



AI-assisted writing's differential impact on EFL learners and implications for adaptive instruction

Shaobo Liang, Yingyu Wang, and Dan Wu

DOI: <https://doi.org/10.47989/ir31iConf64196>

Abstract

Introduction. While AI-assisted writing tools gain widespread educational adoption, their differential impacts on English as a foreign language (EFL) learners across proficiency levels remain underexplored. This study investigates how AI writing assistance affects EFL learners at different proficiency levels to inform differentiated instructional design.

Method. A controlled experiment with 17 Chinese university students divided into low-, intermediate-, and high-proficiency groups employed mixed methods, analyzing 85 writing samples and 56 human-AI interaction dialogues using SPSS and NVivo.

Analysis. Findings revealed striking proficiency-dependent effects: low-proficiency learners showed significant writing improvements after AI-assisted training, while intermediate- and high-proficiency learners demonstrated minimal gains. Interaction analysis uncovered distinct usage patterns—low- and intermediate-proficiency learners relied on AI primarily for linguistic compensation, whereas high-proficiency learners engaged in extensive prompt refinement to align AI output with their writing goals.

Conclusion. Results challenge one-size-fits-all approaches to AI writing instruction. Current AI tools excel at addressing basic linguistic needs but struggle with higher-order writing skills required by advanced learners. Findings emphasise the need for proficiency-responsive AI interventions and enhanced explainability to transform AI assistance into sustainable writing competence.

Introduction

Writing assignments serve as fundamental components of knowledge production and information processing, demanding that students effectively synthesise, organise, and communicate complex concepts. In educational settings, writing functions not merely as an output mechanism but as a crucial cognitive tool that promotes deeper understanding and skill development across academic disciplines (Hedgcock, 2005; Polio & Lee, 2017). The multifaceted nature of writing encompasses various performance dimensions, including content organisation, coherence, linguistic accuracy, and task effectiveness. This complexity makes writing a cognitively demanding activity that presents significant challenges for English as a foreign language (EFL) learners at all proficiency levels (Hoang et al., 2024).

Contemporary EFL learners face substantial obstacles in developing effective writing skills. Simultaneously, the rise of artificial intelligence technologies has opened new avenues for EFL learners to overcome these challenges (Yan, 2023). Current AI writing tools, such as ChatGPT, Grammarly, and Writefull, offer immediate feedback, provide targeted improvement suggestions, and support content generation. These tools have gained widespread adoption among EFL learners, creating novel opportunities for enhancing writing proficiency (Lee et al., 2024).

As AI technologies become increasingly integrated into English writing instruction, their impact on learners' writing competence has attracted growing scholarly interest. While research has demonstrated that effective AI integration can improve writing performance (Ouyang et al., 2023; Shahzad & Zahid, 2025), concerns persist that excessive dependence may undermine learner autonomy and impede long-term skill development (Xia et al., 2024; Nguyen et al., 2024).

Although previous studies have explored AI's role in writing instruction, most have concentrated on learners within a single proficiency level (Chen & Gong, 2025; Lai, 2025). This limited scope fails to capture group-based variations in AI assistance effects, thereby hindering the development of effective AI usage strategies in educational practice and making it challenging for instructional interventions to achieve differentiated and targeted implementation.

To bridge this gap, this study employs a mixed-methods approach that combines experimental design with human-AI interaction dialogue analysis. The research investigates the differential effects of AI-assisted writing across learners of varying proficiency levels while identifying the underlying factors that drive these differences. Specifically, the study addresses the following research questions:

RQ1: How does AI-assisted writing produce differential effects across learners of different proficiency levels?

RQ2: What factors account for these observed differences?

This research contributes to AI-enhanced language learning literature by providing empirical evidence of proficiency-dependent effects in AI-assisted writing. The findings offer practical guidance for educators implementing differentiated AI writing interventions. By examining underlying mechanisms through interaction analysis, this study advances understanding of how learners at different proficiency levels engage with AI writing tools, informing the development of more adaptive writing instruction.

