



The Effect of Artificial Intelligence Tools on EFL Students' Writing Skills: A Mixed-Methods Study at Universitas Panca Sakti Bekasi

Subadar^{1*}, Eryandi², Deka Meuthia Novari³

^{1,2}Universitas Panca Sakti Bekasi, Jl. Raya Hankam No.54, RT.005/RW.002, Jatirahayu, Kec. Pd. Melati, Kota Bekasi, Jawa Barat, Indonesia 17414.

³Institut Al-Ma'arif Way Kanan, Jl. Veteran No. 03, Tiuh Balak / Bhakti Negara, Kecamatan Baradatu, Kabupaten Way Kanan, Provinsi Lampung, Indonesia 34761.

Email Correspondence: badarsetiko@gmail.com

Abstract

Although artificial intelligence (AI) has increasingly been incorporated into EFL writing instruction, previous research has largely focused on overall writing performance, with less attention to changes across specific writing dimensions and students' actual experiences of using AI during the writing process. This study examined the effect of AI-assisted instruction on EFL students' writing performance, with particular attention to five writing dimensions content, organization, vocabulary, language use, and mechanics and explored how students used and experienced AI during writing. Using a quasi-experimental explanatory sequential mixed-methods design, the study involved 60 English Language Education students at Universitas Panca Sakti Bekasi, equally divided into experimental and control groups. The experimental group engaged in AI-assisted writing activities integrated into regular writing tasks, while the control group received conventional instruction. Pre- and post-test writing scores were analyzed using descriptive statistics and ANCOVA, complemented by semi-structured interviews and AI Usage Logs. The experimental group demonstrated significantly greater improvement in overall writing performance than the control group ($F = 381.24$, $p < .001$, partial $\eta^2 = .870$), with significant differences across all five dimensions. Language use showed the largest raw gain, whereas mechanics yielded the largest effect size. AI Usage Logs further showed that students primarily used AI for language feedback, idea generation, vocabulary enrichment, and revision. The findings indicate a substantial effect within the specific instructional context and demonstrate the value of examining writing outcomes alongside students' actual AI-use practices.

Keywords: AI Tools; Chat GPT; English; Writing.

How to Cite: Subadar, S., Eryandi, E., & Novari, D. M. (2026). The Effect of Artificial Intelligence Tools on EFL Students' Writing Skills: A Mixed-Methods Study at Universitas Panca Sakti Bekasi. *Empiricism Journal*, 7(3), 3265-3274. <https://doi.org/10.36312/ej.v7i3.6699>



<https://doi.org/10.36312/ej.v7i3.6699>

Copyright© 2026, Subadar et al.

This is an open-access article under the CC-BY-SA License.



INTRODUCTION

Learners of English as a Foreign Language (EFL) are expected to put across ideas that are clear, accurate, and purposeful, which is why writing sits at the heart of EFL education. Several competencies must operate simultaneously for writing to succeed: generating ideas, ordering information coherently, choosing suitable words, building sentences that are grammatically acceptable, and respecting the conventions of written text. As these competencies constantly influence one another, writing is more accurately described as a complex and multidimensional skill than as a matter of turning out sentences without errors. The skill carries added weight in higher education, where students are regularly called on to put forward arguments, clarify concepts, condense information, and share academic knowledge in writing. Writing instruction should accordingly let students draft, assess, and revise their own texts while they receive constructive feedback.

The wider availability of artificial intelligence (AI) has brought additional ways of helping writers. Such tools can support grammar checking, choice of vocabulary, sentence construction, organization, and revision. In second-language writing, automated evaluation and written feedback have taken on greater importance, since learners can get responses again and again without a teacher having to address each linguistic problem individually. According to research on automated writing evaluation, these technologies do more than assess: they can also foster writing development and learning in the classroom, though learners' reactions can differ depending on personal and contextual factors (Karatay &

Karatay, 2024; Shi & Aryadoust, 2024). This is why AI-supported writing now draws considerable attention in EFL research.

