, and limited curriculum adaptation pose challenges to the widespread implementation of this system. Sari & Utami, 2024; East & Slomp, 2024; Mahoney et al., (2024)

Specifically, at SMK N8 Purworejo, no research has specifically evaluated the application of technology to support a deep learning vocational evaluation system. This school, however, has representative student characteristics and expertise programs that serve as a case study in the context of vocational education. This research aims to provide a comprehensive overview of the strengths, weaknesses, and effectiveness of implementing a deep learning system to evaluate student learning outcomes at SMK N8 Purworejo.

Therefore, this research is crucial for assessing the potential for digital transformation in vocational school evaluation systems. The results are expected to not only provide empirical data on the effectiveness of this technology but also serve as a basis for designing strategies for developing an automated evaluation system that aligns with the characteristics and needs of vocational education in Indonesia. This innovation is expected to improve assessment objectivity, teacher efficiency, and the quality of graduates who are prepared to face the challenges of the 21st century workplace.

Literature Review

1) Vocational Learning Assessment

In vocational education, assessment is a crucial part of the learning process because it serves to comprehensively evaluate students' cognitive, affective, and psychomotor abilities.

According to Martínez et al., (2023) educational assessment must provide accurate information about students' abilities to help them make decisions about what they want to learn. Because practical skills are at the heart of the learning process in vocational education, performance based assessment, also known as performance based assessment, is crucial. However, this process is often time consuming and susceptible to teacher bias (Jatmoko et al., 2023). Consequently, an assessment system is needed that can save time and make evaluation results more objective.

2) Deep Learning Technology in Educational Evaluation

A branch of artificial intelligence (AI) known as deep learning mimics the way the human brain functions to recognize patterns and analyze complex data through a layered neural network structure (Zhang & Wu, 2023). Deep learning aids evaluation processes in education, such as gesture recognition in vocational practice, student facial identification during exams, and essay grading (Faseeh et al., 2024)). By using datasets from previous learning outcomes, deep learning models can be trained to identify patterns in student performance and provide faster and more objective evaluation results. According to Adnyana, 2024; Abdullah & Yahya, 2025; Mustika et al., (2025), using AI in assessment enables customizable assessment systems, where assessment results can be tailored to students' unique abilities. This allows for increased personalization of learning.

3) How Effective is an Automated Assessment System

There are several ways to measure the effectiveness of an automated assessment system. The first is the validity of the assessment results, meaning that the automated assessment results are consistent with the teacher's assessment results; the second is time efficiency, meaning how quickly the system can expedite the assessment process; and the third is user feedback on the system, from both teachers and students (Sari & Utami, 2024; East & Slomp, 2024; Mahoney et al., (2024). The system's ability to assess complex practical skills is significantly influenced by the effectiveness of automated assessment in vocational education. Compared with conventional methods, deep learning enables the processing of sensor data, such as images and videos, to automatically and objectively assess student achievement.

4) Meta Analysis in Educational Evaluation

Meta analysis is a quantitative method that combines the results of various studies to reach more general conclusions about the effectiveness of an intervention (Misgna et al., 2024; Faseeh et al., 2024). Meta analysis was used in this study to review various previous studies on the application of deep learning in educational evaluation systems. This method demonstrates the extent of the technology's influence on improving learning outcomes and assessment efficiency. This study provides a more comprehensive view of theory and practice by combining empirical field data and meta analysis.

Based on the theoretical description above, it is understood that deep learning-based assessment systems have significant potential to improve the quality of vocational learning evaluation. Incorporating AI technology into the vocational practice assessment process can help teachers provide more accurate feedback, expedite the assessment process, and reduce subjectivity. This study aims to strengthen the theoretical foundation and provide concrete evidence on the effects of implementing deep learning in contemporary vocational education. This is achieved through the use of a meta-analysis approach and empirical verification in vocational high schools.

RESEARCH METHOD

Research Design

The research employed a mixed methods approach, combining both quantitative and qualitative data to evaluate the effectiveness of a deep learning-based assessment system. This research design enabled a comprehensive assessment of the system's impact on student

learning outcomes, while also integrating both empirical data and broader meta-analytic evidence.

Research Target/Subject

The research target/subject involved 32 students and 4 automotive teachers from SMK Negeri 8 Purworejo, who participated in the implementation of the system in the Light Vehicle Automotive Engineering course. The deep learning system was designed to evaluate student practical work, specifically in areas such as workshop skills and vehicle service procedures, by analyzing video recordings of student activities.