Research design

This study employs a mixed-methods approach to examine AI-assisted writing effects across different proficiency levels. The quantitative component involved learners participating in three AI-assisted writing experiments within a controlled environment, with performance analyses conducted to assess differential effects across proficiency groups. The qualitative component

analyzed human-AI interaction dialogues to explore how learners at different proficiency levels vary in their AI prompt usage patterns.

By integrating quantitative and qualitative findings, this study investigates the differential effects of AI-assisted writing on EFL learners with varying proficiency levels and identifies the underlying factors driving these differences. The following sections detail the research methods and findings.

Participants

The sample comprised Chinese university students from various non-English disciplines with experience in AI-assisted writing. Participants were compensated with 180 RMB each. Questionnaires were first distributed to recruit participants with the same proficiency level in AI-assisted English writing. Before the experiment, all participants completed a pre-test. To minimise the potential influence of prior knowledge on writing topics, candidates first underwent a familiarity assessment, and the Social Life topic was ultimately selected for the experiment, with a difficulty level comparable to the CET-6. After determining the topic, participants were instructed to complete the writing pre-test without AI or any other tool assistance. The writing samples were independently scored by two professional instructors with CET-6 grading experience.

A total of 17 participants meeting the criteria were selected, including 7 males and 10 females. Based on the tertiles of the pre-test scores, participants were divided into three proficiency groups. While the sample size is limited, it remains appropriate given the experimental context and established precedents (Nguyen et al., 2024). Due to score distribution differences, the group sizes were slightly uneven, but this did not affect the overall validity of the experiment. The groupings were as follows:

- (1) Low proficiency group (Group A): Pre-test scores below the first tertile, n=5.
- (2) Intermediate proficiency group (Group B): Pre-test scores between the first and third tertiles, n=7.
- (3) High proficiency group (Group C): Pre-test scores above the third tertile, n=5.

Experimental procedures

The complete research design is shown in Figure 1. In this experiment, participants were required to complete writing tasks on the topic of Social Life, submitting one essay every three days. The study protocol involved three sessions of AI-assisted writing training followed by a writing test without AI assistance. During the AI-assisted training sessions, participants used AI as a supportive tool and were required to critically engage with the AI-generated content through comprehension, reflection, and integration processes, while refraining from directly using or copying the generated text. After each writing task, participants submitted both their completed essays and the complete human-AI interaction dialogues. To ensure the authenticity and reliability of the experimental data, participants were required to keep their cameras on throughout the entire writing process for comprehensive video recording.

All essays were scored anonymously by two professional instructors with extensive CET-6 grading experience, according to the standardised CET-6 writing evaluation criteria. The average of the two scores was taken as the participant's final writing score for analysis.

This study maintained strict control over the experimental environment, ensuring that AI was the sole assistive tool employed throughout the process. Specifically, Kimi was chosen for all AI-assisted writing tasks due to the following practical and technical considerations: its wide accessibility in mainland China, comprehensive bilingual support (Chinese and English), strong alignment with university students' established usage habits, and superior access speed and stability compared to foreign AI tools. Kimi provides comprehensive context-aware support

capabilities, including idea expansion, language refinement, grammar correction, and structural suggestions, making it particularly well-suited for investigating the differential effects of AI usage intensity on writing performance across different proficiency levels.

