The Effect Of Generative Ai Assistance On Indonesian University Students' Academic Writing

Sofia Dwi Sanzain1✉, Ilham[2], Rima Rahmaniah[3]
(1,2,3) Department of English Education, Muhammadiyah University of Mataram

(Received: 17-04-2026 Revision: 23-05-2026 Accepted: 07-08-2026 Published: 22-08-2026)

✉ Sofia Dwi Sanzain
(sanzainsofiadwi@gmail.com)
DOI: 10.31605/ijes.v8i1.6843

Abstract

The use of Generative Artificial Intelligence (AI) in academic writing instruction has increasingly attracted attention because of its potential to support idea generation, paragraph organization, and linguistic accuracy. Despite its growing use, empirical evidence regarding its use in Indonesian EFL higher education remains limited. This study aimed to examine changes in students' academic writing performance following Generative AI-assisted writing instruction. The study employed a pre-experimental one-group pretest–posttest design involving 25 undergraduate EFL students. Data were collected through an academic writing test and a student perception questionnaire. Since the difference scores were not normally distributed, the Wilcoxon Signed-Rank Test was used to compare the pre-test and post-test scores. The findings revealed a statistically significant improvement in students' academic writing performance following the intervention ($Z = -4.374$, $p < .001$). In addition, students reported generally positive perceptions of Generative AI, particularly in generating ideas, organizing arguments, and increasing their confidence in writing. These findings suggest that AI-assisted writing activities may support the development of academic writing skills in Indonesian EFL higher education. However, because this study employed a pre-experimental one-group pretest–posttest design without a control group, the findings should be interpreted with caution and cannot establish a causal relationship between the intervention and the observed improvement. Future studies involving larger samples, control groups, and longer intervention periods are recommended.

Keyword: *Generative Artificial Intelligence; Academic Writing; Writing Performance; EFL students'; Higher Education*

1. INTRODUCTION

The rapid advancement of generative artificial intelligence (AI) has significantly transformed teaching and learning practices across higher education, particularly in academic writing (Krause et al., 2025). The emergence of AI-powered writing assistants, such as ChatGPT, has changed the way university students generate ideas, organize arguments, revise drafts, and refine linguistic accuracy throughout the writing process (Kim et al., 2025). By leveraging large language models trained on extensive textual data, these technologies provide immediate support for grammar correction, vocabulary enhancement, sentence restructuring, and text organization, allowing learners to improve the quality of their written work more efficiently. As a result, generative AI has become an integral component of digital learning environments and is increasingly incorporated into English language education to facilitate academic writing development. Despite these educational benefits, the rapid adoption of AI-assisted writing tools has also generated growing concerns regarding academic integrity, learner autonomy, originality, and the development of higher-order thinking skills. Consequently, understanding the educational implications of generative AI has become an important research priority in technology-enhanced language learning.

Academic writing is widely recognized as one of the most essential competencies for university students because it serves as the primary medium for communicating scholarly

knowledge, disseminating research findings, and participating in academic discourse (Mason & Merga, 2022). Producing high-quality academic writing requires students to demonstrate not only grammatical accuracy but also critical thinking, logical reasoning, coherent argumentation, effective source integration, and appropriate use of academic conventions (Feng & Mei, 2022). These competencies enable learners to construct evidence-based arguments and communicate ideas systematically in accordance with disciplinary standards. However, mastering academic writing remains particularly challenging for students learning English as a Foreign Language (EFL), who frequently experience difficulties in organizing ideas, maintaining textual coherence, integrating relevant literature, and employing appropriate academic vocabulary. These persistent challenges often limit students' ability to produce academically acceptable written work and highlight the need for innovative instructional approaches capable of supporting both linguistic development and higher-order writing skills.

The increasing integration of generative AI into higher education has encouraged educators and researchers to explore its potential as a pedagogical tool for improving students' academic writing (Qian, 2025). Existing studies have demonstrated that generative AI supports multiple stages of the writing process, including idea generation, language refinement, grammatical correction, vocabulary development, and text organization (Mekheimer, 2025). Unlike conventional proofreading tools, generative AI provides interactive feedback that enables learners to revise their writing iteratively and receive immediate suggestions tailored to their textual input. Such features promote more efficient writing practices while encouraging students to reflect on language use and rhetorical organization during the revision process. Consequently, generative AI is increasingly viewed not merely as a technological innovation but also as an instructional resource capable of supporting the development of academic writing skills in English language education.