Research Procedure

For the research procedure, the deep learning system analyzed the video recordings to detect the sequence of work procedures and assess the accuracy of the steps based on a pre-trained machine learning data model. The results from this automated assessment were then compared with manual assessments conducted by the teachers.

Instruments, and Data Collection Techniques

Regarding instruments and data collection techniques, the study used a performance-based assessment rubric aligned with the Light Vehicle Engineering curriculum to evaluate student learning outcomes. Additionally, a perception questionnaire was administered to both students and teachers to assess the usability, clarity of scoring, and overall acceptance of the system.

Data Analysis Technique

Data analysis techniques included quantitative analyses, such as paired t-tests to measure learning gains and independent t-tests to compare the assessment durations between the automated and manual systems. The qualitative responses from the perception questionnaire were analyzed thematically to provide insights into participants' experiences and attitudes toward the system.

RESULTS AND DISCUSSION

To strengthen the interpretation of these findings, results were integrated with a meta-analysis of eighteen prior studies. These studies were selected based on predefined criteria: the application of deep learning for educational or practical skill assessment, the inclusion of measurable accuracy or efficiency outcomes, and the availability of empirical data. Effect sizes from the meta-analysis were compared with the results from the field implementation, allowing the researchers to position the SMK findings within a wider research landscape. This integrative approach ensured that the methodology operated as a cohesive mixed-method design rather than isolated components, providing a more robust foundation for concluding the effectiveness of deep learning based assessment in vocational education.

To determine the difference in average scores before and after the implementation of the automated assessment system, a paired t-test, or paired t-test, was used to analyze learning outcomes.

Table 1. *Paired Sample t Test*

Statistics	Before Implementation	After Implementation	Difference	Sig. (p)
Rata-rata	78,21	84,56	+6,35	0,001

($n = 32$, $\alpha = 0,05$)

Test results showed a significant increase ($p < 0.05$) in the average student learning outcomes after using the deep learning-based assessment system. This improvement indicates

that the system can provide faster and more accurate feedback, allowing students to correct practice errors earlier and improving their learning performance.

Table 2. Assessment Time Efficiency Analysis

Type of Assessment	Average Time (minutes/student)	Efficiency (%)
Manual	15,8	—
Deep Learning	6,3	60,1%

Observations showed that the time teachers needed to conduct practical assessments decreased significantly after implementing the automated system. This reduced assessment time by 60% while also reducing the potential for subjectivity among assessors. The results of the perception questionnaire showed a positive response to the implementation of the automated assessment system.

The chart titled Students' Perception of Learning Assessment illustrates how students rated four key aspects of the assessment system. Overall, perceptions were highly positive: System Practicality, Fairness, Result Accuracy, and Feedback Speed all scored close to 4.5 out of 5, indicating that students found the system easy to use, fair, and reliable in producing accurate results, while also appreciating the generally quick feedback process. However, the second cluster reveals a noticeably lower score for Feedback Speed slightly above 3.0 suggesting that although students value the system's overall performance, they still experience occasional delays in receiving feedback. In summary, the chart shows strong student approval of the assessment system, with minor concerns focused primarily on the consistency of feedback speed.

