

A Review of Research on the Application of Generative Artificial Intelligence in Second Language Writing Feedback

Liu Yu

Xi'an International Studies University, Xi'an, Shaanxi Province, 710000;

Abstract: Generative artificial intelligence (GAI), represented by ChatGPT, is profoundly transforming the field of feedback in second language (L2) writing instruction. This paper systematically reviews twenty-one recent core research articles to clarify the research landscape and developmental trajectory of this domain. The literature review is structured around four core dimensions: comparative studies of AI and teacher feedback, learner reception and engagement with feedback, exploration of pedagogical integration and application models, and technical characteristics and optimization pathways. The review finds that GAI demonstrates clear advantages in feedback efficiency, coverage breadth and immediacy, particularly excelling in error correction of linguistic forms and vocabulary enhancement. However, it still cannot replace teachers in areas requiring in-depth logical comprehension, personalized emotional support and heuristic guidance. Learner engagement with feedback is significantly influenced by their self-regulation capabilities and AI literacy. The “human-AI collaboration” model emerges as the most promising application paradigm, within which the teacher’s role is shifting from a “knowledge transmitter” to a “learning facilitator and AI literacy coach”. Future research should prioritize the long-term effects of GAI feedback and its cross-cultural adaptability, commit to developing more intelligent and integrated learning systems, and strengthen the construction of relevant theoretical frameworks.

Key Words: Generative Artificial Intelligence (GAI); Large Language Models (LLMs); Second Language Writing Feedback; Human-AI Collaboration; AI Literacy; Learner Engagement

DOI: 10.69979/3041-0843.26.01.068

1 Introduction

With the explosive development of generative artificial intelligence (GAI) technologies, such as ChatGPT, the field of education is undergoing a profound paradigmatic shift. In the context of second language (L2) writing instruction, feedback as a crucial link connecting “writing production” with “language internalization” has long been plagued by multiple challenges. Traditional teacher feedback, constrained by time and energy, struggles to deliver immediate, frequent, and deeply personalized support in large-scale classes. While early Automated Essay Scoring (AES) systems improved efficiency, they often failed to meet complex pedagogical needs due to their narrow feedback dimensions, lack of explanatory power, and rigid models. Generative AI, particularly large language models (LLMs), offers a novel technological pathway and pedagogical potential to tackle these longstanding challenges by leveraging their powerful natural language understanding and generation capabilities, flexible “few-shot learning” characteristics, and human-like interactive conversational capabilities (Guo & Wang, 2024).

This study aims to systematically review and synthesize twenty-one recent empirical and theoretical studies that focus on the application of generative AI in second language (L2) writing feedback. It thus carries significant theoretical and practical implications. At the theoretical level, by conducting an in-depth comparison of the similarities and differences between AI and human feedback in terms of efficacy, strategies, learner and instructor attitudes, and learner acceptance, it contributes to a deeper understanding of core theoretical issues such as the cognitive mechanisms of human-AI collaboration, learner agency mediated by technology, and the transformation of teacher roles in the digital age (Chen & Lyu, 2024; Zou et al., 2024). At the practical level, this review provides evidence-based decision-making guidance for frontline teachers, curriculum designers, and educational administrators. It guides them in integrating AI tools into writing instruction scientifically and prudently, optimizing the feedback ecosystem, thereby enhancing teaching efficiency while safeguarding and even deepening learning outcomes. Of particular note is that for regions or schools with relatively scarce educational resources, low-cost and easily accessible AI-assisted feedback tools may act as a key lever to promote equity in educational processes (Huang & Guo, 2024; Liu et al., 2025).

2 Literature Review

The literature selected for this review covers the period from 2023 to 2025. The sources are diverse and comprehensive, including authoritative domestic and international journals such as *Foreign Language Teaching Theory and Practice*, *Foreign Languages in China*, *System*, *Computer Assisted Language Learning*, as well as proceedings from notable academic conferences, thereby demonstrating the depth and breadth of the research.