Several studies have reported potential benefits. Fan (2023) applied a mixed-methods design to automated written corrective feedback from Grammarly and reported that EFL writing quality improved. After students used Wordtune, Rad et al. (2024) saw favorable changes in writing outcomes, engagement, and feedback literacy. Research on generative AI points the same way. Song and Song (2023) found that ChatGPT-supported instruction lifted EFL students' writing in content, organization, and language use; Mahapatra (2024) reported a significant positive effect of ChatGPT on undergraduate academic writing in a mixed-methods intervention; and Poláková and Ivenz (2024) noted improvements in various aspects of EFL writing once students were given feedback produced by ChatGPT. Read together, these studies indicate that AI is more than a text generator; it can also offer support and feedback at different points of the writing process.

Still, one overall writing score cannot capture everything AI contributes. Writing quality is made up of several components, and AI's influence may not be equal across them. Marzuki et al. (2023), for example, reported that EFL teachers at Indonesian universities judged AI writing tools to be particularly helpful for content and organization, which call for higher-level decisions than grammatical correction does. Findings on automated feedback also suggest that learners more readily act on explicit corrections than on wider suggestions that require interpretation and revision (Yang et al., 2025). To find out whether AI-assisted instruction brings consistent improvement across writing proficiency, content, organization, vocabulary, language use, and mechanics therefore need to be examined one by one.

The way students engage with AI may also affect how much educational value it holds. AI-generated feedback does not lead to learning on its own; learners must make sense of the suggestions, decide whether they are suitable, and select which recommendations to use. Xiao and Zhi (2023) observed that EFL learners considered ChatGPT useful for language-learning tasks but remained aware of what it cannot do. Woo et al. (2024) used a mixed-methods study to relate students' experiences of writing with ChatGPT to motivation, cognitive load, and satisfaction. Meniado et al. (2024) also found mostly positive perceptions among EFL learners in Thailand and Vietnam, notably about brainstorming, organizing ideas, developing outlines, and editing drafts. These results show why students' perceptions and interaction patterns need to be taken into account when the role of AI in EFL writing is studied.

The benefits are not without concerns, however, when AI enters academic writing. Barrot (2023) highlighted problems concerning the accuracy of generated information, response quality, pedagogical implications, and academic integrity, even while acknowledging that ChatGPT can provide adaptive support. Related concerns have been raised elsewhere, including overreliance, the reliability of suggestions produced by AI, and whether students still feel that their writing is their own. Teng (2024), reviewing the literature on ChatGPT in EFL writing systematically, argued that generative AI's opportunities must be considered together with pedagogical and ethical issues. Hence, what matters is not just whether students use AI but also how they apply what it suggests.

Despite the steady growth of research on AI-assisted writing, some important gaps remain. Previous studies have frequently centered on a single AI application, a single kind of automated feedback, or students' general perceptions; comparatively few have considered several writing dimensions together with learners' qualitative experiences. In Indonesian higher education this gap is particularly relevant, because AI writing tools are becoming easier to access while empirical evidence on how they affect specific components of EFL writing is still limited. Jaya and Susyla (2024) found that Indonesian students considered AI tools useful for diverse writing activities, yet they also recognized shortcomings in coherence, creativity, critical thinking, and dependence. Such observations lend further support to research that addresses overall writing performance, individual writing dimensions, and students' experiences of AI-assisted writing.

Given this context, the present study looks at how artificial intelligence tools affect the writing skills of EFL students at Universitas Panca Sakti Bekasi, employing a mixed-methods design. In its quantitative phase, the study checks whether AI-assisted instruction produces significant changes in five dimensions: content, organization, vocabulary, language use, and mechanics. In its qualitative phase, it explores how students experience and perceive the use

of AI in supporting their writing development. Using both kinds of evidence allows measurable performance changes to be examined alongside what students experienced with the technology as they learned.

The study thus has three objectives: (1) to determine if AI tools significantly improve EFL students' writing skills; (2) to look into the degree to which AI-assisted writing influences content, organization, vocabulary, language use, and mechanics; and (3) to explore how EFL students experience and perceive the use of AI tools for developing their writing skills.

