

Repositioning AI in ELT: Supporting teachers in transforming English language evaluation

Fitria Lapele^{1*}, Abel Yohanis Romrome², Eko Wahyunanto Prihono³

¹English Tadris Department, Faculty of Tarbiyah, A.M. Sangadji State Islamic University, Ambon, Indonesia; email: fitrialapele@iainambon.ac.id

²English Education Department, Wamena Christian Teacher Training and Education College, Papua, Indonesia; email: giovannymoreira10@gmail.com

³Master of Mathematics Tadris Department, Postgraduate Program, A.M. Sangadji State Islamic University, Ambon, Indonesia; email: ekoprihono@iainambon.ac.id

ABSTRACT

The rapid advancement of artificial intelligence (AI) has significantly transformed pedagogical and evaluative practices in higher education, yet its role in English language evaluation remains relatively underexplored. Previous studies predominantly examine general classroom applications or student perspectives, providing limited insight into how English teachers employ AI within evaluation practices. This study addresses this gap by investigating teachers' perceptions of AI as a teaching assistant in English language evaluation. Employing a qualitative design, data were gathered through semi-structured interviews with seven English teachers, which were then analyzed thematically. The findings reveal four interrelated dimensions of AI-supported evaluation: (1) efficiency in evaluation design, (2) empowerment in formative and reflective practices, (3) ethical and valid use of AI, and (4) institutional readiness for sustainable integration. The study contributes new empirical evidence by showing how teachers actively negotiate the pedagogical, ethical, and policy-related challenges of integrating AI, an aspect often overlooked in earlier research. These insights highlight that AI functions most effectively as a collaborative partner that augments, rather than replaces, human evaluative judgment. The study offers both theoretical implications for AI-mediated evaluation literature and practical guidance for institutions seeking responsible and sustainable AI integration.

Keywords: AI integration; English language evaluation; ethical evaluation; formative feedback; institutional readiness; teacher's assistant

*Corresponding Author:
fitrialapele@iainambon.ac.id

Article history:

Received: 28 October 2025

Revised: 28 November 2025

Accepted: 23 December 2025

Published: 28 February 2026

Citation: Lapele, F., Romrome, A.Y., & Prihono, E.W. (2026). Repositioning AI in ELT: Supporting teachers in transforming English language evaluation. *English Learning Innovation*, 7(1). 291-303. <https://doi.org/10.22219/englie.v7i1.42623>

Copyright © 2026: Lapele et al.,
This is an open access article
under the CC-BY-SA license



P-ISSN 2723-7400

E-ISSN 2723-7419

INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) has significantly transformed higher education worldwide, altering how teaching, learning, and evaluation are conceived and implemented. In recent years, the integration of AI tools such as ChatGPT, Grammarly, and QuillBot has become increasingly common in English language education, particularly for supporting evaluation design and feedback processes (Abdullah, 2025; Al-khresheh, 2024; Pratama & Hastuti, 2024). Scholars assert that AI's true potential lies in automating tasks and enhancing educators' pedagogical capabilities to innovate, reflect, and engage learners more effectively (Alam & Mohanty, 2023; Cotos et al., 2020). However, these developments also raise critical questions about ethics, transparency, and the continued necessity of human judgment in assessment contexts (Kilinc & Mansiz, 2024; Radanliev, 2025).

Recent research in applied linguistics and educational technology highlights the dual role of AI in language education, acting as both a catalyst for pedagogical creativity and a source of ethical and procedural tension. Studies have shown that while AI can increase efficiency in task preparation and feedback generation, its integration into assessment remains shaped by issues of authenticity, teacher interpretive authority, and institutional capacity (Ahmed et al., 2025; Perkins et al., 2024; Radovan et al., 2025). Research also indicates substantial variation in institutional readiness, with disparities in training support, policy clarity, and digital infrastructure influencing how responsibly AI can be used in evaluative decision-making (Chan, 2023). Taken together, this body of work suggests that AI's pedagogical promise is tempered by ethical, procedural, and contextual constraints.

