



The influence of ai-assisted writing tools and digital feedback on EFL students' linguistic accuracy and writing performance in higher education contexts

Untung Nopriansyah✉

UIN Raden Intan Lampung, Indonesia

✉ untung.nopri@radenintan.ac.id

Article Information

Submitted April 5, 2026

Revised May 15, 2026

Accepted June 9, 2026

Keywords:

Artificial Intelligence-Assisted Writing;
Digital Feedback;
EFL Writing;
Grammatical Accuracy;
Higher Education;
Writing Performance.

Abstract

Background: The integration of artificial intelligence (AI) into language education has accelerated markedly in recent years, reshaping how feedback is delivered in second-language (L2) writing classrooms.

Aim: This study examined the influence of AI-assisted writing tools and structured digital feedback on the grammatical accuracy of English as a Foreign Language (EFL) students in higher education.

Method: A quasi-experimental, pre-test/post-test control-group design was used, comparing an experimental group receiving AI-based feedback ($n = 36$) with a control group receiving conventional teacher feedback ($n = 36$) across a six-week intervention. Grammatical accuracy scores were obtained at pre-test and post-test for both groups.

Results: The experimental group improved from a pre-test mean of 70.5 to a post-test mean of 84.2 ($SD = 5.4$), while the control group improved from 71.1 to 76.1 ($SD = 6.0$). An independent-samples t-test indicated a statistically significant between-group difference favoring the experimental group ($t = 3.87, p < .05$). These findings suggest that AI-assisted feedback was associated with a larger gain in grammatical accuracy than conventional teacher feedback over the same intervention period.

Conclusion: Fluency and revision-process outcomes originally planned as secondary measures are not yet available and are flagged as such rather than estimated. The results are discussed in relation to prior automated-feedback research and the broader debate on the relative contributions of AI-generated and human feedback in second-language writing instruction, together with the methodological caveats that necessarily accompany a two-group comparison of this scope.

I. Introduction

The integration of artificial intelligence (AI) into language education has accelerated markedly in recent years, reshaping how feedback is delivered in second-language (L2) writing classrooms. Automated writing evaluation (AWE) systems and generative AI tools now offer learners immediate, algorithmically generated feedback on grammar, lexical choice, and mechanics, a shift that has generated both enthusiasm and caution among applied linguists and writing instructors.

A substantial body of research has documented positive effects of automated written corrective feedback on L2 writing accuracy. Barrot (2023), for instance, found that automated written corrective feedback delivered measurable gains in learners' grammatical accuracy, while Guo, Feng, and Hua (2021) reported that EFL students were able to use automated written corrective feedback effectively during research writing tasks, though uptake varied by error type. Randomized evidence from Wei, Wang, and Dong (2023) further indicated that automated writing evaluation improved second-language writing outcomes for Chinese EFL learners relative to conventional instruction alone. Complementing these findings, Yan (2023)

demonstrated that ChatGPT-mediated feedback supported measurable gains in EFL learners' writing skills within an experimental design.

At the same time, researchers caution against treating AI feedback as a wholesale substitute for human instruction. Malik et al. (2023) found that higher-education students held mixed perceptions of AI in academic essay writing, valuing its efficiency while expressing concern about over-reliance and reduced critical engagement. Similarly, Liang, Huang, and Teo (2024) reported that Chinese university EFL learners' perceptions of AI-assisted writing were shaped by trust, perceived accuracy, and the extent to which feedback addressed meaning rather than form alone. Earlier work by Waer (2021) showed that automated writing evaluation could reduce writing apprehension and strengthen grammatical knowledge, yet noted that such tools address surface-level accuracy more consistently than higher-order communicative competence. Karim and Nassaji (2019), in their critical synthesis of decades of written corrective feedback research, emphasize that the effectiveness of any feedback source depends heavily on how learners process and act upon it, a caution equally relevant to AI-generated feedback.

Despite this growing literature, direct experimental comparisons between AI-based feedback and conventional teacher feedback remain comparatively scarce in Indonesian higher-education EFL contexts, where classroom conditions, learner proficiency profiles, and institutional access to AI tools differ from those in the predominantly Chinese, Egyptian, and Anglophone settings that dominate the existing evidence base. Most available studies also concentrate on a single outcome domain, typically accuracy, without jointly examining fluency and the process-level revision behavior that may explain why accuracy gains occur.