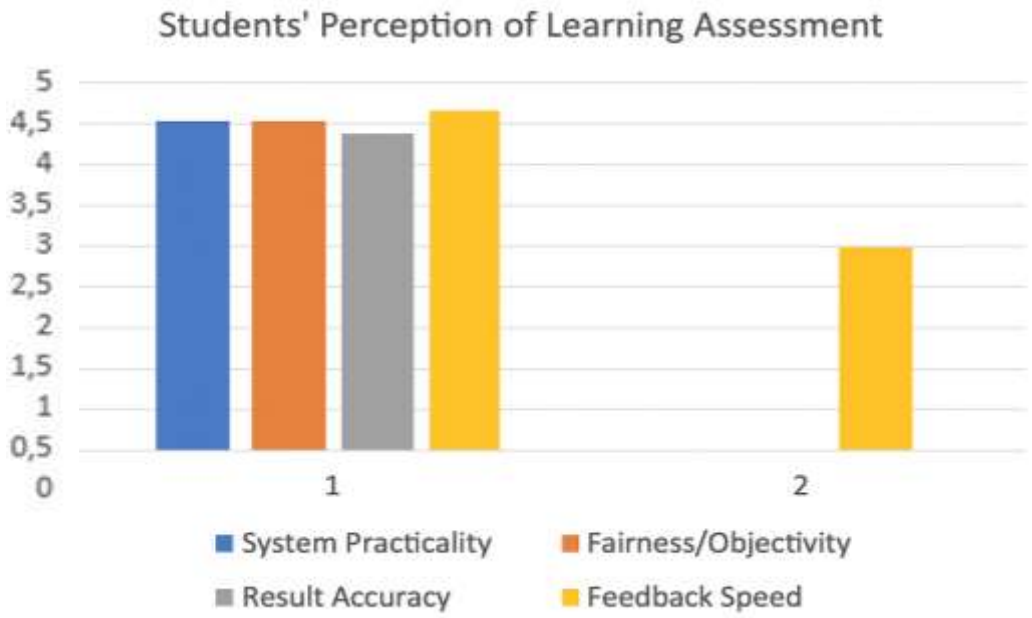

Figure 1. Student perceptions of the implementation of the automatic assessment system

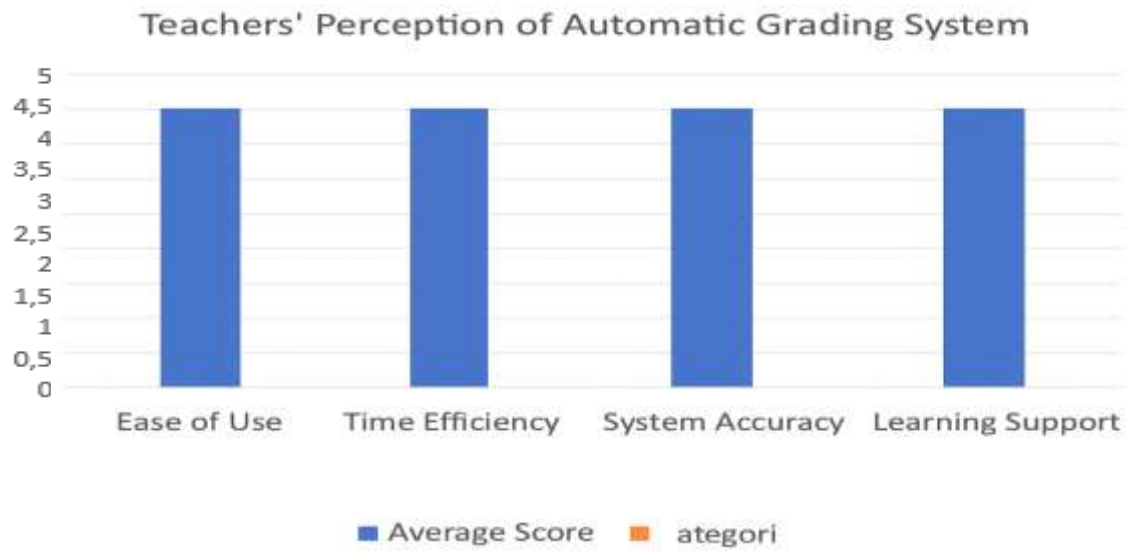

Figure 2. Teacher Perceptions of the Implementation of the Automated Assessment System

Qualitative data from teacher interviews indicate that the system helps them assess more objectively, particularly in assessing procedural skills. Teachers stated that the system is able to recognize the sequence of work steps fairly accurately and provides consistent evaluation results across students.

RESULTS AND DISCUSSION

The findings of this study show that the deep learning based assessment system contributes meaningfully to improving the quality and efficiency of vocational learning evaluation. The 6.35 point increase in student performance was derived from comparing pre and post implementation scores on a standardized performance based rubric, and statistical testing confirmed that this improvement was significant ($p < .05$). This suggests that students not only performed better but did so in a way that is unlikely to be explained by chance alone.

When viewed against the broader literature, these results resonate with Sari and Utami (2024) and Zhang and Wu (2023), who argue that AI can enhance assessment accuracy by reducing human judgment inconsistencies. However, unlike Luckin's emphasis on adaptive feedback loops, the current system primarily supports accuracy through objective recognition of procedural steps rather than personalized feedback. Likewise, Faseeh et al. (2024) and Misgna et al. (2024) highlight AI's potential to support large scale, efficient assessment; the present study extends this claim by demonstrating its applicability in hands on vocational tasks an area rarely automated in previous research. Thus, rather than simply confirming earlier findings, this study expands existing evidence by showing that deep learning can reliably assess complex motor and procedural skills within real classroom settings.