This study retrieved relevant literature from 2023 to 2025 through three databases: CNKI, Web of Science Core Collection, and Scopus. The search keywords included “generative artificial intelligence/GAI”, “second language writing feedback”, and “ChatGPT/large language models/LLMs”. A total of 50 initial documents were obtained. After applying inclusion criteria (focusing on L2 writing feedback scenarios; including comparisons between AI and teacher feedback or analysis of learner usage effects; published in SSCI and CSSCI core journals or authoritative academic conferences) and exclusion criteria (non-L2 teaching scenarios; pure technology development studies without

pedagogical application analysis), 21 core articles were finally included (13 Chinese, 8 English; 18 empirical studies, 3 theoretical studies), forming the basis for this review.

2.1 Comparative Studies on AI and Teacher Feedback: Efficacy, Strategy, and Attitudinal Differences

A large body of empirical research has systematically analyzed the similarities and differences between feedback generated by Generative AI and that provided by teachers through controlled comparative experiments, primarily focusing on three core dimensions: rating consistency, feedback content, and feedback strategies.

Regarding rating consistency and feedback efficacy, studies indicate that AI, particularly GPT-4 when optimized with appropriate prompt engineering, demonstrates moderate to high correlation with teacher ratings. For instance, Huang and Guo (2024) found that under an optimal prompt strategy combining essay topic, scoring rubric and human-graded sample, the Pearson correlation coefficient between GPT-4 and human raters reached 0.96, with a Quadratic Weighted Kappa (QWK) coefficient of 0.56, indicating a “moderate level of agreement.” Ren et al. (2024) also confirmed a degree of consistency between ChatGPT and teachers in error identification. However, such consistency is not without limitations. Multiple studies point out potential systematic biases in AI scoring, such as a tendency to assign higher scores compared to teachers (Huang & Guo, 2024; Zhang et al., 2025). Furthermore, AI’s consistency and accuracy require further improvement when handling writing tasks that demand deep contextual understanding, cultural background knowledge, or complex logical argumentation (Wei & Li, 2023; Zhou et al., 2023).

Concerning feedback quantity, dimensions, and strategies, research findings show a high degree of convergence: AI-generated feedback is significantly more abundant than teacher feedback (Ren et al., 2024; Guo & Wang, 2024). For example, in the study by Ren et al. (2024), ChatGPT produced 449 feedback items for 50 essays, far exceeding the 348 items provided by the teacher. Besides, AI feedback tends to cover various dimensions more evenly. Guo and Wang (2024) suggests that while teacher feedback primarily focuses on content and language issues, ChatGPT distributes its attention relatively equally across the foci of content, organization, and language. In terms of strategy types, AI exhibits a strong reliance on explicit feedback, such as direct correction, with its feedback expressions often being straightforward and specific (Zhang et al., 2025; Ren et al., 2024). In contrast, teachers more frequently employ guiding and implicit strategies like metalinguistic comments and clarification requests, aiming to stimulate students’ capacity for self-reflection and revision (Ren et al., 2024; Zeevy-Solovey, 2024).

In terms of feedback attitude and affective orientation, although both parties primarily provide corrective (negative) feedback, AI offers a significantly higher proportion of positive (encouraging) feedback than teachers at the structural and content levels of feedback, displaying a stronger “motivational orientation” (Ren et al., 2024). However, this efficient positive feedback has also been critiqued by some researchers as “depersonalized,” “lacking emotional warmth,” and “mechanical” (Wang et al., 2025). In comparison, teacher feedback retains its irreplaceable value in providing emotional support, understanding students’ writing intentions and specific contexts, and facilitating affective interpersonal interaction (Alsofyani & Barzanji, 2025; Zeevy-Solovey, 2024).

2.2 Learner Perspectives: Acceptance, Engagement, and Individual Differences

Learners’ perceptions, preferences, and usage strategies towards feedback present complex characteristics. Research indicates that students generally hold open and positive attitudes towards artificial intelligence (AI) feedback, and highly value its efficiency and accuracy in grammatical error correction, vocabulary enhancement, and the cultivation of idiomatic expression (Alsofyani & Barzanji, 2025; Mao et al., 2024). However, in dimensions such as in-depth logical reasoning, content originality, cultural appropriateness, and personal emotional expression, students demonstrate a stronger sense of trust in teacher feedback (Zou et al., 2024; Xu et al., 2024). Multiple preference surveys show that pure teacher feedback or the teacher and AI combined feedback model is the most favored by students (Escalante et al., 2023; Zeevy-Solovey, 2024). In terms of feedback usage behavior, students exhibit distinct instrumental characteristics and selective tendencies: they readily accept and adopt AI feedback suggestions at the linguistic form level, while at the level of content creativity and personal viewpoint expression, they are more inclined to retain their own judgment or seek professional guidance from teachers (Wang et al., 2025; Yan & Zhang, 2024).