In Indonesia, the adoption of AI in higher education has gained momentum, fueled by the government's digital transformation agenda and the increasing integration of generative AI in academic settings. English teachers are increasingly exploring AI for rubric development, item generation, and formative feedback, particularly through generative and feedback-oriented tools (Kanchana & Saha, 2025; Yavuz et al., 2025). However, empirical studies in this area are still limited, and existing research often focuses on technical or policy aspects rather than teachers' lived pedagogical experiences in using AI for evaluation (Al-kfairy, 2024; Asad et al., 2021). Additionally, concerns about academic integrity, fairness, and the potential risks of overreliance on AI-driven evaluations remain prevalent in institutional conversations (Kim, 2024; Yan & Liu, 2025). These patterns indicate a growing need for evidence-based insights into how educators navigate AI within the practical, ethical, and institutional realities of evaluation work.

Despite the increasing volume of literature on AI in education, several limitations remain. Existing studies frequently prioritize student perspectives, technical performance, or classroom-based applications rather than teachers' evaluative practices (Aini et al., 2024; Filiz et al., 2025). Moreover, few studies examine how teachers negotiate institutional constraints or ethical considerations when integrating AI into evaluation (Tubella et al., 2024; Schiff, 2022). Consequently, a clear gap persists in understanding how English teachers balance efficiency with ethical judgment, and how they maintain evaluative integrity while adopting new technologies in higher education contexts. This gap is important because AI-supported evaluation remains understudied compared to its broader use in teaching and learning.

This study aims to address the existing gap by examining English teachers' perceptions and experiences regarding the use of AI as a teaching assistant in language evaluation practices. Specifically, it investigates how teachers view AI's contributions to efficiency, formative feedback, ethical judgment, and institutional readiness. By positioning AI not as a replacement but as a pedagogical collaborator, this research enhances our understanding of AI-assisted evaluation in higher education. On a theoretical level, it contributes to the discourse on human–AI collaboration in pedagogy and educational ethics. Practically, it provides guidance for institutions and educators striving to balance technological innovation with pedagogical integrity and ethical accountability. Building on these aims, this study is guided by the following research questions:

1. How do English teachers perceive the role of AI as a teaching assistant in English language evaluation practices?
2. How do teachers describe AI's contributions to efficiency, formative feedback, ethical judgment, and institutional readiness in their evaluative work?

METHODS

This study employed a qualitative case study design because it examined a bounded system of English teachers' AI-assisted evaluation practices within a specific higher-education context (Creswell & Creswell, 2017; Yin, 2018). In this research, the "case" refers to the shared evaluative practices and decision-making processes of English teachers who incorporated AI tools into evaluation tasks in Indonesian universities, rather than individual narratives or life-world experiences. A case study approach was chosen over phenomenology or narrative inquiry because the aim was to understand how a group of practitioners collectively navigated pedagogical, ethical, and institutional realities when integrating AI into evaluation, making this design more suitable for capturing contextual dynamics.

Purposive sampling was used to recruit seven English teachers who met the following inclusion criteria: (a) at least three years of teaching experience in higher education; (b) demonstrable use of AI tools specifically for evaluation-related activities; and (c) involvement in designing or conducting evaluative practices in English language learning. These criteria ensured that participants were genuinely engaged in AI-supported evaluation rather than general digital tool use. The sample size of seven teachers was deemed adequate because thematic saturation was reached during the sixth interview, with no new substantive insights emerging in the seventh; and while not statistically representative, the sample reflects varied academic ranks and levels of AI usage.

Data were collected through semi-structured interviews consisting of 10 core questions and supplementary probes exploring teachers' experiences with AI in evaluation, perceived benefits and constraints, and institutional support mechanisms. The interviews were conducted online via Zoom between April and June 2025 due to geographical constraints, lasted 45–60 minutes, and were audio-recorded with participants' consent before being transcribed verbatim. Written informed consent was obtained from all participants, who were assured of anonymity, confidentiality, and their right to withdraw at any stage of the study.

The data were analyzed using Braun and Clarke's (2021) six-phase thematic analysis

framework, which included familiarization, generating codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. The analysis followed a primarily inductive approach so that themes emerged from the data, while the evaluative framework was applied only in later interpretive stages to avoid imposing pre-existing categories during initial coding. Coding was conducted manually, and emerging interpretations were discussed among the research team to enhance credibility.

RESULTS

The thematic analysis of interview data from seven English lecturers generated four themes describing their experiences with using AI in English language evaluation. These themes present participants' perspectives on how AI supports their evaluative work. The themes are: (1) Efficiency in Language Evaluation Design, (2) Empowerment in Formative Feedback Practices, (3) Ethical Judgment in AI-Supported Evaluation, and (4) Institutional Readiness and Policy Influence.

