

Generative AI in EFL Writing Instruction: An Empirical Review of Pedagogical Benefits, Challenges, and Future Directions

Nguyen Tat Hiep, PhD candidate

Faculty of Foreign Languages, University of Labor and Social Affairs, Ho Chi Minh City, Vietnam.

Email: hiepnt@ldxh.edu.vn

ARTICLE INFO	ABSTRACT
<p>Received: April 1, 2026 Accepted: April 27, 2026 Volume: 7 Issue: 1</p> <hr/> <p>KEYWORDS</p> <p>Generative artificial intelligence; EFL writing education; artificial intelligence; academic integrity; literature review</p>	<p>This paper examines and integrates empirical research about the application of generative artificial intelligence (AI) in English as a Foreign Language (EFL) writing teaching. Generative AI has grown more common in writing schools to help with idea creation, linguistic accuracy, giving feedback, and rewriting. This is because huge language models like ChatGPT are becoming more common very quickly. However, scientific findings continue to be disjointed across various instructional situations and learner demographics. Utilizing a literature review methodology, this study analyzes recent empirical research to ascertain the categories of generative AI tools utilized, the educational contexts in which they are applied, and their documented impacts on the writing performance and learning experiences of EFL learners. The review emphasizes significant advantages, including increased language precision, less writing apprehension, and heightened learner involvement, while also acknowledging ongoing issues such as excessive dependence on AI, concerns regarding academic integrity, and disparities in technological access. This work synthesizes current information to present an evidence-based review of generative AI as both a pedagogical potential and a challenge in EFL writing education, providing implications for educators, researchers, and policymakers.</p>

1. Introduction

1.1 Background of study

Generative artificial intelligence (AI), especially large language models like ChatGPT, has grown very quickly and has had a big effect on how people teach, including how they teach English as a Foreign Language (EFL). Writing has garnered significant attention among the four language skills due to its cognitive intricacy and the difficulties EFL learners encounter in idea generation, content organization, and the production of linguistically precise writings. In recent years, generative AI techniques have been increasingly utilized to help EFL writing by helping with things like coming up with ideas, modeling language, giving automated feedback, and helping with revisions.

Empirical research has indicated both encouraging results and significant apprehensions over the incorporation of generative AI in EFL writing education. On the one hand, AI-assisted writing tools have been demonstrated to improve language accuracy, minimize writing anxiety, and give quick feedback, especially to students who are not very good at writing. On the other side, concerns about over-reliance, academic honesty, and the quality of AI-generated feedback have led to debates about the moral and educational usefulness of these technologies. Consequently, generative AI has emerged as a pivotal subject in contemporary discourse on its application in EFL writing, viewed as either an educational advantage or a prospective disadvantage.

1.2 Rationale of the study

Even while more real-world studies have looked into the use of generative AI in EFL writing instruction, the results are still spread out across different settings, groups of learners, and research methods. Many individual studies look at only one tool, a short-term intervention, or one learning result, which makes it hard for instructors and researchers to get a full picture of how well generative AI works and where it falls short in EFL writing classrooms. Furthermore, the swift advancement of AI technology has created a need to integrate contemporary empirical information to guide pedagogical decision-making.

There is still a paucity of targeted literature reviews that look at empirical research on generative AI in EFL writing education, even though there are more studies in this area. It is necessary to fill this vacuum to make present research trends clearer, uncover consistent results, and point out areas that need more study.

1.3 Significance of the study

There are many reasons why this literature review is important. First, it gives EFL teachers an overview of how generative AI has been utilized in writing teaching and what results they may realistically expect. Second, it gives curriculum designers and policymakers ideas for how to set rules for the ethical and pedagogical sound use of AI in language education. Finally, the study adds to the larger academic conversation on AI in education by looking at real-world evidence instead than just making guesses or using only theoretical arguments.