This study therefore reports a controlled, comparative investigation designed to examine whether AI-assisted feedback produced greater gains in grammatical accuracy than conventional teacher feedback among Indonesian undergraduate EFL students over a six-week instructional period. In doing so, the study aims to contribute evidence beyond the East Asian and Middle Eastern contexts that dominate the existing literature on this comparison.

II. Literature Review

2.1 Artificial Intelligence and Automated Feedback in Language Learning

AI-based systems, ranging from rule-based grammar checkers to large language model (LLM) writing assistants, are increasingly embedded in language-learning environments to support adaptive, learner-paced correction. Malik et al. (2023) and Liang et al. (2024) both note that students generally perceive such tools as efficient and accessible, though acceptance is moderated by perceived reliability and the tool's capacity to explain, rather than merely flag, an error. This distinction between corrective and explanatory feedback is central to the theoretical rationale of the present study, since explanatory depth is hypothesized to influence whether learners merely correct or genuinely internalize a linguistic rule.

2.2 Feedback and Second Language Writing Development

Feedback has long occupied a central position in second-language acquisition theory. Karim and Nassaji's (2019) synthesis of written corrective feedback research traces the field's trajectory from Truscott's (1996) skepticism about the value of error correction to the substantial subsequent evidence, summarized across numerous experimental studies, that focused and timely corrective feedback can support measurable gains in accuracy. Within this tradition, noticing is frequently proposed as the mechanism through which feedback translates into learning: learners must first notice a gap between their output and the target form before restructuring their interlanguage system. AI-generated feedback, by virtue of its immediacy,

may enhance the salience of such gaps, a possibility supported indirectly by Barrot (2023) and Guo et al. (2021).

2.3 Comparative Evidence on AI versus Teacher Feedback

Direct comparisons between AI and teacher feedback remain limited but growing. Waer (2021) found that automated writing evaluation reduced learners' writing apprehension while improving grammatical knowledge, yet did not directly benchmark these gains against a teacher-feedback control condition matched in intensity and duration. Wei et al. (2023) employed a randomized controlled design and found automated writing evaluation effective relative to standard instruction, lending methodological precedent to the design proposed here. However, none of the reviewed studies simultaneously tracked accuracy, fluency, and revision-process indicators within a single Indonesian higher-education sample, which constitutes the specific gap this proposal addresses.

2.4 Research Gap and Rationale

Three gaps motivate the present study. First, comparative experimental evidence contrasting AI feedback with teacher feedback under matched conditions is scarce outside a small number of national contexts. Second, most studies report outcome-level accuracy gains without examining the revision process that presumably mediates them. Third, few studies have been conducted with Indonesian undergraduate EFL learners, whose prior exposure to AI writing tools and classroom feedback norms may differ from the populations studied elsewhere. This study addresses the first and third of these gaps directly; the second, concerning revision process, remains only partially addressed here given the unavailability of revision-log data at the time of analysis.

III. Method

3.1 Research Design

The study adopted a quasi-experimental, non-equivalent control-group design with pre-test and post-test measurement. Intact undergraduate EFL writing classes were assigned, at the class level, to either the experimental (AI-assisted feedback) or control (teacher feedback) condition, in order to preserve naturalistic classroom instruction while enabling controlled comparison.

3.2 Participants

The sample comprised 72 undergraduate students enrolled in an EFL academic writing course (experimental group, $n = 36$; control group, $n = 36$). Participants were screened for comparable baseline English proficiency using an existing institutional placement measure prior to group assignment. Participation was voluntary and followed institutional research-ethics procedures, including informed consent and the right to withdraw without academic penalty.

3.3 Instruments

- Writing tasks: parallel argumentative-essay prompts of comparable difficulty for the pre-test and post-test, scored using an analytic rubric covering grammatical accuracy, lexical range, coherence, and fluency.
- Grammatical accuracy index: error-free clause ratio and error frequency per 100 words, coded by trained raters blind to group assignment.
- AI feedback tool: an AI-based writing assistant providing automated grammar, lexical, and structural feedback, used by the experimental group during drafting and revision.
- Teacher feedback protocol: a standardized feedback checklist used by instructors in the control group to ensure comparable feedback frequency and coverage to the experimental condition.