In addition to improving the linguistic quality of students' writing, generative AI has also been shown to support learners' cognitive and affective development throughout the writing process. Previous studies have reported that AI-assisted feedback can reduce writing anxiety, strengthen learners' confidence, and increase motivation by providing immediate, personalized, and non-judgmental assistance during drafting and revision. Through adaptive feedback and contextual explanations, AI allows learners to identify weaknesses in their writing, evaluate alternative expressions, and make informed revisions independently (Khojasteh et al., 2025). These opportunities encourage greater learner engagement and promote self-regulated learning, enabling students to become more active participants in their own writing development. Nevertheless, the educational value of AI depends largely on students' ability to evaluate and critically interpret AI-generated suggestions rather than accepting them uncritically(Darwin et al., 2024). Therefore, the integration of AI into writing instruction should be viewed as a complementary learning strategy that enhances, rather than replaces, students' analytical thinking and independent academic writing.

Although the educational benefits of generative AI have been widely acknowledged, empirical evidence regarding its effectiveness in improving academic writing remains inconclusive (Farrokhnia et al., 2026). Existing studies have predominantly emphasized improvements in grammatical accuracy, vocabulary use, and students' perceptions of AI-assisted writing, while relatively limited attention has been devoted to higher-order aspects of writing, including argument development, discourse coherence, logical organization, and critical reasoning (Khan et al., 2025). Furthermore, research examining the implementation of generative AI in Indonesian higher education remains scarce, particularly studies employing experimental approaches capable of establishing causal

relationships between AI-assisted learning and students' academic writing performance. These limitations indicate the need for more rigorous empirical studies that examine the effectiveness of generative AI in improving students' overall academic writing performance while also exploring learners' perceptions of AI-assisted writing in EFL higher education contexts.

Despite the growing body of literature on generative AI in education, important research gaps remain. Previous studies have primarily focused on students' perceptions, technology acceptance, and improvements in grammatical accuracy. Comparatively few studies have investigated the effect of generative AI on overall academic writing performance, particularly higher-order writing skills such as organization, coherence, argument development, and academic reasoning. Furthermore, experimental studies examining the effectiveness of generative AI in Indonesian EFL higher education contexts remain limited. To address these gaps, this study employed a pre-experimental one-group pretest–posttest design to examine changes in students' academic writing performance following Generative AI-assisted writing instruction. In addition to examining changes in students' overall academic writing performance through pre-test and post-test scores, this study also explored students' perceptions of using generative AI in academic writing. The findings are expected to contribute empirical evidence regarding the effectiveness of generative AI in supporting academic writing and provide practical insights for its integration into EFL writing instruction in Indonesian higher education.

2. RESEARCH METHODS

This study employed a quantitative approach using a pre-experimental one-group pretest-posttest design to examine changes in students' academic writing performance following the use of generative artificial intelligence (AI) assistance (Jmaiel et al., 2025). This research design was selected because it enables the measurement of changes in students' writing performance before and after the intervention within the same group, allowing the researchers to evaluate differences in writing performance following the AI-assisted instructional activities (Chung et al., 2021). The study was conducted in the English Language Education Study Program at Universitas Muhammadiyah Mataram. The participants consisted of 25 fifth-semester students enrolled in the Academic Writing course. Purposive sampling was employed because the selected class had already been enrolled in the Academic Writing course and represented the target population of the study. The participants were selected because they were enrolled in the Academic Writing course and were available to participate throughout the study. Although the students demonstrated varying levels of initial writing proficiency, they shared the same instructional context, received identical learning activities, and completed the same AI-assisted writing intervention, making them appropriate participants for this study.