1.4 Objectives of this study

The goal of this study is to look at and combine real-world studies that look at how generative AI might be used to teach writing in English as a Foreign Language (EFL). This review seeks to delineate the varieties of generative AI tools utilized, the educational context in which they are applied, and their documented impact on the writing development of EFL learners through an analysis of available research. Moreover, the study aims to encapsulate the advantages and obstacles linked to AI-assisted writing habits as documented in empirical research.

1.5 The research questions

To fulfill the articulated objective, this literature evaluation is directed by the subsequent research inquiries:

- 1) What kinds of generative AI technologies are employed in real-world research on teaching EFL writing?
- 2) What impact do generative AI technologies exert on the writing performance and educational experiences of EFL learners?
- 3) What advantages and obstacles associated with the utilization of generative AI in EFL writing teaching have been documented in empirical studies?

2. Literature Review

The incorporation of Generative Artificial Intelligence (GenAI) into English as a Foreign Language (EFL) writing instruction is presently occurring at the pivotal nexus of the Fourth Industrial Revolution and assertive national digital transformation initiatives. This change is based on the "National Digital Transformation Program by 2025, with a vision to 2030" in the Socialist Republic of Vietnam. This program sees education as a key driver of modernization (Nguyen & Pham, 2025). This policy's goal is to create a highly skilled workforce that can help Vietnam reach its goal of becoming a global leader in semiconductors and advanced computing power by 2050.

In this context, GenAI has evolved from a futuristic novelty to a ubiquitous "cognitive co-pilot" in English Language Teaching (ELT), specifically engineered to address enduring challenges like the creation of formulaic, superficially topical writing (Webb & Şenaydın, 2025). GenAI helps us migrate from static information repositories to dynamic, adaptive learning ecosystems. This makes it possible to personalize and provide real-time scaffolding that meets the different cognitive needs of language learners. To understand this technological infusion, you need to make a complete list of the precise tools and functional affordances that are now being used to support the recursive stages of the writing process.

2.1 Classification of Generative AI Tools in Empirical Studies of EFL Writing

A detailed comprehension of AI tool classifications is crucial for strategic curriculum development, enabling educators to synchronize technical capabilities with educational goals across the many tiers of Bloom's Taxonomy (Elim, 2024; Jain, 2025). Present empirical literature differentiates between fundamental digital resources and novel learning platforms that employ Large Language Models (LLMs) to enhance various phases of the writing process. LLMs like GPT-4, Claude, and Writesonic are very important for automated content generation and pre-writing (Cainoi, 2026). These tools make it easier for students to come up with ideas for lessons, find theme structures, and create rough drafts of essays (Lee et al., 2025; Webb & Şenaydın, 2025). Linguistic refinement tools like Grammarly and QuillBot are useful for more than just the first draft. They help with correcting surface-level mistakes, adding variety to vocabulary, and changing the tone of writing. This lets students try out different ways of talking while getting instant feedback on their syntax (Lee et al., 2023).

Advanced instructional frameworks employ specific tools to enhance higher-order evaluative thinking. For example, the "AI-CRITIQUE" protocol uses AI-generated Socratic questioning to get students to think about their own argumentation assumptions. This turns the AI from just a content generator into a conversation partner (Hong et al., 2025). Additionally, specialized multimodal scaffolds such as ELSA Speak target phonetic deficiencies that indirectly affect the language confidence necessary for writing (Kholis, 2021). For Vietnamese learners, these tools give feedback on specific phonetic problems at the syllable level, such as how to pronounce terminal consonants, how to master complicated consonant clusters, and how to use English stress patterns correctly (Nguyen & Pham, 2025). By distinguishing between fundamental "foundational resources" such as PowerPoint and "innovative learning resources" that provide adaptive feedback, educators can more effectively incorporate these tools into the cognitive scaffolding and dialogic mediation components of the GenAI-CT paradigm. The diversity of these tools directly impacts quantifiable results in learner performance and the subtleties of the psychological learning experience.