- Revision-tracking log: intended as a record of the number, type, and timing of revisions made between drafts; log data are not included in the results reported here.
- Background questionnaire: captured demographic information, prior exposure to AI writing tools, and self-reported writing confidence.

3.4 Procedure

Following baseline data collection (pre-test essay), the six-week intervention proceeded with both groups completing one guided writing-and-revision cycle per week. The experimental group drafted, submitted work to the AI writing assistant, reviewed its feedback, and revised; the control group followed an equivalent draft–feedback–revise cycle using the standardized teacher feedback protocol, matched in frequency and turnaround time to minimize confounds related to feedback quantity. A post-test essay, using a topic parallel in difficulty to the pre-test, was administered in the week following the intervention.

3.5 Ethical Considerations

The study followed institutional research-ethics procedures. Participant data were anonymized at the point of analysis, and no student was denied instructional feedback as a condition of participation, since both conditions provided a recognized form of feedback rather than a no-feedback control.

IV. Result and Discussion

Result

Grammatical accuracy scores (0–100 scale) were compared between the experimental (AI-feedback) and control (teacher-feedback) groups at pre-test and post-test. Descriptive statistics are reported in Table 1.

Table 1. Pre-test and post-test grammatical accuracy scores by group

Group	Pre-test M	Post-test M	Post-test SD
Experimental (AI feedback), n = 36	70.5	84.2	5.4
Control (teacher feedback), n = 36	71.1	76.1	6.0

At pre-test, the two groups had comparable mean accuracy scores (experimental M = 70.5; control M = 71.1), suggesting reasonable baseline equivalence between conditions prior to the intervention. Following the six-week intervention, the experimental group showed a larger increase in mean accuracy (+13.7 points, from 70.5 to 84.2) than the control group (+5.0 points, from 71.1 to 76.1).

An independent-samples t-test comparing post-test accuracy scores between groups indicated a statistically significant difference favoring the experimental group, $t = 3.87$, $p < .05$. Only the t-value and significance level ($p < .05$) were supplied for this comparison; exact degrees of freedom, the precise p-value, and an effect-size estimate (e.g., Cohen’s d) were not reported and are therefore not stated numerically here. Readers should note this limitation when evaluating the magnitude, as opposed to the statistical significance, of the observed difference.

Fluency and revision-log outcomes, both included in the original methodological plan, were not available at the time of this analysis and are consequently not reported. Their omission means the present results speak only to grammatical accuracy and should not be generalized to writing fluency or to the revision process without further data.

Discussion

The observed post-test advantage for the AI-feedback group is broadly consistent with prior research reporting benefits of automated feedback for L2 grammatical accuracy. Barrot (2023) similarly found that automated written corrective feedback produced measurable gains in learners' accuracy, and Guo et al. (2021) reported that EFL students were able to make effective use of automated feedback during writing tasks. The randomized-trial evidence from Wei et al. (2023), which found automated writing evaluation more effective than standard instruction for Chinese EFL learners, offers a close methodological parallel to the present comparison, despite differences in learner population and institutional context.

The magnitude of the gain difference between groups (13.7 points versus 5.0 points) is larger than the accuracy gains typically reported in comparable automated-feedback studies, which raises an interpretive question rather than settling one. Karim and Nassaji (2019) caution that the effectiveness of any corrective feedback source depends substantially on how learners notice and act upon it, not solely on the feedback source itself; the present design cannot rule out that the AI-feedback group also received more frequent or more immediately actionable feedback cycles than the control group, given that feedback frequency, while intended to be matched, was not independently verified through the revision-log measure originally planned. This is a meaningful limitation, since a confound of feedback frequency, rather than feedback source per se, could plausibly account for part of the observed difference.

At the same time, the finding that AI feedback outperformed teacher feedback specifically on a form-focused accuracy measure aligns with the broader pattern noted by Malik et al. (2023) and Liang et al. (2024), who observed that students and instructors alike tend to view AI tools as particularly well-suited to surface-level, rule-based correction. Whether the same advantage would extend to fluency, coherence, or communicative adequacy, the outcomes originally planned as secondary measures in this study, remains an open question that the present results cannot answer, since those data were not available for analysis.

