



The International Journal of Language Studies

ISSN: 3078-2244 (Online)

2026, Vol. 3, No. 2, pp. 55–68

DOI: <https://doi.org/10.60087/ijls.v3.n2.004>

Journal homepage: <https://ijlangstudies.org>



The Impact of Next-Generation AI Chatbots on Academic Writing Skills and Academic Integrity Among English-Majored Students at HUIT: An Experimental Study

Nguyen Minh Nhat, Nguyen Thao My, Tran Thi Thuan, Le Thi Quynh Nhu

Faculty of Foreign Languages, Ho Chi Minh City University of Industry and Trade (HUIT), Vietnam

*Corresponding author: Nguyen Minh Nhat (minhnhat.9a7.2013@gmail.com)

Abstract

This study examines the impact of next-generation AI chatbots on the academic writing performance of English-majored students at Ho Chi Minh City University of Industry and Trade (HUIT), situating the findings within broader questions of academic integrity raised by AI-assisted writing. Using a quasi-experimental pretest-posttest design, 287 students completed academic writing tasks before and after an AI-assisted writing intervention; 150 students received AI assistance during the intervention while 137 completed the task without AI support. Writing performance was scored using a five-dimension analytic rubric (organization, language accuracy, syntactic complexity, coherence and cohesion, and critical thinking). Results showed a statistically significant improvement from pretest ($M = 6.64$) to posttest ($M = 8.07$), $t(286) = -51.16$, $p < .001$. Students who received AI assistance achieved significantly higher posttest scores ($M = 8.33$) than those who did not ($M = 7.79$), $t(285) = -5.46$, $p < .001$. After controlling for pretest scores, ANCOVA confirmed a large effect of AI assistance on posttest performance, $F(1, 284) = 218.70$, $p < .001$, partial $\eta^2 = .435$. These findings indicate that AI-assisted writing produced substantial, statistically robust gains in students' writing performance beyond what pre-existing ability could explain. The discussion situates these performance gains within ongoing debates about authorship, originality, and academic integrity in AI-assisted academic work, arguing that improvements in measurable writing quality do not by themselves resolve, and may in fact sharpen, questions about what should count as a student's own intellectual contribution. Implications for students, lecturers, and institutional assessment policy are discussed.

Keywords: AI chatbots, Academic writing performance, ANCOVA, Academic integrity, English as a foreign language, Higher education

Article information: Received: 04/08/2026;

Accepted: 10/08/2026;

Published: 25/06/2026

Copyright © 2026 The Author(s). Published by the International Journal of Language Studies. This is an open-access article distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Introduction

Background of the study

Next-generation AI chatbots such as ChatGPT, Gemini, and Claude have introduced new possibilities for supporting academic writing, offering assistance with idea generation, language revision, and organizational planning at a level of interactivity unavailable in earlier writing-support software (Kasneji et al., 2023). Built on large language models capable of producing coherent, contextually appropriate text, these systems can now engage with writers at nearly

every stage of composition, from brainstorming an argument to polishing its final phrasing (Lo, 2023).

Academic writing is widely regarded as one of the most demanding competencies expected of university students, requiring not only linguistic accuracy but also logical organization, disciplinary awareness, and the capacity for independent critical reasoning (Hyland, 2019). From a cognitive perspective, Flower and Hayes (1981) characterized writing as a recursive process involving planning, translating, and reviewing, rather than a simple linear sequence of steps. For students writing in English as a foreign language, these cognitive demands are often compounded by limitations in vocabulary, grammar, and sentence construction (Leki, 1992), which is precisely where AI chatbots have been proposed as a source of meaningful support.

For English-majored students, whose academic programs place disproportionate weight on writing, these developments raise a fundamental empirical question: do measurable gains in writing performance actually materialize when students write with AI assistance, and if so, how large are those gains once prior writing ability is taken into account? Anecdotal enthusiasm and self-reported perceptions of usefulness, however common in the literature, cannot by themselves answer this question; only a controlled comparison of writing outcomes can.