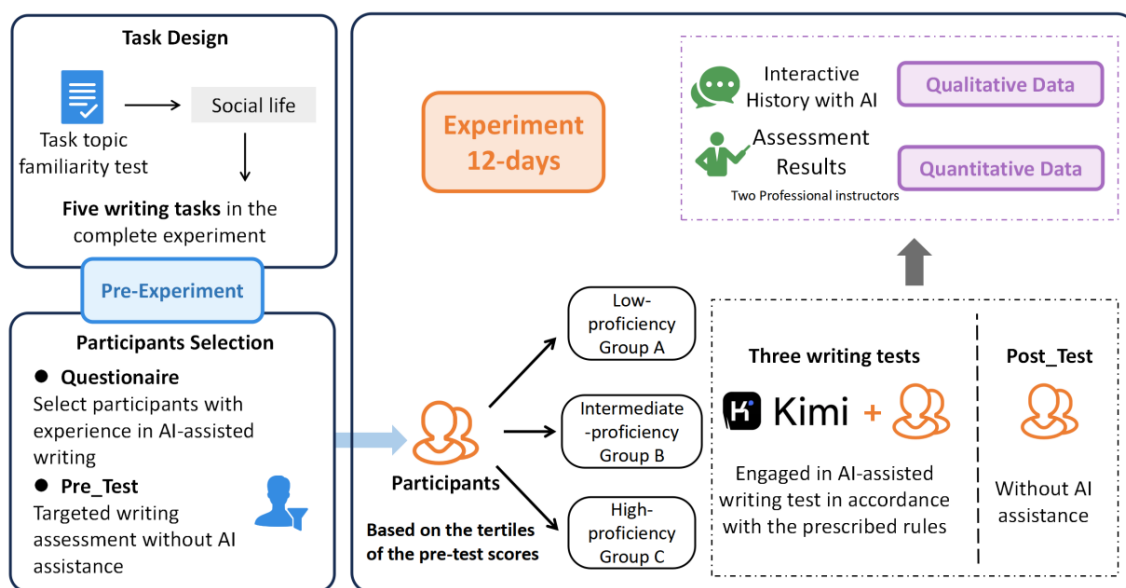


Figure 1. Complete experimental process.

Data analysis

Data sets

Data collection in this study encompassed both quantitative and qualitative dimensions, with all data obtained firsthand from the experiment. Quantitative data primarily consisted of the scores of participants' essays for each writing task, while qualitative data were derived from participants' human-AI interaction dialogues. The essay scores obtained during the pre-experiment phase were retained as baseline measures for longitudinal assessment of participants' English writing performance. The main data collected in this experiment were as follows:

- (1) Essays and scores: A total of 85 English essays along with their corresponding scores (including pre-test data).
- (2) Human-AI dialogues: A total of 56 dialogue records across the three proficiency groups (A, B, and C) over three rounds of AI-assisted writing sessions, totaling 104,397 words.

All participants in this study signed informed consent forms prior to the experiment, and data collection was conducted in accordance with ethical guidelines for human subject research. Additionally, all personally identifiable information was removed during data processing, and the content of essays and dialogues was fully anonymised.

Data analysis methods

For quantitative analysis, paired-sample t-tests were conducted using SPSS to compare changes in writing performance across the three experimental groups. For qualitative analysis, 56 human-AI dialogues were analyzed using a directed content analysis approach (Hsieh & Shannon, 2005). We applied a deductive coding scheme based on the seven categories derived from the cognitive model of writing (Flower & Hayes, 1981) and AI-assisted micro-processes (Nguyen et al., 2024). NVivo was utilised to systematically categorise the prompt data into these predefined dimensions. To ensure rigor in the classification process, researchers independently completed the coding and

documented their decisions, followed by iterative comparisons and discussions to resolve any discrepancies in the categorisation of prompts.

Results

AI-assisted writing performance of participant

To examine differences in learners' writing performance during AI-assisted training, paired-samples t-tests were conducted between pre-test scores and average scores during AI-assisted training (Table 1). Results revealed a significant difference between the pre-test and AI-assisted training performance in Group A ($M_{diff} = 3.999$, $SD = 2.590$, $t(4) = 3.453$, $p = .026$, 95% CI [0.784, 7.216]), suggesting that AI support generated a substantial immediate learning gain for low-proficiency learners. In contrast, no significant differences were observed for Group B ($M_{diff} = -0.595$, $SD = 0.699$, $t(6) = -2.251$, $p = .065$) or Group C ($M_{diff} = -0.799$, $SD = 1.076$, $t(4) = -1.662$, $p = .172$), both with $p > 0.05$. These findings suggest that while AI-assisted training significantly enhances immediate writing performance for low-proficiency learners, its effect on medium-proficiency and high-proficiency learners is not statistically significant.