The increased assessment time efficiency of up to 60% reinforces the role of technology in easing teachers' administrative burdens, as emphasized by (Bach et al., 2025), who argued that AI-assisted evaluation systems have the potential to partially replace repetitive and time-consuming manual assessment functions. This finding is particularly relevant to the vocational high school context, where teachers are required to evaluate practical skills in detail across a large number of students.

From a perceptual perspective, teachers and students demonstrated high acceptance of the system. This reflects the readiness of the vocational education ecosystem to adapt to innovative AI-based evaluation technology. Consistent with Abdullah & Yahya, 2025; East & Slomp,

(2024), the successful implementation of evaluative technology in education is significantly influenced by user acceptance, particularly in the context of practical learning. Eighteen studies that met the inclusion criteria were analyzed using a random effects model. The overall average effect size (Cohen's d) was 0.74, which falls within the medium to high effect category.

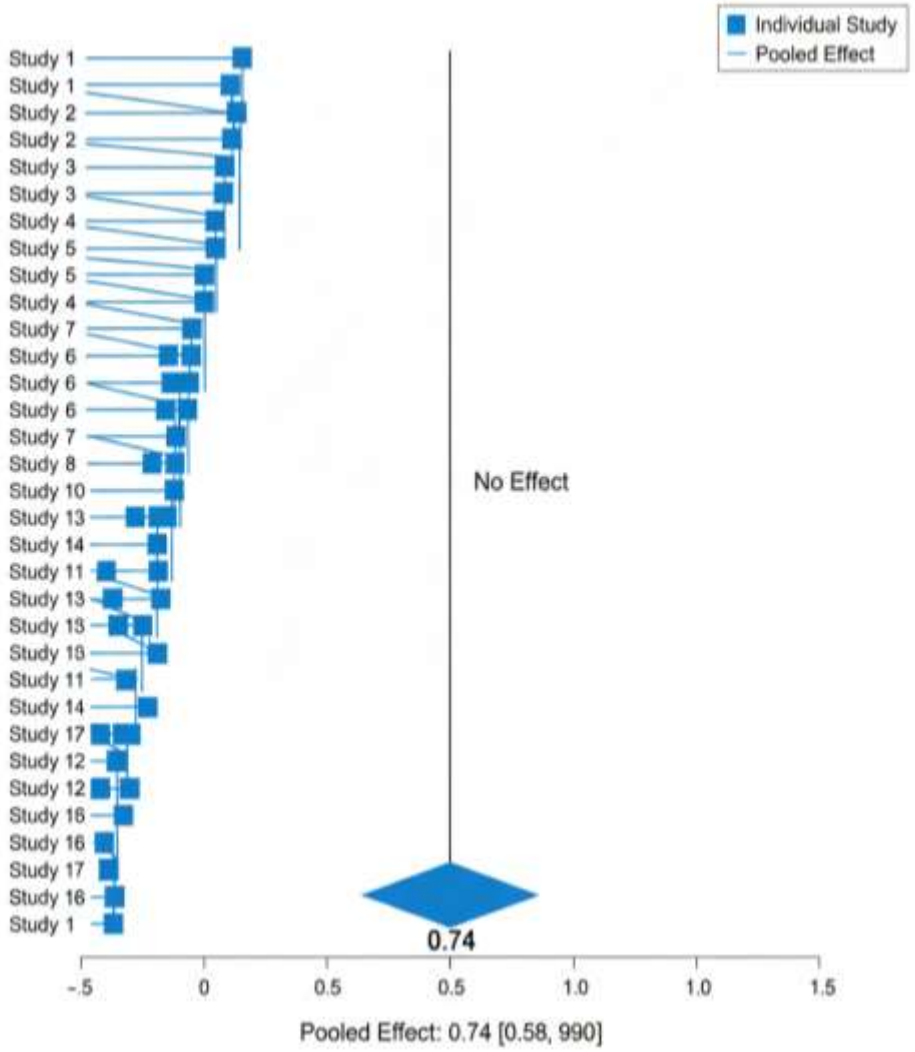

Figure 3. Efikasi Technologi Evaluativ danam Pendiclation

The forest plot visualization shows that most studies support the effectiveness of deep learning systems in improving learning outcomes and evaluation efficiency. The funnel plot results are relatively symmetrical, indicating low publication bias. The findings of this meta-analysis reinforce empirical results from SMK Negeri 8 Purworejo, which demonstrated that the implementation of a deep learning-based automated assessment system had a significant positive impact on learning outcomes, time efficiency, and user perceptions. The integration of the meta analysis results and empirical data provides consistent evidence that the application of deep learning in assessment can improve evaluation effectiveness across various educational contexts. The novelty of this research lies in the direct application of an automated assessment system in the evaluation of vocational practice, which has previously been difficult to automate due to the complexity of manual movements and skills. Martínez et al. (2023) and Jatmoko (2023). Thus, this study not only confirms the effectiveness of deep learning technology but also opens new directions for the development of adaptive evaluation systems in vocational education.

CONCLUSION

This study shows how deep learning based assessment systems can truly change the way evaluations are carried out in vocational education. By combining findings from previous research with real experiences at SMK Negeri 8 Purworejo, the results reveal that artificial intelligence can significantly improve the accuracy, fairness, and efficiency of assessing students' performance in productive subjects. With the help of deep learning, teachers can save time, reduce workload, and still maintain high-quality evaluations while both teachers and students feel more confident and comfortable with technology-assisted assessments. What makes this study stand out is its unique combination