The problem of the study

Much of the existing literature on AI-assisted writing relies on self-reported perceptions of usefulness rather than direct measurement of writing outcomes (Chan & Hu, 2023). Self-reported benefit is a useful but indirect indicator of impact; it does not establish whether AI assistance produces genuine, measurable improvement in writing quality, nor how large such improvement is relative to what students' pre-existing ability would predict. Without a controlled comparison, it remains unclear whether reported gains reflect the AI tool's actual contribution or simply reflect students' generally favorable attitudes toward new technology.

At the same time, any demonstrated improvement in AI-assisted writing quality raises immediate follow-on questions about authorship and academic integrity: if a text scores substantially higher with AI assistance, whose achievement does that score represent (Cotton et al., 2023)? As generative AI becomes more deeply embedded in students' writing practices, the reliability of a writing score as an indicator of a student's own competence can no longer be assumed by default (Dwivedi et al., 2023). This study addresses the empirical question directly through an experimental design, and uses the results to inform a discussion of the conceptual, integrity-related question that follows from them.

Aim of the study

This study investigates the following research question:

RQ: To what extent does AI-assisted writing influence English-majored students' academic writing performance, and does this influence remain significant after accounting for students' initial writing ability?

In addressing this question, the study also considers what a demonstrated performance effect implies for how academic integrity should be understood and safeguarded once AI assistance becomes part of the writing process.

Significance of the study

For students, the findings clarify what AI assistance can be reasonably expected to contribute to writing performance, helping to set realistic expectations about its role in their own skill development. For lecturers, the results offer an evidence base for deciding how much weight to place on AI-assisted writing scores when assessing genuine student competence, and for designing tasks or rubrics that remain valid under AI-assisted conditions. For institutions, the study offers a concrete empirical anchor, an effect size, for policy discussions about disclosure requirements and assessment reform that would otherwise proceed largely on the basis of anecdote or principle alone.

Scope of the study

This article reports the experimental, performance-measurement component of a larger university-funded research project on next-generation AI chatbots, academic writing, and academic integrity conducted at HUIT. A separate publication from the same project reports a survey-based analysis of students' AI-use patterns, perceived benefits and limitations, academic integrity awareness, and AI literacy (based on a 28-item questionnaire, Q4-Q28); that survey data is not reproduced in this article. The present study is concerned exclusively with the pretest-posttest writing-performance experiment and its implications, and uses those results, rather than survey responses, as the empirical basis for its discussion of academic integrity.

Research context: HUIT and its English-majored program

HUIT's English-majored program requires students to complete a sequence of academic writing courses across their four years of study. Because writing assessment in this program relies heavily on graded written products, often completed outside strictly supervised conditions, the program is particularly exposed to the assessment-validity concerns that motivate this study: if AI assistance can substantially raise a writing score, an unsupervised graded assignment may no longer reliably distinguish a student's own competence from AI-supported output. This institutional context makes HUIT a particularly relevant setting in which to measure, rather than simply theorize about, the size of the AI-assisted writing effect.

Literature Review

Academic writing as a cognitive and social practice

Academic writing has been theorized both as an individual cognitive activity and as a socially situated practice. Flower and Hayes' (1981) cognitive process theory frames writing as a set of recursive activities, planning, translating, and reviewing, that writers move between rather than complete in strict sequence. Complementing this cognitive account, Lea and Street (1998) argue that academic writing is embedded within institutional practices, disciplinary expectations, and power relations, meaning that what counts as competent writing is partly defined by the conventions of a given academic community rather than by linguistic accuracy alone. Hyland (2019) similarly emphasizes that successful academic writing requires disciplinary awareness and rhetorical sensitivity, not merely grammatical correctness. Taken together, these perspectives suggest that any tool that intervenes in the writing process,

including AI chatbots, is likely to affect not just the surface features of a text but also the process through which a writer arrives at it.

Challenges of second-language academic writing

For learners writing in English as a foreign or second language, the cognitive demands of academic writing are frequently compounded by limited vocabulary, grammatical inaccuracy, and difficulty organizing and developing ideas within unfamiliar disciplinary conventions (Leki, 1992). These learners must manage linguistic accuracy and content development simultaneously, a dual burden that can strain working memory and reduce the cognitive resources available for higher-order tasks such as argument construction. This tension between linguistic demands and content demands provides a plausible mechanism by which AI-assisted language support could free up cognitive resources for other aspects of writing, a possibility explored empirically in the present study.