Efficiency in language evaluation design

Teachers reported that AI tools contributed to increased efficiency in the preparation of English language evaluation tasks. AI applications such as ChatGPT, Grammarly, and QuillBot were described as supporting the generation of prompts, the drafting of rubric descriptors, and the development of sample materials for assessing students' language performance. Through these functions, AI was perceived as reducing the time required for routine preparatory work and enabling teachers to concentrate more on evaluative judgment. This efficiency was particularly evident during the initial stages of task and rubric development, as reflected in the following accounts:

"AI helps me generate examples for writing or grammar evaluations almost instantly. Instead of spending a lot of time creating sample texts, I can focus more on reviewing students' actual language use and performance" (L3).

"Sometimes I use AI to draft or rephrase rubric descriptors, especially when I want clearer distinctions between performance levels. The suggestions give me new perspectives on how to describe assessment criteria" (L6).

These statements indicate that AI functions as a preparatory support tool rather than an autonomous evaluator. Teachers used AI-generated materials as preliminary drafts, which were subsequently refined to align with course objectives and students' proficiency levels. In addition to improving efficiency, teachers noted that AI broadened their evaluative options by offering alternative wording and task variations:

"Sometimes AI gives me different task ideas that make my evaluations more varied. I do not use them directly, but they help me rethink how to design evaluation tasks" (L2).

"AI helps me think of new prompts, especially for writing tasks, which I can later adjust to fit my class context" (L5).

Taken together, these accounts indicate that AI was primarily used to enhance efficiency and flexibility in evaluation design. Teachers described AI as reducing routine preparatory workload and supporting the generation of diverse tasks and

rubric formulations, while decisions regarding contextual adaptation and evaluative alignment remained under teacher control.

Empowerment in formative feedback practices

Teachers reported that AI tools supported their ability to provide more efficient and timely formative feedback. Applications such as ChatGPT and Grammarly were described as assisting in generating initial feedback drafts, identifying common language issues, and offering general improvement suggestions that teachers could later refine for individual learners. Through these functions, AI was perceived as facilitating faster feedback cycles while maintaining teacher involvement in evaluative decisions. The role of AI in accelerating feedback delivery was reflected in the following accounts:

“AI helps me provide quicker feedback on writing tasks. Students can revise their work immediately, while I focus more on qualitative comments related to content and organization” (L4).

“I often use AI to outline initial feedback or rubric language, but I always revise it to fit each student’s context and learning trajectory” (L1).

These responses suggest that AI-generated feedback functions as an initial scaffold rather than a final evaluative judgment. Teachers retained responsibility for adapting feedback to students’ individual needs and developmental stages. Teachers also reported that AI contributed to greater consistency in feedback provision and reduced the time spent on repetitive corrections:

“AI can highlight recurring mistakes that I might miss when checking many papers, especially when the class size is large” (L5).

“Using AI saves time on basic corrections, so I can pay more attention to students’ specific weaknesses and progress” (L3).

Collectively, the findings show that AI supported formative feedback practices by accelerating feedback delivery and improving consistency across student work. At the same time, teachers maintained an active role in interpreting and personalizing feedback to ensure its relevance to individual learning needs.

Ethical judgment in AI-supported evaluation

Teachers consistently emphasized the importance of ethical judgment when incorporating AI into language evaluation. They described AI-generated comments and suggested scores as useful initial references; however, such outputs were not considered sufficient to represent students’ actual performance without careful human review. Teachers highlighted the need to mediate AI assistance through professional judgment to ensure that evaluation remains accurate and meaningful. This ethical consideration was reflected in teachers’ accounts regarding the limitations of relying solely on AI-generated evaluative outputs:

“AI can suggest scores or comments, but I never rely on them fully. Evaluation must reflect students’ real effort and growth, not just what the system generates” (L7).

“If teachers depend too much on AI feedback, we might lose our awareness of students’ learning struggles and developmental processes” (L3).

These responses indicate that teachers perceived AI outputs as partial

representations of student learning. Human interpretation was viewed as essential for capturing developmental progress and learning effort that may not be evident in automated feedback. Teachers also raised concerns related to fairness and the risk of over-reliance on automated tools, particularly in relation to the contextual and creative dimensions of language use:

“AI sometimes gives comments that are too general, so I still need to check the student’s intention and the context of the writing” (L6).