METHOD

An explanatory sequential mixed-methods design was adopted to explore how artificial intelligence (AI) tools contribute to the writing development of EFL students at Universitas Panca Sakti Bekasi. Two linked stages made up the research. In the first, quantitative data were used to establish whether AI-assisted instruction was associated with changes in writing performance and to evaluate five writing dimensions: content, organization, vocabulary, language use, and mechanics. In the second, qualitative stage, students' experiences and perceptions of using AI were explored so as to account for the patterns seen in the quantitative findings. The design suited the study because it sought to measure change in writing performance and, at the same time, to understand students' experience of AI during writing activities.

Sixty undergraduate students from the English Language Education program at Universitas Panca Sakti Bekasi took part, thirty of them in the experimental group and thirty in the control group. Since the participants came from existing university classes and assigning individuals at random was impractical, a quasi-experimental design was used. AI-assisted writing instruction was given to the experimental group; the control group was taught writing in the conventional way, without AI support. A writing pre-test was administered to both groups before the intervention to set baseline performance and check that the groups were comparable.

Tests of EFL writing, given before and after the intervention, served to assess writing performance. Students wrote argumentative essays of about 300-400 words on topics appropriate for their academic level. An analytic rubric was used to score the texts. It covered five dimensions (content, organization, vocabulary, language use, and mechanics), each rated on a five-point scale and counted equally toward the overall score. Throughout the intervention, students in the experimental group relied on AI as a supporting resource to generate ideas, organize content, develop vocabulary, receive grammar and language feedback, and revise drafts. They also completed an AI Usage Log in which they noted the tool used, why they used it each time, and how they built AI-generated suggestions into their writing.

Both descriptive and inferential procedures made up the quantitative analysis. For each group, means and standard deviations were calculated on pre-test and post-test scores. Normality and homogeneity assumptions were examined before any inferential test was run. The primary analysis was ANCOVA, in which post-test writing performance was the dependent variable and pre-test performance was the covariate. To locate differences in improvement, each of the five dimensions was analyzed separately as well. Effect sizes were reported too, to indicate how large the group differences were in practical terms and not only whether they were statistically significant.

Once the quantitative phase was complete, semi-structured interviews were held with 12 students from the experimental group to collect qualitative data. Purposive selection was used so that participants would represent low, moderate, and high levels of writing improvement. Topics in the interviews included students' experiences with AI, the benefits they perceived, the aspects of writing where AI helped most, the way they used AI feedback when revising, the difficulties they encountered, and how they saw their own ability to write independently. Thematic analysis of the interview data followed these steps: familiarization, initial coding, theme development, review, and interpretation. Afterward, the qualitative findings were integrated with the quantitative results in order to explain more fully how and why AI-assisted writing was related to changes in students' performance.

RESULTS AND DISCUSSION

Result

Quantitative Results

This phase of the quantitative work looked at whether AI-assisted instruction was associated with changes in overall EFL writing performance, and whether the pattern differed among content, organization, vocabulary, language use, and mechanics. The analysis covered 60 students: 30 in the experimental group and 30 in the control group.

Average writing scores were fairly close between the groups at the start. Pre-test means were 63.18 (SD = 3.23) for the experimental group and 62.38 (SD = 2.53) for the control group. After the intervention, the experimental group reached a post-test mean of 79.91 (SD = 3.33) while the control group reached 69.53 (SD = 2.64). On average, therefore, the experimental group gained 16.06 points (SD = 2.08), over twice the control group's gain of 7.15 points (SD = 1.35).

Table 1. Experiment and control gorup result

Group	N	Pre-test M (SD)	Post-test M (SD)	Gain M (SD)
Experimental	30	63.18 (3.23)	79.91 (3.33)	16.06 (2.08)
Control	30	62.38 (2.53)	69.53 (2.64)	7.15 (1.35)

ANCOVA assumptions were tested. Residuals were normally distributed according to the Shapiro-Wilk test ($W = .990$, $p = .915$), and the assumption of homogeneous regression slopes was met, given that the group $\times$ pre-test interaction was non-significant, $F(1, 56) = 0.539$, $p = .466$. With these results, ANCOVA could appropriately be used to compare post-test performance while holding initial ability constant.