2.2 Effects of GenAI on EFL Learners' Writing Performance and Learning Experiences

It is very important to measure both cognitive output and affective emotions in order to prove that AI integration in the language classroom works. Structured GenAI interventions have shown substantial empirical improvements in many critical thinking characteristics, such as analytical depth and content refinement, when evaluated through performance metrics. Data from vocational and academic settings demonstrate that structured interaction with tiered AI prompts yields substantial impact sizes, with Cohen's d values sometimes attaining or surpassing 1.42 (Hong et al., 2025). These gains indicate that students generate writing with markedly enhanced logical coherence when the AI promotes expert-level cognitive methods, including perspective-shifting and argument rearrangement. This technique encourages a transition from passive knowledge consumption to active, learner-centered participation, with the AI acting as a "critical friend" offering immediate feedback that is typically inaccessible in large classes (Webb & Şenaydin, 2025).

The learning experience is further enhanced by alleviating writing fear and fostering self-efficacy. GenAI creates a "low stakes" practice environment where students feel empowered to experiment with complex language without the fear of social judgment (Shen & Tao, 2025). Nonetheless, a significant divergence persists between the "spontaneous use" of AI by students for task clarification and the "structured engagement" mandated by instructional frameworks. Spontaneous use may result in superficial automation, whereas controlled interventions promote metacognitive reflection by necessitating that students evaluate, amend, or dismiss AI recommendations (Hong et al., 2025; Lee et al., 2025). This iterative involvement is important because it helps "fade" the scaffolds, which means that students slowly stop relying on AI and start to anticipate and include critical reasoning in their drafts on their own. These performance and emotional improvements are interesting, but they come with systemic and ethical trade-offs that leaders of institutions must deal with (Baron, 2023).

2.3 Potentials and Risks of Implementing GenAI in EFL Writing

For institutional leaders who want to create a sustainable digital environment, it is important to look at both the pros and cons of GenAI in a balanced way. The main strategic advantage of GenAI is that it can automate simple activities like grammar and spelling, which makes it easier for students to learn. This decrease allows pupils to concentrate their finite cognitive resources on advanced reasoning and the integration of evidence within their Zone of Proximal Development (Hong et al., 2025; Lee et al., 2025). Immediate, round-the-clock feedback serves as a cognitive bridge, facilitating differentiated education that meets the specific needs of each learner in real time (Mote, 2024).

Nonetheless, these advantages are mitigated by significant systemic obstacles, particularly a widespread "perception-usage contradiction" (Barrot, 2023). Students and professors are very worried about academic honesty and plagiarism, yet usage rates are still high. This could lead to a "dependency risk," where students don't question AI-generated information (Webb & Şenaydin, 2025). There is also a big digital divide. A quantitative study of ASEAN countries shows that high-speed internet access is 90% in cities but only 55% in rural areas. This makes it harder for students in resource-poor areas to access LLMs (Nguyen & Pham, 2025). This is made worse by a "policy vacuum," which means that teachers often have to work without clear rules or enough training in how to teach (Schiff, 2022). This lack of teacher competence is a major reason why people don't want to use technology. This is because teachers may feel technostress or a sense of loss for traditional procedures. These systemic difficulties underscore domains that existing literature and policy frameworks have not yet comprehensively addressed.

2.4 Research gaps

A significant deficiency is present in the examination of longitudinal heuristic mastery. Although short-term benefits are well demonstrated, there exists a relative deficiency in research concerning the preservation of critical thinking skills following the removal of the AI scaffold (Hong et al., 2025). Moreover, the prevailing rhetoric is significantly shaped by "Global North" viewpoints, frequently overlooking the socio-economic and cultural contexts of developing countries. For example, research in

Generative AI in EFL Writing Instruction: An Empirical Review of Pedagogical Benefits, Challenges, and Future Directions
Vietnam needs to consider the differences between metropolitan areas and resource-poor areas, such Central Vietnam, where poor infrastructure makes it very hard to get online (Nguyen & Pham, 2025).