Pair	Group	Mdiff	SD	T value	DF	P value	95% CI of the difference	
							Lower	Upper
AI_avg- Pre_test	A	3.999	2.590	3.453*	4	0.026	0.784	7.216
	B	-0.595	0.699	-2.251	6	0.065	-1.242	0.051
	C	-0.799	1.076	-1.662	4	0.172	-2.136	0.536

Note. * $p < 0.05$, ** $p < 0.01$.

Table 1. Paired-Samples t-test of pre-test and AI-assisted writing performance (N=17).

Writing performance of participants without AI assistance

To investigate the impact of AI-assisted training on learners' writing proficiency (i.e., learners' writing performance without AI assistance), this study conducted paired-samples t-tests comparing pre-test and post-test scores (Table 2). The results demonstrated a significant difference in Group A ($M_{diff} = 3.800$, $SD = 1.255$, $t(4) = 6.771$, $p = .002$, 95% CI [2.242, 5.358]), indicating a robust improvement in authentic writing performance following AI-assisted training. In contrast, neither Group B ($M_{diff} = -0.643$, $SD = 0.801$, $t(6) = -2.121$, $p = .078$, 95% CI [-1.384, 0.098]) nor Group C ($M_{diff} = -0.800$, $SD = 0.908$, $t(4) = -1.969$, $p = .120$, 95% CI [-1.928, 0.328]) showed statistically significant differences between pre- and post-test scores ($p > .05$). These findings suggest that AI-assisted writing training produced a sustained learning effect for lower-proficiency learners, significantly enhancing their writing proficiency beyond the training context. However, for medium-proficiency and high-proficiency learners, no significant improvements were observed once the AI assistance was withdrawn.

Pair	Group	Mdiff	SD	T value	DF	P value	95% CI of the difference	
							Lower	Upper
Post_test- Pre_test	A	3.800	1.255	6.771*	4	0.002	2.242	5.358
	B	-0.643	0.801	-2.121	6	0.078	-1.384	0.098
	C	-0.800	0.908	-1.969	4	0.120	-1.928	0.328

Table 2. Paired-Samples t-test of pre-test and post-test(N=17)

Note. *p <0.05, **p <0.01.

Analysis of participants' human-AI interaction dialogues

Building upon the theoretical frameworks of Flower and Hayes (1981) and Nguyen et al. (2024) and synthesised with the specific observations from our experimental human-AI interaction data, this study identified seven distinct categories of participant prompts. (Table 3).

Classification	Description
Role Setting	Users define roles or situational contexts to guide AI in producing text aligned with the intended style and tone.
Outline Structuring	Users engage AI to organise ideas and generate outlines that scaffold subsequent writing.
Content Retrieval	Users prompt AI to provide examples, facts, or references to enrich content and diversify arguments.
Language Guidance	Users seek AI assistance in understanding grammar rules, language usage, and stylistic choices.
Prompt Refinement	Users refine AI outputs through feedback and prompts to align with desired content and style.
Direct Generation	Users prompt AI to produce target text directly for rapid task completion.
Polish & Correction	Users rely on AI to refine language, improve style, and correct errors in existing text.

Table 3. Categorisation of prompts in human-AI dialogues.

Based on the above classification framework, a statistical analysis was conducted on the average number of prompts in human-AI dialogues across groups; the results are presented in Figure 2. The pie chart indicates that while there was little difference in the overall average number of prompts across groups, notable variations emerged in the distribution of specific behavioral categories.

Participants in the low- and intermediate-proficiency groups concentrated their prompts primarily on **language guidance** and **polish&correction**, reflecting their tendency to rely on AI to compensate for weaknesses in linguistic expression and grammatical accuracy, thereby ensuring the basic correctness and fluency of their writing.