With pre-test scores held constant, ANCOVA found a statistically significant group effect on post-test writing performance, $F(1, 57) = 381.24$, $p < .001$, partial $\eta^2 = .870$. Pre-test score was a significant covariate as well, $F(1, 57) = 139.73$, $p < .001$. In the adjusted analysis, students who received AI-assisted instruction therefore held a substantial advantage.

Improvement Across the Five Writing Dimensions

Each writing dimension (content, organization, vocabulary, language use, and mechanics) was looked at on its own. In descriptive terms, the experimental group improved on all five. The mean increase was highest for language use, then vocabulary, organization, mechanics, and finally content.

Table 2. Improvement Across the Five Writing Dimensions

Writing Dimension	Experimental Pre	Experimental Post	Gain	Control Pre	Control Post	Gain
Content	12.16	15.10	2.94	12.13	13.22	1.09
Organization	12.95	16.19	3.23	12.75	14.53	1.78
Vocabulary	12.68	15.95	3.27	12.07	13.34	1.27
Language Use	13.00	16.50	3.50	12.60	14.39	1.79
Mechanics	12.40	15.51	3.11	12.84	14.05	1.22

Significant group differences were confirmed by ANCOVA for all five components. The results were as follows: content, $F(1, 57) = 65.59$, $p < .001$, partial $\eta^2 = .535$; organization, $F(1, 57) = 38.95$, $p < .001$, partial $\eta^2 = .406$; vocabulary, $F(1, 57) = 49.18$, $p < .001$, partial $\eta^2 = .463$; language use, $F(1, 57) = 47.10$, $p < .001$, partial $\eta^2 = .452$. Mechanics had the largest adjusted effect, $F(1, 57) = 73.40$, $p < .001$, partial $\eta^2 = .563$.

Table 3. The analysis significant difference

Dimension	F(1,57)	p	Partial η^2
Content	65.59	< .001	.535
Organization	38.95	< .001	.406
Vocabulary	49.18	< .001	.463
Language Use	47.10	< .001	.452
Mechanics	73.40	< .001	.563

To summarize, students taught with AI assistance improved significantly on all five dimensions. Unadjusted gains were largest for language use (3.50 points) and smallest for

content (2.94 points). With initial performance controlled, though, the largest adjusted effect belonged to mechanics (partial $\eta^2 = .563$), followed by content (.535), vocabulary (.463), language use (.452), and organization (.406).

Patterns of AI Use

Information on how the experimental group used AI tools came from the AI Usage Log. The 30 students averaged 6.63 sessions (SD = 1.67), with a range of four to nine. Around 64.83% (SD = 14.92) of AI-generated suggestions were accepted on average, and individual acceptance rates fell between 41% and 85%.

Grammar/language feedback was the leading purpose (14 students, 46.7%). Brainstorming and vocabulary development each came next, with six students (20.0%) apiece. Revision/editing was the main use for three students (10.0%), and organization/outlining for one (3.3%).

Twelve students (40.0%) were categorized as highly engaged, 11 (36.7%) as moderately engaged, and seven (23.3%) as having low engagement. From these patterns it appears that students mainly used AI as a resource for language support and revision instead of for producing entire texts.

No significant correlation with writing gain was found for either frequency of AI use ($r = .153$, $p = .419$) or proportion of accepted suggestions ($r = -.095$, $p = .617$). In this sample, using AI more often or accepting more suggestions thus did not necessarily go with larger score gains, which suggests that how well students engage may matter more than how much.

Selection of Participants for the Qualitative Phase

Interviews were held with twelve experimental-group students, chosen from predetermined improvement strata of four low, four moderate, and four high. Writing gains among them ranged from 12.6 to 20.2 points.

Table 4. Qualitative phase

Improvement Stratum	n	Gain Range
Low improvement	4	12.6–12.8
Moderate improvement	4	15.1–15.6
High improvement	4	18.5–20.2

Because of this stratified selection, students at different levels of improvement were represented, which allowed the interviews to explore possible connections between experiences, patterns of AI use, perceived benefits, and the variation seen in the quantitative data.