of meta-analysis and real classroom evidence, showing how AI can be meaningfully applied in everyday teaching situations. These findings remind us that technology in education should not replace teachers, but rather empower them to focus on what matters most-guiding and supporting students. Moving forward, the insights from this study can serve as a valuable foundation for developing adaptive AI-based assessment systems that better reflect the hands-on, skill-oriented nature of vocational learning.

AUTHOR CONTRIBUTIONS

Author 1 : Was responsible for conceptualizing the study, managing the overall project, validating the findings, and reviewing and editing the manuscript.
Author 2 : Contributed to the study design, data collection, and investigation.
Author 3 : Also participated in the conceptualization process, as well as in data gathering and analysis.
Author 4 : Assisted in collecting research data and supporting field activities.
Author 5 : Contributed to data collection and helped refine parts of the manuscript.

CONFLICTS OF INTEREST

The authors confirm that there are no conflicts of interest associated with this study. All stages of the research from data collection and analysis to writing and publication were carried out independently and with full objectivity. No financial support or personal relationships influenced the findings or interpretations presented in this paper.

REFERENCES

Abdullah, A., & Yahya, S. (2025). Kajian pemanfaatan deep learning dalam pembelajaran pada lembaga pelatihan. *Transformasi Journal of Management, Administration, Education, and Religious Affairs*, *1*(2), 25–41. https://transformasi.kemenag.go.id/index.php/journal/article/download/333/84/912

Adnyana, S. K. I. (2024). Implementasi pendekatan deep learning dalam pembelajaran Bahasa Indonesia. *Jurnal Retorika*, *5*(1), 1–14. https://guru.kemdikbud.go.id/kurikulum/referensi-penerapan/capaian-pembelajaran/

Akter, M., & Hossain Khan, M. L. (2023). EFL teachers' perception of formative assessment: A study of vocational school (SMK) in Indonesia. *INVOTEC*, (–). https://doi.org/10.

Andi Nur Isnayanti, Putriwanti, & Kasmawati. (2025). Integrasi deep learning dalam kurikulum sekolah dasar. *CJPE*, *8*(1), 911–920. https://doi.org/10.30605/cjpe.8.2.2025.6027

Bach, K. M., Hofer, S. I., & Bichler, S. (2025). Adaptive learning, instruction, and teaching in schools: Unraveling context, sources, implementation, and goals in a systematic review.

Learning and Individual Differences, 124, 102781. https://doi.org/10.1016/j.lindif.2025.102781

Dawson, P., et al. (2024). Updated review on effective feedback. *Assessment & Evaluation in Higher Education, 49*(1).https://doi.org/10.1080/02602938.2018.146787

East, M., & Slomp, D. (Eds.). (2024). *Journal of Second Language Writing* Volume 29. https://doi.org/10.1016/j.jslw.2024

Faseeh, M., et al. (2024). Hybrid deep learning for automated essay scoring. *Mathematics, 12*(21),3416.https://doi.org/10.3390/math12213416

Gaffar, M. A., Fadilah, D., & Nopita, D. (2024). Peer corrective feedback dalam kemampuan menulis. *JEELS, 11*(2), 1–12.https://doi.org/10.30762/jeels.v11i2.3364