The research instruments comprised an academic writing test and a student perception questionnaire (Feng & Mei, 2022). Prior to data collection, all research instruments were subjected to content validation by an expert in English Language Teaching (ELT). The expert evaluated the clarity, relevance, appropriateness, and alignment of each instrument with the research objectives. Based on the expert's feedback, several revisions were made to improve the wording and organization of the writing test instructions and questionnaire items before they were administered to the participants. The writing test was administered as both a pre-test and a post-test to assess students' academic writing performance before and after the intervention. Students' writing was evaluated based on five assessment criteria: content, organization, vocabulary, language use, and mechanics (Wale & Bogale, 2021). In addition, a structured questionnaire consisting of 25 items was administered at the end of the study to explore students'

perceptions, learning motivation, and attitudes toward the integration of generative AI in academic writing activities. The questionnaire employed a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The questionnaire provided complementary information regarding students' experiences with AI-assisted writing beyond the quantitative results obtained from the writing test.

The research procedure was conducted in four consecutive stages: pre-test, treatment, post-test, and questionnaire administration. During the first stage, all participants completed a pre-test by writing a 150–200-word academic essay on the topic *The Challenges and Opportunities of Using AI in the Modern Era* to assess their initial academic writing performance. The treatment consisted of two instructional sessions, each lasting 90 minutes. During the first session, students were introduced to the basic functions of ChatGPT and received guidance on using AI to generate ideas, develop outlines, and organize their essays. They were allowed to formulate their own prompts based on their writing needs while being encouraged to critically evaluate the AI-generated responses. In the second session, students used ChatGPT to draft, revise, and refine their essays with guidance from the lecturer, who emphasized independent revision and adherence to academic integrity principles. Following the intervention, all participants completed a post-test by writing another 150–200-word academic essay on the same topic using the same assessment rubric to evaluate their writing performance after the treatment. Finally, all participants completed the perception questionnaire to provide feedback regarding their experiences and views on the integration of generative AI into academic writing instruction.

The collected data were analyzed using both descriptive and inferential statistical techniques with IBM SPSS Statistics version 25. Descriptive statistics were calculated to summarize the pre-test and post-test results, including the minimum score, maximum score, mean, standard deviation, and variance. Prior to hypothesis testing, the normality of the difference scores (post-test minus pre-test) was examined using the Shapiro–Wilk test. The results indicated that the difference scores were not normally distributed ($W = 0.918$, $p = .046$). Therefore, the non-parametric Wilcoxon Signed-Rank Test was employed to compare students' academic writing performance before and after the intervention. Statistical significance was established at the .05 level. In addition, responses to the 25-item perception questionnaire were analyzed descriptively by calculating the mean score for each item and interpreting the results according to the corresponding Likert scale categories to provide a comprehensive description of students' perceptions of generative AI-assisted academic writing instruction. Participation in this study was voluntary. Before data collection, all participants were informed about the purpose of the study and agreed to participate. Participants' responses were kept confidential and used solely for research purposes.

3. RESULTS AND DISCUSSION

3.1. Result

Table 1. Descriptive Statistics of Students' Academic Writing Performance Before and After the Intervention

Generative AI Media Digital	Pre-test	Post-test
N	**25**	**25**
Mean	**64.92**	**82.08**
Standard Deviation	**14.04**	**7.90**
Variance	**197.06**	**62.41**
Minimum	**42**	**59**
Maximum	**86**	**93**

Table 1 presents the descriptive statistics of students' academic writing performance before and after the intervention. The pre-test scores of the 25 participants ranged from 42 to 86, with a mean score of 64.92 (SD = 14.04). These results indicate considerable variation in students' initial academic writing performance prior to the intervention.

Following the intervention, the post-test scores ranged from 59 to 93, with a mean score of 82.08 (SD = 7.90). The increase in the mean score indicates improved writing performance after the intervention, while the lower standard deviation suggests less variability in students' scores. This pattern is consistent with more homogeneous writing performance among participants following the AI-assisted writing activities.

Table 2. Table 2. Normality Test Results for the Pre-test Scores

Types of Test	Statistic	Degrees of Freedom	Sig.	Description
Kolmogorov-Semirnov	0.181	25	0.033	Not normal
Shapiro- Wilk	0.907	25	0.026	Not normal

Table 2 presents the results of the normality test for the pre-test scores, indicating that the distribution of students' academic writing performance prior to the intervention did not satisfy the assumption of normality. The Kolmogorov–Smirnov significance value recorded was 0.033, and the Shapiro–Wilk significance value was 0.026, both of which fall below the standard threshold of 0.05. These outcomes indicate that the pre-test data deviate significantly from a normal distribution, meaning that the students' initial writing abilities varied widely and were not clustered around an expected average. The non-normal distribution indicates that the pre-test scores were not evenly distributed across the participants. Therefore, the assumption of normality for the pre-test scores was not met. Considering these results, a non-parametric statistical test was selected for the subsequent analysis. Accordingly, the Wilcoxon Signed-Rank Test was employed to compare students' pre-test and post-test scores.