In contrast, the high-proficiency group exhibited a distinct pattern: their proportion of prompts related to **prompt refinement** was significantly higher, while linguistic prompts were comparatively fewer. This suggests that high-proficiency learners have a solid command of linguistic knowledge and require less AI support at the grammatical and lexical levels. However, their more frequent engagement in prompt refinement indicates a higher expectation for the quality of AI-generated content, as they actively adjusted and refined prompts to guide the AI toward outputs that better aligned with their own ideas and writing objectives.

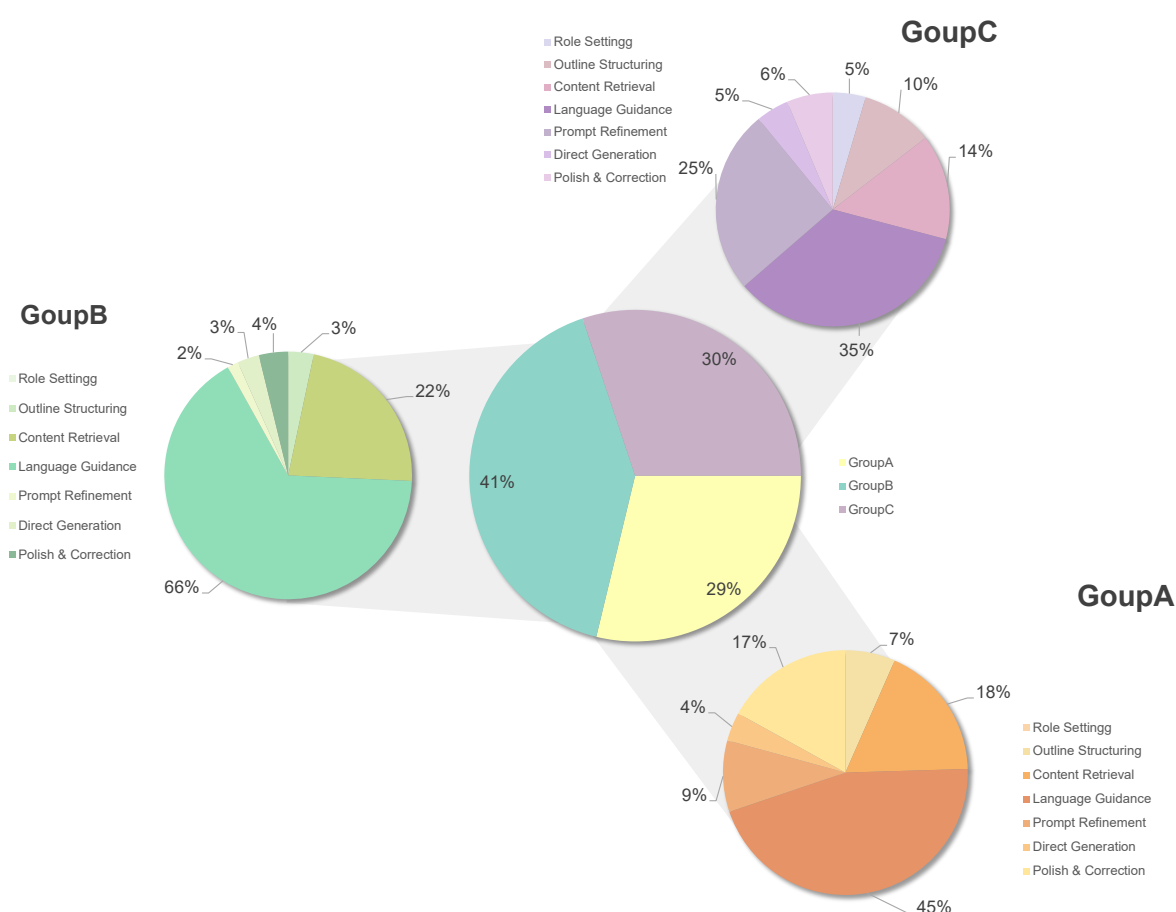


Figure 2. Average number of prompts classification chart.