“AI may miss certain nuances in students’ writing, especially when they use expressions that do not fit standard patterns” (L2).

These findings highlight that ethical judgment remained central in AI-supported evaluation practices. Teachers positioned AI outputs as partial and provisional inputs, emphasizing the necessity of professional interpretation to ensure fairness, contextual sensitivity, and alignment with students’ learning development.

Institutional readiness and policy influence

Teachers reported that institutional policies and support systems strongly influenced how AI was integrated into language evaluation practices. While AI tools were increasingly used for evaluative purposes, participants noted that institutional guidelines regarding their appropriate application were not always clearly articulated. This lack of clarity was described as affecting teachers’ confidence and consistency when using AI in assessment-related activities. The influence of institutional policy on evaluative practice was reflected in the following accounts:

“Our university permits the use of AI for feedback and rubric design, but we still need clear standards regarding how far AI can be involved in grading decisions,” (L2).

“Institutional workshops are invaluable in helping us understand how to integrate AI into evaluation without compromising fairness and validity” (L5).

These statements indicate that while institutional permission enables AI use, the absence of detailed evaluative standards may create uncertainty. Conversely, professional development initiatives were perceived as supporting more informed and responsible integration of AI. Teachers also emphasized the importance of training and technical support in shaping their ability to use AI confidently and consistently:

“Workshops help us understand what we can and cannot do with AI when evaluating students. This guidance makes the process clearer” (L4).

“Without clear policies, each teacher might use AI differently, which can lead to inconsistencies in evaluation” (L1).

These responses suggest that institutional readiness extends beyond access to technology, encompassing policy clarity, training opportunities, and ongoing support. Overall, teachers described institutional readiness as a key factor in enabling responsible, ethical, and consistent use of AI in language evaluation.

Cross-theme analysis

A cross-theme reading of the four findings shows that teachers used AI in ways that were shaped by both individual pedagogical decisions and institutional conditions. Across themes, teachers consistently viewed AI as a practical tool that helps with routine evaluative tasks, such as generating prompts or drafting feedback, while still requiring human judgment to ensure fairness, contextual accuracy, and alignment with learning goals. Teachers also emphasized that their ability to integrate AI responsibly depended on institutional clarity, especially regarding policy, training, and ethical guidelines.

These patterns suggest that AI-supported language evaluation is influenced by the interaction between teachers' evaluative practices and the broader institutional environment. AI tools assist with procedural tasks, but teachers remain responsible for interpreting AI-generated suggestions and making final evaluative decisions. Likewise, institutional readiness shapes how confidently and consistently teachers apply AI in their evaluation work.

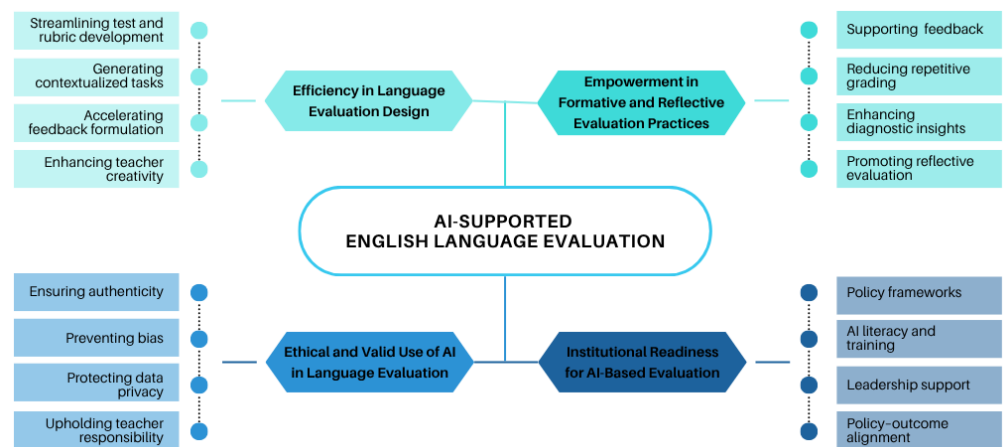


Figure 1. The Structural Dimensions of AI-Supported Language Evaluation

Figure 1 summarizes the four structural dimensions that teachers described when discussing their use of AI in language evaluation. The first dimension reflects how teachers used AI to improve efficiency in designing evaluation tasks. The second dimension shows how AI supported the provision of timely and detailed formative feedback. The third-dimension highlights teachers' emphasis on maintaining ethical judgment and human oversight when interpreting AI-generated suggestions. The fourth dimension represents the influence of institutional policies, training, and support in shaping teachers' confidence and consistency in applying AI in evaluation. Together, these dimensions illustrate the interconnected factors that shape AI-supported evaluation practices in the participants' teaching contexts.