Generative AI chatbots in education

Generative AI chatbots differ from earlier computer-assisted writing tools in their capacity for open-ended, contextually responsive interaction. Rather than simply flagging grammatical errors, systems such as ChatGPT, Gemini, and Claude can generate outlines, suggest alternative phrasing, and produce extended passages of coherent text in response to a prompt (Kasneji et al., 2023). Tlili et al. (2023) argue that such tools support self-directed learning by making writing assistance available outside the constraints of classroom time, while Lo's (2023) review concludes that generative AI can enhance student engagement and extend the reach of individualized academic support beyond what is feasible for a single instructor working with a large class.

Empirical evidence on AI's effect on writing performance

Empirical evidence on whether these theoretical benefits translate into measurable writing improvement remains comparatively limited. Chan and Hu (2023) found that university students perceived generative AI as beneficial for idea generation and language refinement, though their study relied on self-report rather than direct assessment of writing quality. Angwaomaodoko's (2026) systematic review of empirical studies from 2020 to 2025 reports generally positive effects of generative AI on measured student learning outcomes, attributing these gains to adaptive assistance and continuous feedback, but notes that rigorous, controlled comparisons of AI-assisted versus unassisted written output remain relatively scarce. This gap motivates the controlled, quantitative design adopted in the present study.

Risks: over-reliance and academic integrity

Alongside these potential benefits, a substantial body of literature raises concerns about the risks of AI-assisted writing. Cotton et al. (2023) caution that over-reliance on AI-generated content may erode students' independent writing skills and reduce opportunities for critical, original reasoning, framing AI-assisted writing as occupying a "gray area" between legitimate support and academic misconduct. Dwivedi et al. (2023) similarly argue that the increasing sophistication of generative AI complicates conventional notions of authorship, intellectual contribution, and assessment validity. These concerns connect directly to the concept of academic integrity, which Bretag (2013) defines around principles of honesty, trust, fairness,

responsibility, and respect underpinning legitimate academic work. Because AI-generated text often appears original in form while its authorship cannot be unambiguously attributed to a specific human contributor, Perkins (2023) argues that the central ethical question is not whether AI was used, but whether that use was transparent and proportionate to what the assessment is meant to measure.

Institutional and policy responses to AI-assisted writing

In response to these tensions, higher education institutions have increasingly moved away from outright prohibition of AI tools toward governance frameworks that combine clear disclosure expectations with structured pedagogical guidance (Holmes et al., 2023). Zawacki-Richter et al. (2019) similarly argue that digital and AI literacy should be deliberately built into curricula rather than assumed to develop on their own, positioning institutional policy as a necessary complement to, rather than a substitute for, classroom-level assessment design. Perkins (2023) adds that policies framed purely around detecting AI use are unlikely to keep pace with the technology, and that disclosure-based, transparency-oriented frameworks are more sustainable in the long term. This body of policy-oriented literature, however, has developed largely independently of controlled empirical evidence on how large an effect AI assistance actually has on graded writing outcomes, a gap the present study is positioned to help close.

Literature Review

This study is informed by two complementary theoretical perspectives. Flower and Hayes' (1981) cognitive process theory of writing provides a basis for understanding how AI assistance might intervene in the planning, translating, and reviewing stages of composition, potentially altering the writing process even where the final product is the primary object of measurement. Zimmerman's (2002) framework of self-regulated learning contributes a complementary lens, emphasizing that learners actively plan, monitor, and evaluate their own strategies; from this perspective, the effect of AI assistance on writing performance is likely to depend not only on the tool's capabilities but on how deliberately and critically students incorporate its suggestions into their own writing process, an issue that bears directly on how a measured performance gain should be interpreted.

Research gap

Despite growing interest in generative AI and writing, few studies combine a controlled, quantitative measurement of AI's effect on writing performance, adjusted for students' pre-existing ability, with an explicit discussion of what such an effect implies for academic integrity and assessment validity. This gap is particularly evident in Vietnamese higher education, where empirical research on generative AI has more often examined broad applications such as automated learning support than direct, rubric-based writing outcomes (Zawacki-Richter et al., 2019). The present study addresses this gap through an experimental design conducted among English-majored students at HUIT.