Future study should investigate the convergence of AI adoption and regional paradigms, particularly the “face-saving” concerns and profound respect for authority characteristic of Confucian-heritage cultures, which frequently contribute to the “silent classroom” issue (Han, 2022). Lastly, there is a very important need for standardized, "human-first" policy frameworks to go from spontaneous adoption to ethical institutionalization. Conceptual models like the "Traffic Light" (prohibited, limited, or permitted use) and the "Two-Lane" systems have been suggested to uphold academic integrity; nonetheless, they necessitate thorough empirical validation in specific situations (Nguyen & Pham, 2025; Webb & Şenaydın, 2025). In the end, a human-centered approach is needed to make sure that GenAI helps close the digital divide instead of making it worse (Gabriel, 2024)

3. Methodology

3. Methodology

3.1 Research Design

This study adopts a qualitative literature review design to synthesize empirical research on the use of generative AI in EFL writing instruction. A literature review is appropriate because the purpose of the study is not to test a new classroom intervention but to examine and interpret existing evidence across multiple instructional contexts. The review focuses specifically on empirical studies so that the discussion is grounded in observed classroom practices, learner outcomes, and research-based conclusions rather than speculative commentary.

More specifically, the review follows an analytical and thematic orientation. It identifies recurring patterns in the literature related to the types of AI tools used, the educational settings in which they were implemented, the learning outcomes reported, and the pedagogical concerns raised by researchers. By organizing studies around these themes, the review provides a clearer understanding of current knowledge and emerging directions in the field.

3.2 Search Strategy and Selection Criteria

To identify relevant studies, the review prioritizes journal articles and conference papers that examine generative AI in EFL writing instruction. The search process may include major academic databases and scholarly search engines commonly used in educational research, such as Google Scholar, ERIC, Scopus, and other accessible indexing platforms. Key search terms include combinations of generative AI, ChatGPT, large language models, EFL writing, second language writing, AI-assisted writing, and writing instruction.

The inclusion criteria are as follows: first, the study must focus on English language learning, particularly EFL or closely related second-language writing contexts; second, it must examine generative AI tools or AI-supported writing systems with clear instructional relevance; third, it must report empirical

data, whether quantitative, qualitative, or mixed-methods; and fourth, it must be available in English. Studies that focus only on technical system design, opinion pieces, or non-empirical commentary are excluded. This selection strategy ensures that the review remains tightly aligned with the research questions and emphasizes evidence-based findings.

3.3 Data Extraction and Analytical Framework

After relevant studies are identified, each source is read carefully and coded according to a common set of analytical categories. These categories include publication year, educational context, learner level, type of AI tool, instructional purpose, research design, reported benefits, reported challenges, and major implications for teaching and learning. Organizing studies in this way makes comparison across sources more systematic and transparent.

The analysis relies primarily on thematic synthesis. Findings from individual studies are compared and grouped into broader patterns, such as support for idea generation, language accuracy, feedback, learner confidence, ethical concerns, and access-related limitations. This approach allows the review to move beyond simple summary by highlighting convergences and tensions across the literature. Where appropriate, the discussion also considers the Vietnamese educational context as a lens for interpreting the applicability of these findings.

3.4 Procedures

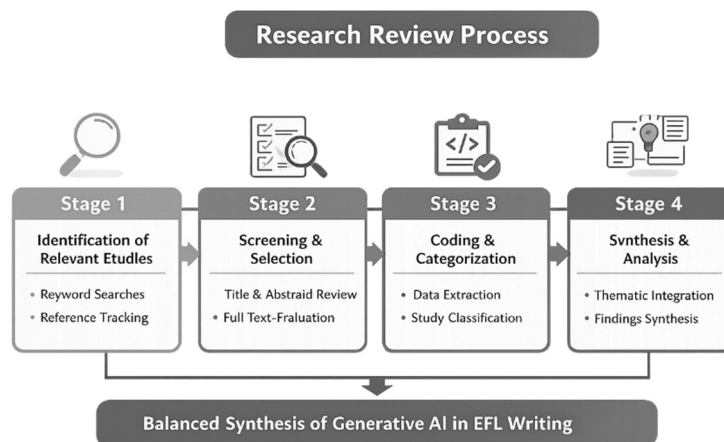


Figure 1. Research Review Process

The review is conducted in four main stages. In the first stage, potentially relevant studies are identified through keyword-based searches and reference tracking. In the second stage, titles, abstracts, and full texts are screened against the inclusion criteria. In the third stage, the selected studies are coded and grouped according to their main characteristics and findings. In the final stage, the coded information is synthesized into thematic sections that directly address the research questions (Figure 1)