DISCUSSION

The findings of this study underscore that teachers position AI as a supportive assistant that enhances efficiency and feedback quality in English language evaluation, rather than a replacement for human judgment. Participants described using AI to streamline routine tasks, such as generating prompts or drafting initial feedback, while consistently emphasizing their own interpretive authority. This finding supports previous research showing that educators tend to use AI to reduce workload while

maintaining oversight (Burner et al., 2025; Cotos et al., 2020; Kaldaras et al., 2024). At the same time, the findings here extend earlier work by showing that teachers' reliance on AI varies substantially depending on task type, perceived student needs, and individual confidence with AI tools. These nuances highlight that efficiency gains do not automatically translate into evaluative accuracy or appropriateness, an important insight that has received limited attention in prior studies.

In relation to formative feedback practices, the results indicate that teachers use AI primarily to accelerate the drafting of feedback rather than to automate evaluative decisions. This finding aligns with studies showing that AI can facilitate timely and language-specific feedback that supports learner revision (Aini et al., 2024; Winstone et al., 2022). However, the participants in this study expressed stronger concerns about overreliance than those documented in earlier research. Some teachers worried that students may accept AI-generated feedback uncritically or depend on automated suggestions instead of engaging in self-regulation and reflection. This contrasts with more optimistic findings reported in studies that portray AI-based feedback as universally beneficial, suggesting that teacher concerns may be shaped by institutional norms, learner profiles, and teachers' own evaluative philosophies (Chang & Sun, 2024; Perkins et al., 2024).

The findings also reinforce ethical considerations as central to AI-supported evaluation. Teachers consistently emphasized the need for fairness, transparency, and contextual sensitivity principles widely discussed in current scholarship on ethical AI in education (Compagnoni et al., 2025; Nguyen et al., 2023). However, this study extends prior work by demonstrating that teachers' ethical reasoning is deeply intertwined with evaluative judgment rather than treated merely as a compliance requirement. Participants described hesitations about automated scoring, concerns about AI's inability to detect creativity or communicative nuance, and a general insistence on retaining full responsibility for final evaluative decisions. These findings diverge from studies that suggest teachers increasingly trust AI outputs in evaluative contexts (Aljabr & Al-Ahdal, 2024; Lazăr et al., 2024). Instead, the teachers in this study expressed a more cautious and context-sensitive stance, shaped by disciplinary expectations in English language education.

Institutional readiness also emerged as a crucial factor in shaping teachers' practices. In line with prior research, participants reported that well-structured policies and training foster confidence in using AI for evaluative purposes (Alqahtani & Wafula, 2025; Chan, 2023; Filiz et al., 2025). This study extends the literature by showing how inconsistencies in institutional guidance leads to fragmented practices, where teachers independently negotiate what is acceptable or ethical. Such divergence underscores that AI integration is not merely a technical shift but a socio-organizational process requiring leadership, policy clarity, and sustained professional development. Teachers in more supportive environments exhibited stronger evaluative literacy and clearer ethical decision-making, supporting the claim that institutional ecosystems mediate AI adoption (Singun, 2025).

Despite general convergences across the themes, several participants articulated contrasting experiences that complicate and nuance the overall findings. While many teachers perceived AI as a practical resource for generating task ideas, rubric wording, and preliminary feedback structures, others reported that AI-generated outputs were often overly generic or insufficiently aligned with students' actual proficiency levels. In

such cases, teachers indicated that substantial revision was required before AI-generated materials could be pedagogically meaningful, echoing prior findings that highlight limitations in the contextual sensitivity of AI-generated pedagogical content (Chang & Sun, 2024; Luo & Yusuf, 2025). These experiences suggest that efficiency gains associated with AI use are not always immediate and may depend on teachers' evaluative literacy and familiarity with adapting AI outputs to specific learning contexts.