Table 3. Test of Normality, Presents the Results of The Normality Test for The Post-Test Scores

Types of Test	Statistic	Degrees of Freedom	Sig.	Description
Kolmogorov-Semirnov	0.164	25	0.081	Normal
Shapiro- Wilk	0.893	25	0.013	Not normal

The normality analysis of the post-test scores presented in Table 3 provides mixed indications regarding the distribution of the dataset. The Kolmogorov–Smirnov test produced a significance value of 0.081, which exceeds the conventional threshold of 0.05, indicating that the post-test scores tend to approximate a normal distribution. This result indicates that the post-test scores satisfied the assumption of normality according to the Kolmogorov–Smirnov test. However, this finding should be interpreted together with the results of the Shapiro–Wilk test

Nevertheless, the Shapiro–Wilk test provides a different indication, reporting a significance value of 0.013, which is lower than the threshold of 0.05 and therefore suggests that the post-test data still deviate from a perfectly normal distribution. Therefore, the two normality tests provide different results regarding the distribution of the post-test scores. Because one of the normality tests indicates non- normality, the post-test data cannot be definitively categorized as normally distributed. Therefore, the

Wilcoxon Signed-Rank Test was used to analyze the difference between the pre-test and post-test scores.

Table 4. Normality Test of Difference Scores

Types of test	Statistic	df	Sig.	Description
Shapiro Wilk	0.918	25	0.046	Not normal

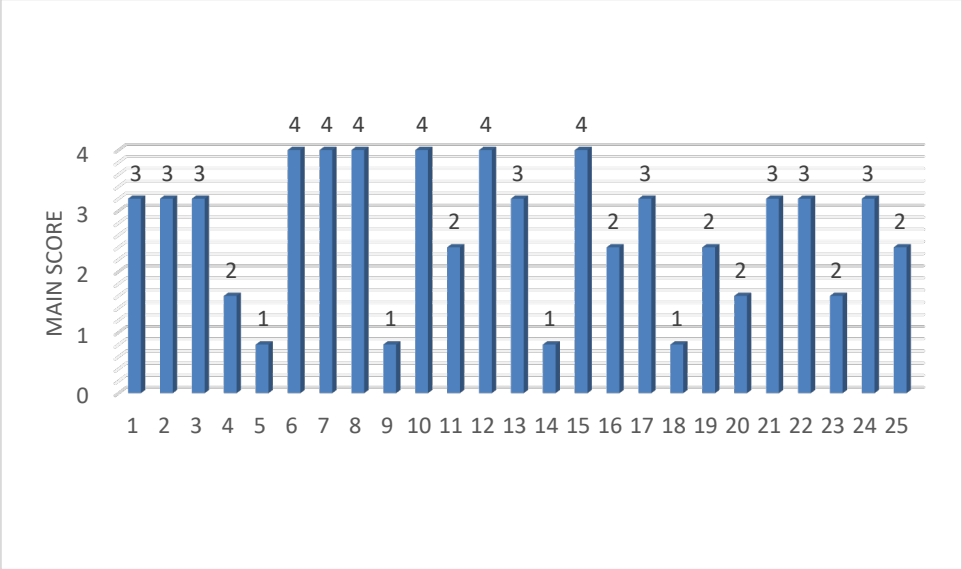
Table 5. Wilcoxon Signed-Rank Test Results

Te	Value	Description
Z	-4.374	Based on Negative Rank
Sig. (2-Tailed)	<0.001	The are Significant

Table 5 presents the results of the Wilcoxon Signed-Rank Test comparing students' pre-test and post-test scores. The analysis indicates a statistically significant difference in academic writing performance following the intervention. The table reports a Z value of -4.374, which reflects the direction and magnitude of change observed in the students' scores. The negative sign does not indicate a decline in performance; rather, it is a procedural outcome that signifies the consistent ranking of post-test scores being higher than pre-test scores. This result indicates that most participants obtained higher post-test scores than pre-test scores, reflecting a consistent pattern of improvement across the sample. The obtained Z value indicates that the observed improvement was statistically significant rather than attributable to random variation.