Materials and Methods

Methodology

This study reports the experimental component of a larger university-funded research project examining the impact of next-generation AI chatbots on academic writing skills and academic integrity, conducted using an explanatory sequential mixed-methods design (Creswell & Creswell, 2018). A separate publication from the same project reports a survey-based analysis of students' AI-use patterns, perceived benefits and limitations, and academic integrity awareness (Q4-Q28 of the project questionnaire); that survey data is not reproduced here. The present article reports exclusively on the pretest-posttest writing-performance experiment, which has not been published elsewhere, and uses its results as an empirical basis for discussing academic integrity implications at a conceptual level.

Research context and participants

The study was conducted at the Faculty of Foreign Languages, Ho Chi Minh City University of Industry and Trade (HUIT), where academic writing constitutes a core component of the English-majored curriculum. Participants were 287 English-majored students spanning Year 2, Year 3, and Year 4 of study, recruited through convenience sampling. For the writing intervention, students were divided into two conditions: an AI-assisted group ($n = 150$), who received AI chatbot support while completing the posttest writing task, and a non-AI group ($n = 137$), who completed the same task independently. Group allocation followed practical classroom groupings rather than full randomization, a design constraint discussed further under Limitations.

Writing assessment instrument

Students completed a pretest and a posttest academic writing task, each requiring a 300-350-word essay on a topic related to education or technology, completed individually within 45 minutes. Written responses were scored using an analytic rubric covering five dimensions, each contributing equally to the overall writing score:

Table 1. Writing Assessment Rubric Dimensions

Dimension	Description
Organization	Logical arrangement and development of ideas throughout the text
Language accuracy	Grammatical correctness, vocabulary use, and linguistic appropriateness
Syntactic complexity	Use of varied and sophisticated sentence structures
Coherence and cohesion	Logical flow of ideas and effectiveness of connections between parts of the text
Critical thinking	Ability to analyze information, evaluate evidence, and construct reasoned arguments

This analytic, multi-dimensional approach was chosen over a single holistic score because it allows the overall writing score to reflect a composite of distinct competencies rather than a

single undifferentiated impression, which strengthens the interpretability of the resulting quantitative comparisons.

Validity and reliability of the writing assessment

Scoring criteria for each of the five rubric dimensions were defined in advance and applied consistently across pretest and posttest responses to support comparability of scores across time and between groups. The use of a multi-dimensional analytic rubric, rather than a single holistic impression, was intended to reduce the influence of rater impressions of overall fluency on the scoring of specific competencies such as critical thinking, which prior research suggests can be masked by superficially polished, AI-assisted language (Dwivedi et al., 2023). While formal inter-rater reliability statistics were not the focus of this article, scoring was conducted using the same rubric and criteria across both conditions and both testing points, supporting the internal comparability of the reported results.

Procedure

All participants completed the pretest task independently, without AI assistance, establishing a baseline measure of writing ability. During the subsequent posttest task, the AI-assisted group used AI chatbots under researcher-provided guidelines specifying permissible forms of assistance (e.g., brainstorming, language revision), while the non-AI group completed the equivalent task without such tools. Both tasks were completed under timed, individual conditions to maintain comparability between pretest and posttest administrations.

Data analysis procedures

Data were analyzed using IBM SPSS Statistics version 26. Prior to the main analyses, the dataset was screened for missing values and incomplete responses; only complete, valid cases ($N = 287$) were retained. A paired-samples t-test compared pretest and posttest scores for the full sample, addressing whether writing performance changed significantly over the course of the study. An independent-samples t-test then compared posttest scores between the AI-assisted and non-AI groups, addressing whether AI assistance was associated with higher writing scores. Finally, an Analysis of Covariance (ANCOVA) was conducted with posttest score as the dependent variable, writing group as the independent variable, and pretest score as a covariate, to determine whether group differences remained significant after statistically controlling for students' initial writing ability. Homogeneity of variance for the independent-samples comparison and the ANCOVA was assessed using Levene's test prior to interpreting each result.

Funding

This research was funded by Ho Chi Minh City University of Industry and Trade (HUIT) under Contract No. 510/HĐ-ĐCT dated 15 December 2025, with a total grant of VND 15,000,000.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Methodology, Project administration, Writing – original draft, Writing –

review & editing: Nguyen Minh Nhat. Investigation, Formal analysis, Data curation: Nguyen Thao My. Methodology, Formal analysis, Writing – original draft: Tran Thi Thuan. Methodology, Investigation: Le Thi Quynh Nhu. All authors have read and agreed to the published version of the manuscript.