Divergence was also evident in teachers' perceptions of AI-assisted formative feedback. Some participants emphasized the benefits of speed, structure, and consistency, particularly when managing large classes or time constraints. Others, however, expressed concern that reliance on AI-generated feedback could dilute the interpersonal and dialogic dimensions of feedback, which they viewed as essential for supporting learners' motivation, reflection, and self-regulation. Such concerns resonate with previous studies that caution against the uncritical use of automated feedback systems, noting their potential to weaken students' engagement with feedback and limit opportunities for developing autonomous learning strategies (Perkins et al., 2024; Ziqi et al., 2024).

These differing perspectives indicate that AI adoption in language evaluation is neither linear nor uniform. Instead, it involves ongoing negotiation shaped by teachers' levels of trust in AI systems, their pedagogical beliefs, and the specific institutional and classroom contexts in which AI is used. Teachers' judgments regarding when, how, and to what extent AI should be employed reflect a context-dependent process in which technological affordances are continually weighed against pedagogical values and ethical considerations. This variability underscores that AI integration in evaluative practice is best understood not as a standardized intervention, but as a situated and interpretive activity mediated by professional expertise.

Taken together, these findings advance current literature by highlighting several forms of nuance not fully captured in earlier studies. Novel contributions include: (1) a clearer articulation of teachers' ethical reasoning as an inseparable part of evaluative judgment, (2) evidence of intra-group divergence that challenges the assumption of uniformly positive teacher perceptions, and (3) a demonstration of how institutional ecosystems directly shape teacher confidence and evaluative consistency when using AI. This study extends existing scholarship by showing how teachers actively negotiate the ethical tension between efficiency and evaluative integrity, an aspect that remains insufficiently explored in AI-in-ELT research. It also contributes contextual insights from Indonesian higher education, where institutional unevenness creates distinct patterns of AI-supported evaluative practice. These insights enrich the broader discourse on AI in education by showing that evaluative practices are shaped not only by technological affordances but also by interpretive, relational, and institutional forces.

These findings carry several implications for practice. Teachers may benefit from targeted professional development that strengthens evaluative literacy, helps them identify appropriate uses of AI, and supports them in managing student expectations. Institutions should prioritize clear and adaptable policies that outline acceptable AI use in evaluation, ensuring transparency, fairness, and alignment with pedagogical goals. At the system level, these results underscore the need for ongoing dialogue between policymakers, educators, and AI developers to address ethical concerns,

improve alignment with language learning outcomes, and reduce variability across contexts.

This study has limitations. The sample size was small and drawn from a limited number of institutions, which may constrain generalizability. The study relied on self-reported data, which may reflect personal beliefs more than actual classroom practices. Future research should include classroom observations or document analysis to triangulate findings, explore institutional differences more systematically, and examine student responses to AI-supported evaluation.

Taken together, the discussion highlights that the value of AI in language evaluation lies not in technological sophistication alone, but in how teachers interpret, regulate, and ethically position its use within their professional practice. These insights provide a foundation for reconsidering the role of AI as a mediated evaluative partner rather than an autonomous authority.

CONCLUSION

This study shows that teachers perceive AI as a supportive tool that can streamline evaluation tasks and enhance the efficiency of feedback processes, without replacing human judgment in English language evaluation. The key contribution of this research is the identification of how teachers integrate ethical reasoning, practical constraints, and contextual considerations into their decisions about using AI, highlighting that evaluation involves more than technical efficiency. The study also demonstrates that institutional conditions, particularly policy clarity and professional development, shape the consistency and confidence with which teachers adopt AI in evaluative work. Together, these insights suggest that AI's educational value depends not only on its technological affordances but also on the interpretive and ethical work carried out by teachers. This study has limitations related to its small and context-specific sample and the reliance on self-reported data. Future research should examine more diverse institutional settings, incorporate classroom observations, and include student perspectives to strengthen empirical grounding. Further investigations may also explore how AI-mediated evaluation practices evolve and how institutional ecosystems shape their sustainability. Overall, the findings indicate that responsible and context-sensitive AI use can support, but not replace, teachers' evaluative judgment in higher education.

ACKNOWLEDGEMENTS

The author would like to express sincere appreciation to the administrations of the three participating universities in Indonesia for granting permission to conduct this study. Special thanks are also extended to the teachers who participated and shared valuable insights that greatly enriched the findings of this research.