Discussions

Proficiency-dependent learning outcomes

This study found that AI-assisted writing training produced markedly different outcomes across proficiency levels. Low-proficiency learners demonstrated significant improvement in writing performance, while intermediate- and high-proficiency groups showed minimal gains. This differential impact can be attributed to the varying nature of writing challenges faced by learners at different proficiency stages and the corresponding alignment with AI capabilities.

Interaction patterns and learning mechanisms

Analysis of human-AI interaction dialogues revealed distinct usage patterns across proficiency groups. Low- and intermediate-level learners primarily relied on AI for language guidance and

error correction functions to ensure basic accuracy and fluency in their writing. These learners utilised AI as a compensatory tool to address their linguistic limitations, allowing them to acquire and internalise linguistic knowledge during the interaction process.

In contrast, high-proficiency learners engaged more frequently in prompt refinement activities, repeatedly adjusting their requests to align AI-generated content with their specific writing intentions and rhetorical goals. This pattern suggests that advanced learners view AI as a collaborative partner rather than a remedial tool, though this collaboration often proved less beneficial than anticipated.

Understanding the compensatory effect for low-proficiency learners

For low-proficiency learners, the primary limitation lies in their weaker linguistic foundation. AI systems excel in language processing, error correction, and expression refinement—precisely the areas where these learners struggle most. Through interaction with AI, low-proficiency learners were able to bridge their linguistic gaps, gradually improving their language skills and overall writing proficiency. This compensatory effect explains why this group showed the most substantial improvement following AI-assisted training.

Limited benefits for advanced learners

For intermediate- and high-proficiency learners, writing challenges extend beyond basic linguistic competence to more complex higher-order skills such as idea organisation, thesis development, and revision strategies (Song & Wang, 2022). Unfortunately, current AI systems generally perform poorly in these sophisticated areas (Zhang et al., 2025). Consequently, AI-assisted training could not provide the type of advanced writing guidance these learners required, resulting in limited observable improvement.

High-proficiency learners, in particular, possess strong autonomous control over cognitive strategies and linguistic resources (Olive, 2014). The findings suggest that AI-generated content often failed to meet their sophisticated writing expectations, compelling them to invest considerable effort in prompt refinement. Rather than enhancing their writing process, AI assistance sometimes introduced additional operational burdens, potentially diminishing the learning benefits, and creating frustration among advanced users.

Implications for AI-enhanced writing instruction

These findings highlight the critical need for differentiated AI writing interventions. Current one-size-fits-all approaches to AI-assisted writing fail to address the diverse needs of learners across proficiency levels. Future pedagogical applications should begin with comprehensive proficiency assessments to identify individual learners' strengths and weaknesses. Based on these diagnostics, AI interventions should be tailored accordingly: lower-proficiency learners would benefit most from grammar and stylistic support, while intermediate- and high-proficiency learners require assistance with idea development, discourse organisation, and rhetorical strategies.

While focusing on the benefits of AI-assisted writing, this study acknowledges potential long-term drawbacks, which warrant further investigation in future research. Crucially, addressing these concerns involves enhancing AI explainability. Instead of providing outputs alone, AI systems should make their reasoning processes and references more transparent. For higher-order writing skills such as argumentation and critical thinking, revealing the underlying rationale behind AI-generated content would enable learners to observe, reflect, and internalise effective strategies. This shift from result-focused support to process-oriented guidance could foster deeper engagement and help transform AI assistance into sustainable writing competence.

Conclusion

This study demonstrates that AI-assisted writing produces proficiency-dependent effects among EFL learners. Low-proficiency learners benefited significantly from AI support through compensatory assistance in linguistic areas, while intermediate- and high-proficiency learners showed limited improvement due to AI's current limitations in addressing higher-order writing skills. These findings underscore the importance of differentiated AI writing interventions tailored to learner proficiency levels. Future developments should focus on enhancing AI explainability and providing process-oriented guidance to maximise learning benefits across all proficiency groups.