Ethics Statement

Participation was voluntary, and students were informed of the study's purpose, procedures, and their right to withdraw at any stage without penalty. Data were collected and reported anonymously and used solely for research purposes. During manuscript preparation, the authors used an AI assistant (Claude, Anthropic) to support formatting and language editing based on the statistical analysis conducted by the research team; all data, statistical results, and conclusions were reviewed, verified, and approved by the authors, who take full responsibility for the content.

Results and Discussion

Pretest-posttest writing performance

Table 2. Paired Samples Statistics (N = 287)

Measure	M	SD
Pretest score	6.64	0.762
Posttest score	8.07	0.877

A paired-samples t-test showed a statistically significant increase from pretest to posttest, $t(286) = -51.16$, $p < .001$, with a mean gain of 1.43 points on the composite rubric scale. This substantial and highly significant increase indicates that students' writing performance improved considerably over the course of the intervention, consistent with the view that AI-assisted revision and feedback can meaningfully enhance measurable writing quality (Kasneci et al., 2023).

Comparison between AI-assisted and non-AI groups

Table 3. Posttest Scores by Group

Group	n	M	SD
Non-AI	137	7.79	0.858
AI-assisted	150	8.33	0.815

An independent-samples t-test showed that the AI-assisted group scored significantly higher on the posttest than the non-AI group, $t(285) = -5.46$, $p < .001$, mean difference = 0.54. Levene's test indicated equal variances, $F = 0.081$, $p = .776$, supporting the validity of this comparison. This finding indicates a statistically significant advantage associated with AI assistance during the writing task, beyond whatever general improvement occurred simply through repeated exposure to a similar writing task.

ANCOVA controlling for initial writing ability

Table 4. ANCOVA Results for Posttest Writing Scores

Source	df	F	p	Partial η^2
Pretest score (covariate)	1	1278.14	<.001	.818
Writing group	1	218.70	<.001	.435

Levene's test confirmed homogeneity of variance for the ANCOVA model, $F(1, 285) = 0.849$, $p = .358$. As expected, pretest score was a strong predictor of posttest performance, confirming that students' initial writing ability substantially shaped their later scores. More importantly, after controlling for this covariate, the effect of writing group on posttest performance remained statistically significant and, per Cohen's (1988) conventions, large in magnitude (partial $\eta^2 = .435$; R^2 for the overall model = .835). This indicates that the advantage associated with AI assistance cannot be explained by pre-existing differences in students' writing ability; it represents an effect over and above what prior competence would predict.

Discussion of the writing-performance effect

The convergence of three independent analyses, a paired-samples comparison, a between-groups comparison, and a covariate-adjusted comparison, provides robust evidence that AI-assisted writing produced a substantial, statistically significant improvement in students' academic writing performance. This is consistent with prior reports that AI chatbots can enhance measurable dimensions of writing quality such as language accuracy and organization (Kasneji et al., 2023; Lo, 2023), and extends this literature by demonstrating that the effect persists even after accounting for students' baseline writing ability, a stronger evidentiary standard than the self-reported perceptions that dominate much of the existing literature (Chan & Hu, 2023).

From the standpoint of Flower and Hayes' (1981) cognitive process theory, this improvement is plausibly explained by AI's capacity to intervene across multiple stages of the writing process simultaneously, easing the burden of generating and organizing ideas while also refining language, thereby freeing cognitive resources that second-language writers might otherwise devote entirely to managing grammar and vocabulary (Leki, 1992). At the same time, from a self-regulated learning perspective (Zimmerman, 2002), the magnitude of this effect likely depends on how actively students evaluated and integrated AI-generated suggestions rather than accepting them passively, a distinction this experimental design cannot directly disentangle but which bears directly on how the resulting gain should be interpreted.