To enhance coherence, the review gives particular attention to studies that provide clear descriptions of instructional use, learner response, and writing-related outcomes. The final synthesis therefore emphasizes not only whether AI tools appear beneficial, but also under what conditions they are most pedagogically meaningful. Through this procedure, the study aims to produce a balanced and well-structured account of generative AI as both an opportunity and a challenge in EFL writing instruction.

4. Recommendations

4.1 Educational Organizations

The synthesis of empirical studies reviewed in this paper indicates that generative artificial intelligence has the potential to substantially reshape EFL writing instruction, provided that its integration is pedagogically grounded, ethically regulated, and contextually responsive. At the institutional level, clear and context-sensitive academic integrity policies are urgently required. Higher education institutions should therefore develop transparent frameworks that distinguish acceptable AI-assisted learning from unacceptable AI-generated submission practices. Such policies should not merely prohibit misuse; they should also educate students and teachers about appropriate forms of disclosure, citation, and accountability.

Educational organizations should also invest in infrastructure, licensing, and support systems that make AI use more equitable. If access to paid platforms, stable internet, or institutional guidance remains uneven, then the benefits of AI will be distributed unevenly as well. Universities and language centers may therefore consider shared subscriptions, institutional workshops, and low-bandwidth alternatives so that GenAI becomes a support mechanism for inclusion rather than a source of widening disparity.

4.2 Teachers

At the pedagogical level, teachers should incorporate GenAI into writing instruction through structured and purpose-driven tasks rather than unregulated student use. Empirical evidence suggests that when GenAI is used as a scaffold for brainstorming, outlining, drafting, and revision, learners benefit from reduced

cognitive load and increased engagement. However, teachers should design follow-up activities that require learners to evaluate, modify, justify, and sometimes reject AI-generated suggestions. In this way, AI becomes a tool for inquiry rather than a substitute for thinking.

Teacher's professional development is equally important. Many language instructors may be interested in AI-supported pedagogy but still feel uncertain about its classroom management, ethical boundaries, or assessment implications. Professional development initiatives should therefore address both technical literacy and pedagogical judgment. Training should include prompt design, feedback comparison, reflective writing tasks, and classroom discussion strategies that help students use AI critically and responsibly.

4.3 Learners

For learners, the most important recommendation is to develop AI literacy alongside writing competence. Students should be taught that AI systems are useful but imperfect tools whose outputs must be checked for relevance, accuracy, tone, and appropriateness. Rather than copying AI-generated content directly, learners should use it to explore ideas, identify language alternatives, and reflect on revision choices. This practice can help students build metacognitive awareness and maintain ownership of their texts.

Learners should also be encouraged to document their writing process. Short reflection notes, revision logs, or prompt-response records can make AI-supported writing more transparent and can help students recognize how their own thinking develops across drafts. Such practices are especially valuable in EFL contexts because they reinforce learner autonomy and make writing growth visible to both students and teachers.

5. Conclusion

The rapid emergence of generative artificial intelligence has introduced both significant opportunities and complex challenges for EFL writing instruction. This literature review has synthesized empirical studies examining the use of generative AI tools, particularly large language models such as ChatGPT, in diverse EFL writing contexts. The findings indicate that, when pedagogically guided, generative AI can function as an effective scaffold that supports idea generation, linguistic accuracy, and learner confidence, especially during the early stages of the writing process. At the same time, unregulated or purely instrumental use of these tools raises serious concerns related to academic integrity, learner dependency, and the erosion of individual authorial voice.