Overall, the results offer preliminary support for the position that AI-assisted feedback can accelerate gains in grammatical accuracy over a short instructional period, while underscoring, consistent with Waer (2021) and the broader written corrective feedback literature (Karim & Nassaji, 2019), that accuracy is only one dimension of writing competence and should not be treated as a proxy for overall writing quality.

Limitations and Future Directions

Several limitations qualify the interpretation of these results. First, the use of intact classes rather than full random assignment limits causal inference relative to a true experimental design. Second, only a t-value and significance threshold were available for the primary comparison; the absence of exact p-values, degrees of freedom, and effect-size statistics limits the precision with which the magnitude of the effect can be characterized and prevents a full assessment of practical significance. Third, the fluency and revision-log measures originally planned as part of the design were not available for this analysis, so the present findings should be read as accuracy-specific rather than as evidence of a broader writing-performance advantage. Fourth, the six-week intervention window is relatively brief, and it remains unknown whether the observed advantage would persist, narrow, or reverse over a longer period or after the removal of AI support.

Future reporting of this study would benefit from the inclusion of exact test statistics (df, p-value, and Cohen's d) for the accuracy comparison, analysis of the fluency and revision-log

data already planned in the methodology, and a delayed post-test to assess retention. A hybrid condition combining AI and teacher feedback, an approach suggested by several of the studies reviewed here as potentially outperforming either source alone, would also be a natural extension of this design.

V. Conclusion

This study examined the influence of AI-assisted writing tools and structured digital feedback on the grammatical accuracy of Indonesian EFL undergraduate students through a six-week quasi-experimental intervention. The findings indicate that students who received AI-assisted feedback demonstrated significantly greater improvement in grammatical accuracy than those who received conventional teacher feedback. While both groups improved between the pre-test and post-test, the experimental group achieved a substantially larger gain, suggesting that AI-generated feedback can effectively support learners in identifying and correcting grammatical errors during the writing process.

These findings contribute to the growing body of evidence that AI-assisted writing tools can complement language instruction by providing immediate and consistent corrective feedback. Nevertheless, the results should be interpreted within the limitations of the present study. The conclusions are restricted to grammatical accuracy because data on writing fluency and revision processes were not available for analysis. In addition, the quasi-experimental design and relatively short intervention period limit the extent to which causal and long-term conclusions can be drawn.

Overall, the study supports the pedagogical value of integrating AI-assisted feedback into higher education EFL writing classrooms as a supplement rather than a replacement for teacher feedback. Future research should investigate longer intervention periods, include broader dimensions of writing performance, and explore hybrid feedback models that combine the strengths of AI-generated and teacher-mediated feedback to promote more comprehensive writing development.

VI. Acknowledgments

The author would like to express sincere gratitude to the undergraduate EFL students who voluntarily participated in this study and to the lecturers who facilitated the implementation of the research. Appreciation is also extended to the academic staff of UIN Raden Intan Lampung for their administrative and institutional support throughout the research process. Finally, the author acknowledges all researchers whose previous studies have provided valuable theoretical and empirical foundations for this work.

VII. References

- Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, 36(4), 584–607. <https://doi.org/10.1080/09588221.2021.1936071>
- Guo, Q., Feng, R., & Hua, Y. (2021). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*, 35(9), 2312–2331. <https://doi.org/10.1080/09588221.2021.1879161>
- Karim, K., & Nassaji, H. (2019). The effects of written corrective feedback: A critical synthesis of past and present research. *Instructed Second Language Acquisition*, 3(1), 28–52.

- Liang, J., Huang, F., & Teo, T. (2024). Understanding Chinese university EFL learners' perceptions of AI in English writing. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1), 1–16.
- Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki, M. (2023). Exploring artificial intelligence in academic essay: Higher education student's perspective. *International Journal of Educational Research Open*, 5, 100296.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327–369.
- Waer, H. (2021). The effect of integrating automated writing evaluation on EFL writing apprehension and grammatical knowledge. *Innovation in Language Learning and Teaching*, 17(1), 47–71. <https://doi.org/10.1080/17501229.2021.1914062>
- Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. *Frontiers in Psychology*, 14, Article 1249991. <https://doi.org/10.3389/fpsyg.2023.1249991>
- Yan, C. (2023). Enhancing EFL learners' writing skills through ChatGPT: An experimental study. *Frontiers in Psychology*, 14, Article 1260843.