Further confirmation of this improvement is provided by the Asymptotic. Sig. (2-tailed) value of < 0.001, which is far below the widely accepted threshold of 0.05 for determining statistical significance. This exceptionally low significance value provides strong evidence that the observed increase in students' scores is meaningful and not attributable to random variations or chance factors. The effect size was large (r = 0.88), indicating that the magnitude of the observed improvement was substantial. These findings suggest that students' academic writing performance improved following the AI-assisted writing intervention. However, because the present study employed a pre-experimental one-group pretest-posttest design, the observed improvement should be interpreted cautiously and cannot be attributed solely to the intervention. Nevertheless, the findings provide empirical evidence that AI-assisted writing activities were associated with improved academic writing performance in this group of students

Figure 1. Mean Scores for Each of the 25 Questionnaire Items (Items 1–25) Completed by 25 Students (n = 25).



4. DISCUSSION

The quantitative findings of this research show a clear and substantial improvement in students' academic writing performance after the use of Generative AI in classroom instruction. Descriptive statistics indicate that the mean pre-test score of the 25 students was 64.92 with a relatively large standard deviation of 14.038, suggesting notable variation in initial writing proficiency. After the intervention, the mean post-test score increased to 82.08, while the standard deviation decreased to 7.900, indicating that the students' performance became more clustered around a higher average. This pattern suggests that students' writing scores became more homogeneous after the intervention, as reflected by the lower post-test standard deviation. The reduced variability in post-test scores may indicate that students with lower initial performance also showed improvement following the intervention; however, this interpretation should be made cautiously because the present study did not analyze individual learning trajectories. These findings suggest a reduction in score variability following the AI-assisted writing intervention, although the study design does not allow firm conclusions about the underlying causes of this pattern. Overall, the descriptive statistics indicate an improvement in students' academic writing performance following the AI-assisted writing intervention.

The inferential analysis further strengthens this conclusion by demonstrating that the improvement between pre-test and post-test is statistically significant. Since the normality test indicated that the data were not normally distributed, as reflected by the Shapiro–Wilk significance values of .026 in the pre-test and .013 in the post-test, the study therefore employed the Wilcoxon Signed-Rank Test to analyze the paired non-parametric data. The test produced a Z value of −4.374 with an Asymptotic. Sig. (2-tailed) < .001, which confirms that the increase in scores is highly unlikely to be due to chance alone. This result indicates a consistent pattern of positive change across the 25 paired observations rather than random fluctuation. The strong level of significance indicates that students' writing performance improved significantly following the AI-assisted writing intervention. At the same time, the choice of analysis respects statistical assumptions and supports the methodological rigor of the study. Overall, the combination of descriptive and inferential evidence suggests that the AI-assisted writing intervention was associated with improved academic writing performance in this group of students.

From a cognitive perspective, the improvement can be understood by examining how Generative AI interacts with the mental processes involved in academic writing (M. Liu et al., 2024). Writing in a foreign language requires students to generate ideas, structure

arguments, select appropriate vocabulary, and maintain grammatical accuracy at the same time, which creates a heavy cognitive load. In this study, the AI tool provided examples of topic sentences, suggestions for lexical variety, and alternative sentence structures that students could adapt and evaluate. By offloading part of the burden of language formulation to the AI system, students were able to allocate more mental resources to organizing their arguments and clarifying their main ideas. Over repeated cycles of drafting and revising, they were exposed to patterns of academic discourse that they could gradually internalize. The movement from an average of 64.92 to 82.08 can therefore be interpreted not only as a numerical change but as evidence that students were beginning to appropriate these patterns into their own writing. This cognitive support mechanism helps explain why the gains were not limited to surface-level accuracy but also reflected in more coherent and better-structured texts.