Discussion

Overall writing performance among EFL students improved clearly when instruction was supported by artificial intelligence. Whereas the control group advanced only from 62.38 to 69.53, the experimental group moved from a pre-test mean of 63.18 to 79.91. What matters more is that, with pre-test scores controlled, ANCOVA revealed a statistically significant group effect, $F(1, 57) = 381.24$, $p < .001$, partial $\eta^2 = .870$, and a magnitude like this indicates the difference between groups was practically meaningful, not merely statistical. Other recent work points the same direction, showing that AI-supported writing interventions can lead to substantial gains in EFL learners' writing ability. Yin et al. (2025), for instance, found that a ChatGPT-based, AI-assisted critical-thinking program improved undergraduate EFL students' writing proficiency; another recent study found significantly higher post-test scores for learners receiving generative AI feedback than for those taught by conventional methods (Wang et al., 2025); and in interventions combining ChatGPT with teacher feedback, the experimental group outperformed students given only traditional feedback (Liu et al., 2025). The present findings thus strengthen the growing empirical argument that AI is effective instructional support when placed within a structured writing process and not used only to generate text.

One way to understand the experimental group's marked progress is through the interactive quality of AI-assisted feedback. Students used AI in the intervention to generate ideas, organize content, expand vocabulary, get language feedback, and revise drafts; these varied uses probably offered repeated opportunities for them to notice weaknesses and improve their texts. Part of generative AI's value in second-language writing, according to

current research, comes from its interactive and iterative nature. Learners are not restricted to feedback delivered once a task has ended; they can ask follow-up questions, seek clarification, compare options, and revise over successive exchanges. Zhang and Liu (2026) argue, for example, that ChatGPT-mediated writing activities work best when they encourage cognitive engagement, self-assessment, and a critical stance toward feedback generated by AI. Studies of second-language writers' engagement with ChatGPT feedback point similarly to learners who actively accept, reject, and negotiate suggestions and do not simply adopt every recommendation without question (Zhang & Saeed, 2025). It follows that the gains found here may be due less to AI being available than to students having the chance to treat it as an interactive resource at several stages of writing.

The influence of AI support on the five writing dimensions is the second key result. Between-group differences were statistically significant in content, organization, vocabulary, language use, and mechanics alike, which is noteworthy since it indicates that AI-assisted writing benefited more than surface-level language correction. Within the experimental group, unadjusted gains were greatest in language use (+3.50) and then, in order, vocabulary (+3.27), organization (+3.23), mechanics (+3.11), and content (+2.94). One possible reading is that AI was most useful in areas where students could respond directly to linguistic suggestions and to alternative wordings of their ideas. Recent research supports a similar reading. Pratiwig and colleagues (2024) noted that Indonesian EFL students' interaction with ChatGPT could potentially improve lexical resources, grammatical range and accuracy, and coherence and cohesion, and Poláková and Ivenz (2024) documented progress in several facets of EFL writing after students got ChatGPT-generated feedback. Within a quasi-experimental comparison of AI-assisted and conventional instruction, the current study adds to this earlier evidence by showing improvement in five analytically distinct dimensions at once.

Comparing unadjusted gains with ANCOVA effect sizes, however, shows a contrasting pattern. Language use had the biggest raw increase, yet the biggest adjusted effect size was for mechanics (partial $\eta^2 = .563$), ahead of content (.535), vocabulary (.463), language use (.452), and organization (.406). The mismatch means that mean gains alone should not be used to judge how large an improvement is. A possible reason why mechanics responds so well to AI support is that spelling, punctuation, grammatical conventions, and other formal features are comparatively clear-cut, so AI systems can detect them efficiently. Content and organization, on the other hand, involve more intricate rhetorical choices, and students may need to judge relevance, logical flow, and how an argument develops. Recent investigations of ChatGPT feedback support this contrast: learners usually respond more readily to language-related suggestions, while responses to feedback on organization are more varied, being accepted in some cases and rejected in others (Zhang & Saeed, 2025). Kang (2025) studied direct and reformulation feedback given through ChatGPT and likewise found gains in the same five writing dimensions used in this study, although effects depended on feedback type and revision stage. Taken together, the results imply that AI can be very effective with clearly identifiable linguistic problems, whereas higher-order aspects of writing may require more deliberate judgment from learners and more instructional support.