Reconsidering academic integrity in light of a measured effect

This robust empirical finding sharpens, rather than resolves, the academic integrity question raised in the Introduction. A partial η^2 of .435 means that writing condition alone accounted for a substantial share of variance in posttest scores beyond what pretest ability explained. If a comparable share of a graded assignment's quality is attributable to AI assistance rather than the student's own developing competence, this raises exactly the fairness and interpretability

concerns that Cotton et al. (2023) and Bretag (2013) describe: two students may receive similar scores while having made very different personal intellectual contributions to the final text.

Consistent with Perkins' (2023) argument, this suggests that the central issue for academic integrity is not whether AI assistance measurably improves scores, since these results indicate that it clearly can, but whether the extent of that assistance is transparent to instructors and reflected in how such writing is assessed. Instructors and institutions cannot assume that a posttest score, once AI assistance is introduced, straightforwardly indexes a student's own writing competence in the way it might have before such tools existed. This concern is compounded by Dwivedi et al.'s (2023) observation that generative AI complicates conventional notions of authorship: the very fluency and coherence that make AI-assisted text score well on an analytic rubric are also what make it difficult to trace which portions reflect the student's own reasoning.

Comparison with international findings

The magnitude of the effect observed here, a large partial η^2 after covariate adjustment, is notably stronger evidence than the predominantly self-reported benefits documented in prior international research (Chan & Hu, 2023; Tlili et al., 2023). Whereas Angwaomaodoko's (2026) review of the broader literature found generally positive but often indirectly measured effects of generative AI on learning outcomes, the present design isolates the effect of AI assistance on a directly measured writing product while statistically accounting for pre-existing ability. This suggests that, at least for the analytic writing dimensions captured by the rubric used here, the benefits reported qualitatively in earlier studies can also be observed as a robust quantitative effect in a Vietnamese EFL context, extending this line of research beyond the Western and international settings on which much of the existing literature has focused (Zawacki-Richter et al., 2019).

Boundary conditions of the observed effect

It is worth emphasizing what this effect does, and does not, establish. The design compared writing produced with AI assistance against writing produced without it under timed, single-session conditions; it does not speak to how the effect might change over a longer sequence of assignments, nor to whether repeated AI-assisted practice eventually strengthens students' unassisted writing ability through a form of scaffolded learning, or instead entrenches dependence on the tool. Both possibilities are consistent with the self-regulated learning framework outlined earlier (Zimmerman, 2002): a student who critically evaluates and internalizes AI-generated suggestions may plausibly improve their own underlying competence over time, whereas a student who accepts suggestions uncritically may show a similarly large posttest gain without any corresponding growth in independent ability. The present data cannot distinguish between these two trajectories, which is precisely why the size of the measured effect should prompt caution in assessment design rather than straightforward celebration of AI's pedagogical value.

Implications for students, lecturers, and institutions

For students, these findings suggest that AI assistance can meaningfully improve the measurable quality of a piece of writing, but that this improvement does not automatically

translate into equivalent growth in the student's own underlying writing competence; students should therefore be encouraged to treat AI-assisted revision as a learning opportunity rather than a shortcut to a higher score.

For lecturers, the size of the AI-assisted effect implies that a single final-product score can no longer be assumed to transparently reflect a student's own competence once AI tools are in common use. This supports incorporating process-based evidence, such as drafts, revision histories, or disclosed AI use, alongside final scores, and considering rubric weightings that place additional emphasis on dimensions such as critical thinking, which are less directly amenable to AI-generated improvement.

For institutions, these results provide a concrete empirical basis, rather than a purely principled one, for developing disclosure-based policies on AI-assisted writing. Rather than prohibiting AI use outright, which prior research suggests is difficult to enforce and pedagogically counterproductive (Holmes et al., 2023), institutions may be better served by requiring transparent disclosure of AI assistance and adapting assessment criteria to remain valid under AI-assisted conditions.

Limitations

This study is limited by its single-institution, quasi-experimental design, which relied on existing classroom groupings rather than full random assignment to condition, leaving open the possibility of unmeasured group differences beyond pretest score. The intervention period was relatively short, which may understate longer-term effects, whether beneficial or detrimental, of sustained AI-assisted writing practice. Because this article deliberately excludes the project's survey-based data on students' academic integrity perceptions and AI-use patterns (reported separately), the discussion of academic integrity here is necessarily conceptual rather than empirically measured within this particular study. Finally, the writing rubric, while covering five distinct dimensions, produced a single composite score in the analyses reported here; a dimension-by-dimension quantitative breakdown was beyond the scope of this article and is a natural direction for future analysis.