The study also has some limitations, including the short experimental duration and an insufficiently granular analysis of human-AI dialogues. Future research should address these constraints through extended training periods, expand the sample size, and more sensitive evaluation methods. Moreover, future studies could employ qualitative approaches such as grounded theory to conduct a more in-depth exploration of the textual nuances within human-AI interactions.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 72574176), the Natural Science Foundation of Hubei Province (No. 2025AFB770) and the Young Elite Scientists Sponsorship Program by CAST (No. 2023QNRC001).

About the authors

Shaobo Liang is an Associate Professor in School of Information Management, Wuhan University, China. He received his PhD from Wuhan University, and his research interests include the retrieval of information through intelligent systems. He can be contacted at liangshaobo.whu.edu.cn

Yingyu Wang is an undergraduate student at the School of Information Management, Wuhan University. Her research interests primarily include human-computer interaction (HCI) and digital humanities. She can be contacted at yingyuwang@whu.edu.cn

Dan Wu is a Professor in School of Information Management, Central China Normal University. She received her PhD from Peking University, and her research interests include human-computer interaction, human-AI interaction, and information behaviour. She can be contacted at woodan@whu.edu.cn

References

- Chen, C., & Gong, Y. (2025). The role of AI-assisted learning in academic writing: A mixed-methods study on Chinese as a second language students. *Education Sciences*, 15(2), Article 141. <https://doi.org/10.3390/educsci15020141>
- Hedgcock, J. S. (2005). Taking stock of research and pedagogy in L2 writing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 621–638). Routledge
- Hoang, D. T. N., & Hoang, T. (2024). Enhancing EFL students' academic writing skills in online learning via Google Docs-based collaboration: A mixed-methods study. *Computer Assisted Language Learning*, 37(7), 1504–1526. <https://doi.org/10.1080/09588221.2022.2083176>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>

- Lai, Z. C. C. (2025). The impact of AI-assisted blended learning on writing efficacy and resilience. *International Journal of Computer-Assisted Language Learning and Teaching*, 15(1), 1-21. <https://doi.org/10.4018/IJCALLT.377174>
- Lee, H. Y., Chen, P. H., Wang, W. S., Huang, Y. M., & Wu, T. T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education*, 21(1), 16. <https://doi.org/10.1186/s41239-024-00447-4>
- Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 49(5), 847-864. <https://doi.org/10.1080/03075079.2024.2323593>
- Olive, T. (2014). Toward a parallel and cascading model of the writing system: A review of research on writing processes coordination. *Journal of Writing Research*, 6(2), 173-194. <https://doi.org/10.17239/jowr-2014.06.02.4>
- Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, 20(1), 4. <https://doi.org/10.1186/s41239-022-00372-4>
- Polio, C., & Lee, J. (2017). Written language learning. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 299-317). Routledge.
- Shahzad, M. F., Xu, S., & Zahid, H. (2025). Exploring the impact of generative AI-based technologies on learning performance through self-efficacy, fairness & ethics, creativity, and trust in higher education. *Education and Information Technologies*, 30(3), 3691-3716. <https://doi.org/10.1007/s10639-024-12949-9>
- Song, Y., & Wang, S. (2022). The application of pragmatic presupposition in developing critical thinking of writing. *Chinese Journal of Applied Linguistics*, 45, 583-595. <https://doi.org/10.1515/CJAL-2022-0406>
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21(1), 40. <https://doi.org/10.1186/s41239-024-00468-z>
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28, 13943-13967. <https://doi.org/10.1007/s10639-023-11742-4>
- Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. *Innovation in Language Learning and Teaching*, 1-19. <https://doi.org/10.1080/17501229.2025.2503890>

© [CC-BY-NC 4.0](#) The Author(s). For more information, see our [Open Access Policy](#).