That content and organization improved is particularly significant, given that these dimensions cannot easily be improved just by correcting errors. Among experimental-group students, content went up by 2.94 points and organization by 3.23, which suggests that AI may have supported higher-order elements of composition as well as linguistic accuracy. Students were directed to use AI for brainstorming and outlining before and during writing, and this may plausibly explain the result. Such use differs from asking an AI system to write a complete essay, because students still decide which ideas to keep, how to arrange information, and how to turn suggestions into their own writing. Meniado et al. (2024) showed that EFL students often turn to generative AI to brainstorm, organize ideas, refine outlines, and edit drafts, and research with Indonesian EFL learners similarly indicates that ChatGPT can improve coherence and cohesion when students engage actively with it (Pratiwig et al., 2024). The content and organization gains recorded here may therefore reflect the scaffolding role of AI in planning and revision, not AI-generated text directly replacing students' own writing.

How the intervention operated is clarified further by the AI Usage Log. Students reported using AI in 6.63 sessions on average, and grammar and language feedback was cited most often as the purpose (46.7%), with brainstorming and vocabulary development next (20.0% each). This distribution suggests that students used AI mainly as a source of support and feedback and not chiefly to produce complete texts, which agrees with recent findings that EFL learners can exercise agency in interacting with generative AI. Zhang and Saeed (2025), for example, found that learners mostly used ChatGPT to polish their academic writing while actively questioning and negotiating the feedback it provided. Translanguaging research also suggests that EFL learners may use generative AI in an iterative problem-solving process, reshaping and refining ideas before they reach the final English text (Zhou et al., 2025). Read in light of these observations, the students' AI Usage Logs indicate that AI functioned as a mediating resource in the writing process, especially for solving linguistic problems, and did not merely take the place of students' own composing.

The positive outcome of the intervention needs one important qualification: frequency of AI use was not significantly correlated with writing gain ($r = .153$, $p = .419$), and neither was the proportion of AI suggestions accepted ($r = -.095$, $p = .617$). Frequent AI use, in other words, did not automatically come with greater individual improvement. This carries theoretical weight because it challenges the assumption that access to AI, or the sheer amount of interaction with it, is enough to explain writing development. What learners gain from AI may depend more on how they assess and apply its feedback than on how many times they use it. Recent studies of ChatGPT feedback also show that students do not respond to AI suggestions in a uniform way; feedback on content and language may be accepted fairly readily, but organizational feedback may be accepted or rejected selectively (Zhang & Saeed, 2025). Research on AI-mediated writing likewise emphasizes critical engagement, self-evaluation, and feedback literacy over passive acceptance of AI recommendations (Zhang & Liu, 2026). The non-significant correlations found here should not, therefore, be interpreted as showing that engagement with AI does not matter; they instead suggest that the quality of engagement may be more important than its quantity.

Lastly, the results bear on how AI can be used in practice in EFL writing instruction. Given the very large overall ANCOVA effect and the significant gains in all five writing dimensions, AI can be integrated productively into university-level writing courses. The findings do not, however, justify substituting AI for teacher guidance. Recent research underlines that AI-generated feedback, though prompt, adaptive, and personalized, still has limitations in interpreting context, exercising rhetorical judgment, and supporting advanced writing skills. Nguyen (2026), for instance, notes that AI feedback is especially promising for formative writing support but less dependable for complex rhetorical and culturally embedded aspects of writing. Evidence on AI-assisted teacher feedback also indicates that pairing AI's capabilities with teachers' expertise can improve writing outcomes and lessen problems related to contextual understanding and overreliance (Liu et al., 2025). The contribution of this study is therefore not a claim that AI outperforms teachers in every aspect of writing instruction. It shows, instead, that structured AI-assisted writing can raise students' writing performance considerably when AI serves as a scaffold for planning, feedback, and revision. In future practice, priority should be given to AI literacy, critical evaluation of AI suggestions, and ongoing teacher involvement, so that improvements in writing accuracy and performance are matched by students' development into more independent writers.