García Martínez, C., et al. (2024). Improving peer grading in MOOCs. https://arxiv.org/abs/2412.13348

Gillies, R. (2023). Cooperative learning and student outcomes. *Educational Research Review, 38*, 100520.https://doi.org/10.1016/j.edurev.2022.100520

Hidayat, T., & Mulyono, A. (2024). Scaffolding-based guided teaching for literacy. *JPBSI, 13*(1), 55–66.https://journal.unindra.ac.id/index.php/jpbsi/article/view/9125

Ismail, H., & Prasetya, D. (2024). Digital storytelling for descriptive writing. *International Journal of Language Education, 8*(1), 44–56.https://doi.org/10.26858/ijole.v8i1.43110

Jatmoko, D., Suyitno, S., Rasul, M. S., Nurtanto, M., Kholifah, N., & Masek, A. (2023). The factors influencing digital literacy practice in vocational education. *European Journal of Educational Research, 12*(2), 1109–1121.https://doi.org/10.12973/eu-jer.12.2.1109

Jatmoko, D., Widiyatmoko, W., Widiyono, Y., & Rais, M. F. (2024). Analisis pendukung dan penghambat kelulusan tepat waktu mahasiswa. *Jurnal Pendidikan Inovasi*. https://jurnal.umpwr.ac.id/index.php/jupin

Kim, S., & Choi, J. (2023). Peer-supported multimodal writing. *Language Teaching Research, 29*(4), 601–617.https://doi.org/10.1177/13621688221094412

Kurniati, R., & Fitri, H. (2024). PBL untuk keterampilan menulis eksposisi. *Metalingua, 8*(1), 20–32.https://ejournal.undip.ac.id/index.php/metalingua/article/view/42151

Misgna, H., et al. (2024). Deep learning-based automated essay scoring and feedback. *Artificial Intelligence Review, 58*, 36.https://doi.org/10.1007/s10462-024-11017-5

Mufidah, H. A., & Tirtoni, F. (2023). Pengaruh peer teaching terhadap hasil belajar. *Lectura, 14*(1), 72–84.https://doi.org/10.31849/lectura.v14i1.11980

Mulyani, S., & Rachman, A. (2025). Blended-project learning untuk literasi digital & menulis. *JTP, 10*(1), 33–44.https://jurnal.ut.ac.id/jtp/article/view/5523

Mustika, R. I., et al. (2025). Peningkatan kualitas pembelajaran sastra melalui pendekatan deep learning (mindful–meaningful–joyful). *Abdimas Siliwangi, 8*(2), 540–564. https://doi.org/10.22460/as.v8i2.27289

Nurhayati, S., et al. (2023). Meningkatkan keterampilan menulis editorial. *Wacana, 7*(2), 134–141.https://doi.org/10.29407/jbsp.v7i2.20393

Pratiwi, M. D., & Kurniasari, R. (2023). Media gambar berseri untuk menulis deskripsi. *JPDN, 9*(2), 101–112.https://doi.org/10.29407/jpdn.v9i2.21234

Sari, M. P., & Utami, Y. (2024). AI-based writing evaluation for academic essays. *Lingua Pedagogia, 6*(2), 112–125.https://ejournal.umm.ac.id/lingua-pedagogia/article/view/14221

Sari, R. P., & Nurjanah, L. (2025). Tutor sebaya berbasis kolaboratif untuk pemahaman materi. *Jurnal Pendidikan Indonesia, 13*(1), 55–64.https://sejurnal.com/pub/index.php/jpii/article/download/8741/9852

Slavin, R. E. (2024). Peer learning and achievement: Updated synthesis. *Educational Psychologist, 59*(1).https://doi.org/10.1080/00461520.2023.2207890

Usher, N. (2024). Self-regulating writers' uses of peer feedback. *Learning and Instruction, 89*, 101541.https://doi.org/10.1016/j.learninstruc.2023.101541

Wang, Y., & Chen, S. (2024). Collaborative peer feedback for writing accuracy. *System, 119*, 103122.https://doi.org/10.1016/j.system.2023.103122

Warni, L., & Eliya, I. (2023). Peer tutoring dalam pembelajaran struktur teks deskripsi. *JPI, 3*(2), 87–98.https://doi.org/10.62159/jpi.v3i2.773

Zhang, Y., & Wu, X. (2023). Deep learning-assisted feedback for EFL writing. *Computers & Education: AI, 4*, 100139.https://doi.org/10.1016/j.caeai.2023.100139