An additional limitation concerns the nature of AI assistance itself: because students in the AI-assisted group were free to interact with the chatbot in varied ways within the researcher-provided guidelines, the study cannot isolate which specific forms of AI assistance, for example, language correction versus idea generation, contributed most to the observed performance gain. Future studies could address this by logging or categorizing students' actual interactions with the AI tool during the writing task.

Recommendations for Future Research

Building on these limitations, future research could pursue several directions. First, fully randomized controlled designs would strengthen causal inference beyond what the present quasi-experimental design allows. Second, dimension-level scoring of the writing rubric would clarify whether AI assistance benefits some competencies, such as language accuracy, considerably more than others, such as critical thinking, a question this study's composite scoring approach could not directly answer. Third, longitudinal designs following students

across multiple semesters of AI-assisted writing practice would help determine whether the large effect observed here persists, grows, or diminishes as students gain more experience working alongside AI tools. Finally, combining performance data of the kind reported here with the survey-based perception data reported in a companion publication from the same project could support a fuller, triangulated account of how measured performance gains relate to students' own awareness of academic integrity considerations.

Conclusion

This study found that AI-assisted writing produced a statistically significant and practically large improvement in the academic writing performance of English-majored students at HUIT, an effect that remained robust after controlling for students' initial writing ability. At the same time, the size of this effect underscores why AI-assisted writing complicates rather than simplifies academic integrity in writing assessment: substantial, measurable performance gains attributable to AI assistance mean that final writing scores can no longer be assumed to transparently reflect a student's own competence unless the extent of AI assistance is disclosed and factored into assessment design.

Institutions seeking to adopt AI-assisted writing pedagogically should pair such adoption with assessment reforms that make the human contribution to a piece of writing visible and evaluable in its own right, for example through disclosed-use policies, process-based evidence, and rubric criteria that continue to reward independent critical thinking. Future research might extend this design with fully randomized assignment, a longer intervention period, and dimension-level analysis of the writing rubric to clarify precisely which aspects of writing quality are most, and least, responsive to AI assistance.

Acknowledgements

This work was financially supported by Ho Chi Minh City University of Industry and Trade under Contract No. 510/HD-DCT dated 15 December 2025. The authors also thank Dr. Tran Tin Nghi for his supervision throughout this research project, and all students who voluntarily participated in the writing tasks.

Informed Consent Statement

Informed consent was obtained from all participants involved in the study.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to restrictions to protect participant privacy.

Declaration of Generative AI and AI-Assisted Technologies

During the preparation of this manuscript, the authors used Claude (Anthropic) to support language editing and formatting based on the statistical analysis conducted by the research team. All AI-assisted content was reviewed, verified, and edited by the authors, who take full responsibility for the accuracy, originality, and integrity of the published article.

References

- Angwaomaodoko, E. A. (2026). The Impact of Generative AI on Student Learning Outcomes: A Systematic Review of Empirical Studies (2020–2025). *Journal of Education and Training*, 13(2), 59. <https://doi.org/10.5296/jet.v13i2.23294>
- Bretag, T. (2013). Challenges in Addressing Plagiarism in Education. *PLoS Medicine*, 10(12), e1001574. <https://doi.org/10.1371/journal.pmed.1001574>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00411-8>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). SAGE Publications.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition & Communication*, 32(4), 365–387. <https://doi.org/10.58680/cc198115885>
- Holmes (2023). *Guidance for generative AI in education and research*. <https://doi.org/10.54675/ewzm9535>
- Hyland, K. (2019). *Feedback in Second Language Writing*. Cambridge University Press. <https://doi.org/10.1017/9781108635547>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2), 157–172. <https://doi.org/10.1080/03075079812331380364>
- Leki, I. (1992). *Understanding ESL Writers: A Guide for Teachers*. Boynton/Cook.
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2). <https://doi.org/10.53761/1.20.02.07>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1). <https://doi.org/10.1186/s40561-023-00237-x>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators?.

International Journal of Educational Technology in Higher Education, 16(1).

<https://doi.org/10.1186/s41239-019-0171-0>

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2