CONCLUSION

The conclusion of this study is that integrating artificial intelligence (AI) tools significantly improved the writing skills of EFL students at Universitas Panca Sakti Bekasi. Evidence for this comes from the experimental group's considerably higher gains and from the very large ANCOVA effect size. AI-supported writing instruction also had a positive influence on each of the five dimensions examined: content, organization, vocabulary, language use, and mechanics. Improvement varied in degree; language use recorded the highest unadjusted mean gain, whereas mechanics produced the strongest adjusted effect size. The AI Usage Log indicates that students mostly used AI as a supporting resource for grammar and language feedback, idea generation, vocabulary expansion, and revision, and not primarily to generate complete texts. Moreover, no significant relationships were found between writing

improvement and either the frequency of AI use or the rate at which AI suggestions were accepted, which suggests that more AI interaction does not by itself lead to larger writing gains. In sum, the results suggest that AI tools can serve effectively as a pedagogical scaffold in EFL writing instruction if they are embedded in organized stages of planning, drafting, feedback, and revision. To make sure AI assists and does not replace the development of independent writing competence, teachers must keep providing guidance and students must engage critically.

RECOMMENDATION

Future research is recommended to expand the investigation of the effectiveness of Artificial Intelligence (AI) in enhancing English as a Foreign Language (EFL) students' writing skills by employing more rigorous experimental designs, involving larger sample sizes, and including participants from multiple universities with diverse linguistic backgrounds and proficiency levels. Longitudinal studies are also necessary to examine the sustainability of writing improvements after AI-assisted interventions have been discontinued, particularly across the five writing dimensions: content, organization, vocabulary, language use, and mechanics. Given that the present study found no statistically significant correlations between writing improvement and either the frequency of AI use or the acceptance rate of AI-generated suggestions, future investigations should prioritize the quality of students' interactions with AI, including their ability to critically evaluate feedback, independently revise their writing, and maintain originality. Furthermore, comparative studies should examine the effectiveness of different AI platforms, feedback types, and instructional approaches, including independent AI use, teacher-provided feedback, and integrated AI–teacher feedback, to identify the most effective strategies for improving specific writing dimensions. Future research should also incorporate AI literacy, feedback literacy, critical thinking, writing self-efficacy, and learner autonomy as potential mediating or moderating variables influencing the relationship between AI-assisted instruction and writing performance. To obtain more comprehensive evidence, subsequent studies should integrate AI interaction logs, longitudinal analyses of writing drafts, in-depth interviews, and independent writing assessments conducted without AI assistance to distinguish improvements in AI-supported writing products from the development of autonomous writing competence. Additionally, future investigations should address academic integrity, AI overreliance, intellectual ownership, and inter-rater reliability in writing assessment. Ultimately, these research directions are expected to provide stronger empirical evidence regarding the mechanisms, long-term effectiveness, and pedagogical conditions under which AI can function as an effective instructional scaffold while preserving students' independence, creativity, critical thinking, and authentic academic writing abilities.

REFERENCES

- Agustin, N. Z., Cahyono, B. Y., & Agustina, F. R. (2025). College EFL students' perception toward AI assistance in academic writing assignments. *ELT Forum: Journal of English Language Teaching*, 14(4). <https://doi.org/10.15294/elt.v14i3.23636>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Fan, N. (2023). Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study. *SAGE Open*, 13(3), 1–15. <https://doi.org/10.1177/21582440231181296>
- Jaya, S., & Susyla, D. (2024). Assessing students' perceptions of AI grammar and writing assistance tools: Implications for academic writing instruction. *Teaching English and Language Learning English Journal*, 4(3).
- Kang, D. (2025). Direct vs. reformulation feedback through ChatGPT on EFL learners' writings. *Journal of Korean English Education Society*, 24(3), 1–17. <https://doi.org/10.18649/jkees.2025.24.3.1>
- Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System*, 123, 103332. <https://doi.org/10.1016/j.system.2024.103332>