The use of Generative AI also altered the pedagogical dynamics of the writing classroom in meaningful ways. Prior to the intervention, students relied mainly on teacher corrections, which were limited by time and often focused on final products rather than the writing process. During the intervention, the AI tool allowed students to receive immediate feedback on drafts, experiment with alternative formulations, and revise their work multiple times before submitting it to the teacher. This shifted the teacher's role from sole provider of corrections to facilitator and guide who helped students interpret AI suggestions, make informed choices, and reflect on why certain options were more appropriate than others. Such a shift encouraged learners to see writing as a recursive process of planning, drafting, revising, and editing rather than a one-time task. It also promoted greater learner autonomy, since students could engage with feedback at their own pace instead of waiting passively for teacher comments. The statistical improvement in scores is therefore linked not only to the presence of AI, but to a broader reconfiguration of roles and responsibilities within the classroom.

Students' perceptions of the AI-supported writing intervention, as presented in Figure 1, indicate generally positive responses across the 25 questionnaire items, with most items receiving mean scores between 3 and 4. This pattern suggests that students generally viewed the AI-assisted writing activities positively across the questionnaire items. Although some questionnaire items received higher mean scores than others, students' responses were generally positive across the questionnaire. Overall, these perception results are consistent with the quantitative findings, suggesting that students responded positively to the AI-assisted writing intervention while their academic writing performance also improved following the intervention.

However, the perception data also hint at critical issues that must be acknowledged and addressed if AI is to be integrated responsibly into writing instruction. The lower scores on certain items may reflect concerns about becoming too dependent on AI suggestions, fears that their own ideas might be overshadowed, or uncertainty about what counts as acceptable assistance in academic settings. These concerns resonate with wider debates about the impact of AI on originality, critical thinking, and academic integrity. If students rely on AI-generated text without fully understanding or evaluating it, there is a risk that they may produce writing that is formally correct but conceptually shallow, or that does not genuinely represent their own thinking. The findings of this study therefore point to the need for explicit discussion in class about how to use AI as a support tool rather than as a replacement for personal effort. Introducing clear guidelines, reflective tasks, and opportunities to compare AI-assisted and non-assisted drafts would help students maintain a healthy balance between technological help and independent reasoning.

The interpretation of these findings must also take into account the methodological limitations of the research design. The study used a one-group pretest-posttest model with

25 participants and no control group, which means that the observed improvement cannot be attributed exclusively to the AI intervention with absolute certainty. Other factors such as increased familiarity with the test format, general practice effects, short-term motivation, or the novelty of using a new technology may also have contributed to the gains. In addition, the intervention was implemented over a relatively short period, so the data mainly reflect short-term changes rather than long-term retention of skills. The non-normal distribution of scores, while appropriately handled through non-parametric analysis, further suggests that individual trajectories may have varied more than the group averages reveal. Recognizing these limitations is important for maintaining a balanced and honest interpretation of the results. At the same time, the statistical evidence and the consistent improvement observed across the sample suggest that the AI-assisted writing intervention was associated with improved academic writing performance. However, given the one-group pretest-posttest design, these findings should not be interpreted as establishing a causal relationship. Despite these constraints, the study offers meaningful theoretical and practical contributions to the ongoing discussion about the role of Generative AI in EFL academic writing.

The combination of improved test scores, reduced score dispersion, significant Wilcoxon results, and generally positive student perceptions suggests that AI-assisted writing activities have the potential to support process-based academic writing instruction in EFL contexts. The findings support the view that technology should not be treated merely as an efficiency tool, but as part of a broader ecology of support that includes teachers, peers, tasks, and assessment practices. For future research, it would be valuable to implement similar interventions with control groups, longer treatment durations, and delayed post-tests to examine the sustainability of gains over time. Qualitative analyses of students' drafts, revision histories, and reflective interviews could also deepen understanding of how learners interact with AI at the level of argumentation and critical thinking. In this way, the present study provides preliminary evidence that may inform the development of AI-enhanced academic writing instruction in university EFL contexts.

CONCLUSION

The findings of this study indicate that students' academic writing performance improved following the implementation of AI-assisted writing instruction. The quantitative results showed an increase in the mean score from 64.92 on the pre-test to 82.08 on the post-test, while the Wilcoxon Signed-Rank Test ($Z = -4.374$, $p < .001$) indicated that the improvement was statistically significant. Students also reported generally positive perceptions of Generative AI, particularly in supporting idea generation, text organization, and confidence during the writing process.