The review further demonstrates that learning outcomes are strongly influenced by how generative AI is integrated into instruction. Structured engagement requiring learners to evaluate, revise, and reflect on AI-generated output consistently leads to deeper cognitive engagement than spontaneous or unguided use. In the Vietnamese EFL context, where large class sizes, exam-oriented curricula, and regional disparities in digital access persist, generative AI offers both a practical form of supplementary support and a potential source of inequity if access and policy issues remain unaddressed.

Overall, this review suggests that the pedagogical value of generative AI in EFL writing does not reside in the technology itself, but in the human-centered frameworks that govern its use. Effective integration requires the development of AI literacy, clear institutional policies, thoughtful assessment practices, and sustained teacher professional development. As generative AI technologies continue to evolve, future research must move beyond short-term performance gains to examine long-term cognitive development, learner identity, and culturally responsive practices. Only through such balanced and context-sensitive approaches can generative AI contribute meaningfully to equitable and ethical EFL writing instruction.

References

- Baron, N. S. (2023). *Who wrote this? How AI and the lure of efficiency threaten human writing*. Stanford University Press.
- 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>
- Cainoy, J. P. D. (2026). A Systematic Review of ChatGPT and LLMs in English Academic Writing: Challenges, Opportunities, and the Development of the RAISE Framework. *Empowering Learners With AI: Strategies, Ethics, and Frameworks*, 153-184.
- Elim, E. H. S. Y. (2024). Promoting cognitive skills in AI-supported learning environments: The integration of Bloom's taxonomy. *Education* 3(13), 1-11.
- Gabriel, S. (2024). Generative AI and educational (in) equity. *Proceedings of the 4th International Conference on AI Research, Portugal*, 133-142.
- Han, I. (2022). Contextualization of communicative language teaching in Confucian heritage culture: Challenging pedagogic dichotomization. *Sage Open*, 12(1), 21582440221079895.
- Hong, H., Vate-U-Lan, P., & Viriyavejakul, C. (2025). Generative AI-mediated scaffolds for enhanced critical thinking in EFL writing. *Edelweiss Applied Science and Technology*, 9(6), 43-54. <https://doi.org/10.55214/25768484.v9i6.7751>
- Jain, J., & Samuel, M. (2025). Bloom Meets Gen AI: Reconceptualising Bloom's Taxonomy in the Era of Co-piloted Learning.
- Kholis, A. (2021). Elsa speak app: automatic speech recognition (ASR) for supplementing English pronunciation skills. *Pedagogy: Journal of English Language Teaching*, 9(1), 01-14.
- Lee, Y.-J., Davis, R. O., & Choi, J.-I. (2025). Integrating generative AI into EFL writing: University students' strategies and perceptions. *Online Journal of Communication and Media Technologies*, 15(4), e202541. <https://doi.org/10.30935/ojcm/17545>

- Lee, H., Kim, T., & Jeon, J. (2023). A study of evaluation and feedback analysis of Korean English instructors, native English instructors, and ChatGPT on college students' English writing. *Journal of Modern British and American Language and Literature*, 41(4), 247-270. <https://doi.org/10.21084/jmball.2023.11.41.4.247>
- Mote, N. (2024). Role of technology in EFL students' autonomy. *Harvest: An International Multidisciplinary and Multilingual Research Journal*, 3(5), 193-199.
- Nguyen, T. P. N., & Pham, T. H. (2025). Challenges and opportunities for digital learning resource development: An analysis of AI application in Vietnamese general education. *Advances in Artificial Intelligence and Machine Learning*, 5(3), 4292-4307.
- Schiff, D. (2022). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*, 32(3), 527-563.
- Shen, X., & Tao, Y. (2025). Metacognitive Strategies, AI-based Writing Self-efficacy and Writing Anxiety in AI-assisted Writing Contexts: A Structural Equation Modeling Analysis. *International Journal of TESOL Studies*, 7(1).
- Webb, R., & Şenaydın, F. (2025). Generative AI in English language teaching: Students' voices, teachers' reactions, and needs. *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, 29(3). <https://doi.org/10.55593/ej.29115int3>