- Liu, Y., et al. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity*, 59, 101766. <https://doi.org/10.1016/j.tsc.2025.101766>
- Maghfiroh, T. G., & Akhiriyah, S. (2025). EFL learners' perception of using ChatGPT as a supporting tool for academic writing: Benefits and challenges. *TOFEDU: The Future of Education Journal*, 4(8). <https://doi.org/10.61445/tofedu.v4i8.994>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2), 2236469. <https://doi.org/10.1080/2331186X.2023.2236469>
- Meniado, J. C., Huyen, D. T. T., Panyadilokpong, N., & Lertkomolwit, P. (2024). Using ChatGPT for second language writing: Experiences and perceptions of EFL learners in Thailand and Vietnam. *Computers and Education: Artificial Intelligence*, 7, 100313. <https://doi.org/10.1016/j.caeai.2024.100313>
- Mun, C. (2024). EFL learners' English writing feedback and their perception of using ChatGPT. *STEM Journal*, 25(2), 26–39. <https://doi.org/10.16875/stem.2024.25.2.26>
- Nguyen, D. M. Q. (2026). The role of ChatGPT in providing written corrective feedback for EFL academic writing: A literature review. *International Journal of AI in Language Education*. <https://doi.org/10.54855/ijaile.26313>
- Poláková, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1), 2410101. <https://doi.org/10.1080/2331186X.2024.2410101>
- Prasetya, R., Syarif, A., & Mayof, A. (2026). Comparative analysis of ChatGPT and Grammarly in supporting grammar correction and writing development in EFL students. *Journal of English and Education*, 12(1). <https://doi.org/10.20885/jee.v12i1.44160>
- Pratiwig, A. C., et al. (2024). ChatGPT affordances and Indonesian EFL students' perceptions in L2 writing: A collaborative reflexive thematic analysis. *Changing English*, 32(2), 195–211. <https://doi.org/10.1080/1358684X.2024.2418132>
- Rad, H. S., Alipour, R., & Jafarpour, A. (2024). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, 32(9), 5020–5040. <https://doi.org/10.1080/10494820.2023.2208170>
- Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, 36(2), 187–209. <https://doi.org/10.1017/S0958344023000265>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Suryanto, S., Islami, N., Alsolami, T., & Shinjee, B. (2026). AI-assisted writing in EFL education: Students' experiences with QuillBot. *Teaching English as a Foreign Language Journal*, 5(1). <https://doi.org/10.12928/tefl.v5i1.2120>
- Teng, M. F. (2024). A systematic review of ChatGPT for English as a Foreign Language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, 6(3), 36–57. <https://doi.org/10.58304/ijts.20240304>
- Wang, et al. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. *Discover Education*, 4, 170. <https://doi.org/10.1007/s44217-025-00602-7>
- Woo, D. J., Wang, D., Guo, K., & Susanto, H. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. *Education and Information Technologies*, 29(18), 24963–24990. <https://doi.org/10.1007/s10639-024-12819-4>
- Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3), 212. <https://doi.org/10.3390/languages8030212>

- Yang, Y., Hussin, S., & Hashim, H. (2025). Revising with intelligence: ChatGPT feedback and its impact on EFL students' revision and self-efficacy. *Forum for Linguistic Studies*, 7(7). <https://doi.org/10.30564/fls.v7i7.9845>
- Yin, X., Dou, K., & Han, J. (2025). An AI-assisted critical thinking intervention to enhance undergraduate EFL learners' writing proficiency. *Studies in Educational Evaluation*, 86, 101480. <https://doi.org/10.1016/j.stueduc.2025.101480>
- Zhang, K., & Saeed, M. A. (2025). Chinese EFL learners' engagement with ChatGPT feedback on academic writing: A case study in Malaysia. *Computers and Composition*, 78, 102976. <https://doi.org/10.1016/j.compcom.2025.102976>
- Zhang, Y., & Liu, Y. (2026). Designing ChatGPT-mediated feedback activities in EFL writing: A design-based study of the dialogic feedback triangle. *Assessment & Evaluation in Higher Education*, 51(1), 137–160. <https://doi.org/10.1080/02602938.2025.2571846>
- Zhou, et al. (2025). Translanguaging with generative AI in EFL writing: Students' practices and perceptions. *Journal of Second Language Writing*, 67, 101181. <https://doi.org/10.1016/j.jslw.2025.101181>