CONFLICT OF INTERESTS

The author declares no conflict of interest.

AUTHOR(S) CONTRIBUTION

Lapele, F.: Conceptualization (lead), methodology (lead), writing—original draft (lead), data analysis (supporting), review and editing (supporting). Romrome, A.Y.: Methodology (supporting), data analysis (lead), review (supporting). Prihono, E.W.: data analysis (lead), editing (lead), review (supporting).

REFERENCES

- Abdullah, M. Y. (2025). Probing into EFL students' perceptions about the impact of utilizing AI-powered tools on their academic writing practices. *Education and Information Technologies*, 30(15), 21189–21220. <https://doi.org/10.1007/s10639-025-13601-w>
- Ahmed, I., Ghafoor, A., Liliuara, D., & Akyuningrum, V. Q. (2025). The impact of AI learning tools on ESL learners' motivation and success: A systematic literature review. *English Learning Innovation*, 6(1), 134–142. <https://doi.org/10.22219/ENGLIE.V6I1.38961>
- Aini, N., Kurniarahman, I., Widiati, U., Cahyono, B. Y., & Basthomi, Y. (2024). Indonesian University students' perspectives on integrating AIED into English language learning. *Issues in Educational Research*, 34(3), 803–824.
- Alam, A., & Mohanty, A. (2023). Educational technology: Exploring the convergence of technology and pedagogy through mobility, interactivity, AI, and learning tools. In *Cogent Engineering (Vol. 10)*. Cogent OA. <https://doi.org/10.1080/23311916.2023.2283282>
- Aljabr, F. S., & Al-Ahdal, A. A. M. H. (2024). Ethical and pedagogical implications of AI in language education: An empirical study at Ha'il University. *Acta Psychologica*, 251, 104605. <https://doi.org/10.1016/j.actpsy.2024.104605>
- Al-kfairy, M. (2024). Factors impacting the adoption and acceptance of ChatGPT in educational settings: A Narrative Review of Empirical Studies. *Applied System Innovation*, 7(6), 110. <https://doi.org/10.3390/ASI7060110>
- Al-khresheh, M. H. (2024). Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. *Computers and Education: Artificial Intelligence*, 6, 100218. <https://doi.org/10.1016/J.CAEAI.2024.100218>
- Alqahtani, N., & Wafula, Z. (2025). Artificial Intelligence integration: Pedagogical strategies and policies at leading universities. *Innovative Higher Education*, 50(2), 665–684. <https://doi.org/10.1007/s10755-024-09749-x>
- Asad, M. M., Hussain, N., Wadho, M., Khand, Z. H., & Churi, P. P. (2021). Integration of e-learning technologies for interactive teaching and learning process: An empirical study on higher education institutes of Pakistan. *Journal of Applied Research in Higher Education*, 13(3), 649–663. <https://doi.org/10.1108/JARHE-04-2020-0103>
- Braun, V., & Clarke, V. (2021). *Thematic Analysis: A Practical Guide*. In SAGE Publications.
- Burner, T., Lindvig, Y., & Wærness, J. I. (2025). "We should not be like a dinosaur"—Using AI technologies to provide formative feedback to students. *Education Sciences* 2025, 15(1), 58. <https://doi.org/10.3390/EDUCSCI15010058>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university

- teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1), 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Chang, W. L., & Sun, J. C. Y. (2024). Evaluating AI's impact on self-regulated language learning: A systematic review. *System*, 126, 103484. <https://doi.org/10.1016/J.SYSTEM.2024.103484>
- Compagnoni, I., Oguilve, V., & Wen, W. (2025). Ethical frontiers of AI in language pedagogy: Navigating transparency, accountability, and equity. *International Journal of Linguistics*, 17 (4), 1–25. <https://doi.org/10.5296/IJL.V17I4.22768>
- Cotos, E., Huffman, S., & Link, S. (2020). Understanding graduate writers' interaction with and impact of the research writing tutor during revision. *Journal of Writing Research*, 12(1), 187–232. <https://doi.org/10.17239/JOWR-2020.12.01.07>
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Filiz, O., Kaya, M. H., & Adiguzel, T. (2025). Teachers and AI: Understanding the factors influencing AI integration in K-12 education. *Education and Information Technologies*, 30(13), 17931–17967. <https://doi.org/10.1007/s10639-025-13463-2>
- Kaldaras, L., Haudek, K., & Krajcik, J. (2024). Employing automatic analysis tools aligned to learning progressions to assess knowledge application and support learning in STEM. *International Journal of STEM Education*, 11(1), 57. <https://doi.org/10.1186/s40594-024-00516-0>
- Kanchana, S., & Saha, P. R. (2025). Evolving student assessment: AI-driven rubrics for personalized and equitable English language learning. *Journal of Engineering Education Transformations*, 38(Special Issue), 584–590. <https://doi.org/10.16920/JEET/2025/V38IS2/25072>
- Coskun, T. K., & Alper, A. (2024). Evaluating the evaluators: A comparative study of AI and teacher assessments in higher education. *Digital Education Review*, 45, 124–140. <https://doi.org/10.1344/der.2024.45.124>
- Kim, S. J. (2024). Research ethics and issues regarding the use of ChatGPT-like artificial intelligence platforms by authors and reviewers: A narrative review. *Science Editing*, 11(2), 96–106. <https://doi.org/10.6087/KCSE.343>
- Kılınc, D. D., & Mansız, D. (2024). Examination of the reliability and readability of chatbot generative pretrained transformer's (ChatGPT) responses to questions about orthodontics and the evolution of these responses in an updated version. *American Journal of Orthodontics and Dentofacial Orthopedics*, 165, 546–555. <https://doi.org/10.1016/j.ajodo.2023.11.012>
- Lazăr, A. M., Repanovici, A., Popa, D., Ionas, D. G., & Dobrescu, A. I. (2024). Ethical principles in AI use for assessment: Exploring students' perspectives on ethical principles in academic publishing. *Education Sciences* 2024, 14(11), 1239. <https://doi.org/10.3390/EDUCSCI14111239>
- Luo, L., & Yusuf, A. (2025). Bridging AI and pedagogy: How AI-adaptive feedback shapes Chinese EFL students' writing engagement, metacognitive writing strategies, and writing performance. *Assessment & Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2025.2548919>
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28(4), 4221–4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Perkins, M., Furze, L., Roe, J., & Macvaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A Framework for ethical integration of generative AI

- in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6). <https://doi.org/10.53761/Q3AZDE36>
- Pratama, R. M. D., & Hastuti, D. P. (2024). The use of Artificial Intelligence to improve EFL students' writing skill. *English Learning Innovation*, 5(1), 13–25. <https://doi.org/10.22219/ENGLIE.V5I1.30212>
- Radanliev, P. (2025). AI ethics: Integrating transparency, fairness, and privacy in AI development. *Applied Artificial Intelligence*, 39(1). <https://doi.org/10.1080/08839514.2025.2463722>
- Schiff, D. (2022). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*, 32(3), 527–563. <https://doi.org/10.1007/s40593-021-00270-2>
- Singun, A. (2025). Unveiling the barriers to digital transformation in higher education institutions: A systematic literature review. *Discover Education*, 4(1), 37. <https://doi.org/10.1007/s44217-025-00430-9>
- Tubella, A. A., Mora-Cantalops, M., & Nieves, J. C. (2024). How to teach responsible AI in higher education: Challenges and opportunities. *Ethics and Information Technology*, 26(1), 3. <https://doi.org/10.1007/s10676-023-09733-7>
- Winstone, N. E., Balloo, K., & Carless, D. (2022). Discipline-specific feedback literacies: A framework for curriculum design. *Higher Education*, 83(1), 57–77. <https://doi.org/10.1007/s10734-020-00632-0>
- Yan, Y., & Liu, H. (2025). Ethical framework for AI education based on large language models. *Education and Information Technologies*, 30(8), 10891–10909. <https://doi.org/10.1007/s10639-024-13241-6>
- Yavuz, F., Çelik, Ö., & Yavaş Çelik, G. (2025). Utilizing large language models for EFL essay grading: An examination of reliability and validity in rubric-based assessments. *British Journal of Educational Technology*, 56(1), 150–166. <https://doi.org/10.1111/BJET.13494>
- Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods (6th ed.)*. SAGE Publications.
- Ziqi, C., Xinhua, Z., Qi, L., & Wei, W. (2024). L2 students' barriers in engaging with form and content-focused AI-generated feedback in revising their compositions. *Computer Assisted Language Learning*, 1-21. <https://doi.org/10.1080/09588221.2024.2422478>