Although these findings suggest that AI-assisted writing activities may be beneficial in EFL academic writing instruction, they should be interpreted within the limitations of the study. Because this research employed a pre-experimental one-group pretest–posttest design without a control group, the observed improvements cannot be attributed solely to the AI intervention or interpreted as evidence of a causal relationship. In addition, the relatively small sample drawn from a single institution and the short intervention period limit the generalizability of the findings.

Overall, this study provides preliminary empirical evidence that AI-assisted writing instruction may support students' academic writing development in Indonesian EFL higher education. Future research employing more rigorous experimental designs, larger and more diverse samples, and longer intervention periods is needed to provide stronger evidence regarding the effectiveness of Generative AI in academic writing instruction.

THANKS

The authors would like to express their sincere gratitude to the Department of English Education, Universitas Muhammadiyah Mataram, for its academic support throughout this research. The authors also thank all students who voluntarily participated in this study for their valuable cooperation and contribution. All participants were informed about the purpose of the study and voluntarily agreed to participate. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declare that they have no conflicts of interest.

BIBLIOGRAPHY

Chung, H. Q., Chen, V., & Olson, C. B. (2021). The impact of self-assessment, planning and goal setting, and reflection before and after revision on student self-efficacy and writing performance. *Reading and Writing*, *34*(7), 1885–1913. https://doi.org/https://doi.org/10.1007/s11145-021-10186-x

Darwin, Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, *11*(1), 2290342. https://doi.org/https://doi.org/10.1080/2331186X.2023.2290342

Farrokhnia, M., Latifi, S., Papadopoulos, P. M., Hogenkamp, L., Gijlers, H., Khosravi, H., & Noroozi, O. (2026). Generative AI offers more, but students revise less: comparing the effects of teacher and AI feedback on student essay revisions. *International Journal of Educational Technology in Higher Education*, *23*(1), 6. https://doi.org/https://doi.org/10.1186/s41239-026-00579-9

Feng, M., & Mei, T. (2022). Metacognitive writing strategies, critical thinking skills, and academic writing performance : A structural equation modeling approach. *Metacognition and Learning*, *0123456789*. https://doi.org/10.1007/s11409-022-09328-5

Jmaiel, H. A., Abukhait, R. O., Mohamed, A. M., Shaaban, T. S., Al-khresheh, M. H., & AL-Qadri, A. H. (2025). The role of ChatGPT in enhancing EFL students' ESP writing skills: an experimental study of gender and major differences. *Discover Education*, *4*(1), 240. https://doi.org/https://doi.org/10.1007/s44217-025-00700-6

Khan, R., Qamar, M. T., Ansari, M. S., & Yasmeen, J. (2025). Enhancing or impairing? Exploring Indian EFL learners' academic writing narratives with ChatGPT. *Cogent Education*, *12*(1), 2514329. https://doi.org/https://doi.org/10.1080/2331186X.2025.2514329

Khojasteh, L., Kafipour, R., Pakdel, F., & Mukundan, J. (2025). Empowering medical students with AI writing co-pilots: design and validation of AI self-assessment toolkit. *BMC Medical Education*, *25*(1), 159. https://doi.org/https://doi.org/10.1186/s12909-025-06753-3

Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies*, *30*(1), 1265–1300. https://doi.org/https://doi.org/10.1007/s10639-024-12878-7

Krause, S., Panchal, B. H., & Ubhe, N. (2025). Evolution of learning: Assessing the transformative impact of generative AI on higher education. *Frontiers of Digital Education*, *2*(2), 21. https://doi.org/https://doi.org/10.1007/s44366-025-0058-7

Mason, S., & Merga, M. (2022). Communicating research in academia and beyond: Sources of self-efficacy for early career researchers. *Higher Education Research & Development*,

41(6), 2006–2019. https://doi.org/https://doi.org/10.1080/07294360.2021.1945545

Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: a study on proficiency, revision frequency and writing quality. *Discover Education*, *4*(1), 170. https://doi.org/https://doi.org/10.1007/s44217-025-00602-7

Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, 1–16. https://doi.org/https://doi.org/10.1007/s11528-025-01100-1

Wale, B. D., & Bogale, Y. N. (2021). Using inquiry-based writing instruction to develop students' academic writing skills. *Asian-Pacific Journal of Second and Foreign Language Education*, *6*(1), 4. https://doi.org/https://doi.org/10.1186/s40862-020-00108-9