There exist significant individual differences among learners in terms of behavioral and cognitive engagement with feedback. A case study by Teng et al. (2025) found that learners with a high level of self-regulation ability can interact with AI more efficiently, design and use more professional prompts, conduct in-depth processing of AI feedback such as analysis, evaluation, integration, and implement multi-dimensional and higher-order text revision strategies. In contrast, learners with a low level of self-regulation ability mostly engage in superficial processing and mechanical revision of feedback, and are prone to “cognitive fatigue” after completing multiple relevant learning tasks. Xu (2025) also points out that learners’ cognitive inquiries during interactions with AI mostly remain at the level of lower-order thinking. A multiple-case study by Yan and Zhang (2024) further reveals that students with higher linguistic proficiency and stronger technological application skills exhibit more in-depth and effective behavioral and cognitive engagement with ChatGPT feedback. Notably, the introduction of AI feedback has objectively prompted students to develop metacognitive strategies for the integration and screening of feedback across different types of feedback sources (Wang et al., 2025).

Learners’ individual factors play a crucial moderating role in the effectiveness of AI feedback. Existing research has confirmed that learners’ linguistic proficiency, AI literacy particularly prompt design skills, and familiarity with technology are the core variables influencing the extent to which they benefit from AI feedback (Xu, 2025; Yan & Zhang, 2024). Meanwhile, research has found that technological application ability can compensate for deficiencies in learners’ linguistic proficiency to a certain extent, helping learners with lower linguistic proficiency utilize AI tools more efficiently for learning (Yan & Zhang, 2024).

2.3 Exploration of Pedagogical Integration and Application Models: From Tool Assistance to Ecological

Transformation

Research has progressed beyond merely comparing tool efficacy to delve into how AI can be integrated into and reshape the ecology of writing instruction.

The “human-machine collaboration” feedback model has emerged as a mainstream consensus and a core recommendation in relevant research. Researchers widely agree that AI should serve as an efficient and precise “initial filter” and “resource generator,” responsible for handling a large number of foundational and rule-based error correction and polishing tasks. Teachers, conversely, should assume the central roles of “guide,” “evaluator,” and “emotional supporter,” focusing on areas where AI is incompetent, such as fostering in-depth logical reasoning, providing personalized guidance, motivating learners, and guiding academic ethics (Chen & Lyu, 2024; Han & Li, 2024). For example, the “ChatGPT-supported teacher feedback” model explored by Han and Li (2024), where teachers utilize ChatGPT to generate initial feedback drafts and then conduct professional screening and editing. It combines the strengths of both parties and has achieved positive effects on student revisions.

The integration of AI is driving a shift in the teacher’s role. Certain repetitive and knowledge-transmission tasks have been weakened, while competencies in technology integration, critical evaluation of AI output, cultivation of higher-order thinking, and emotional and ethical guidance have been unprecedentedly strengthened (Chen & Lyu, 2024; Wei & Li, 2023). Teachers need to evolve into “prompt designers” proficient in AI characteristics and “AI literacy coaches” for students.

The instructional processes and assessment systems have been innovated. AI has been integrated into the entire writing process (pre-writing brainstorming, while-writing assistance, post-writing review and reflection), powerfully driving the implementation of process-oriented and formative assessment (Huang & Guo, 2024). Research has also begun to explore AI-based personalized learning path recommendations, the construction of dynamic digital error logs, and multi-source feedback loop systems that integrate input from AI, peers and teachers (Liu et al., 2025; Wei & Li, 2023).

2.4 Technical Characteristics and Optimization Pathways: Prompt Engineering and Model Limitations

Prompt engineering has been identified as a critical lever for determining the quality of AI-generated feedback. Research indicates that prompts incorporating specific role definitions, clear task instructions, detailed scoring rubrics and a small number of high-quality human-scored samples can significantly enhance the consistency of AI scoring and the relevance of its feedback (Huang & Guo, 2024). For instance, among the seven types of prompts designed by Huang and Guo (2024), the combination of essay topic, scoring rubric and human-scored samples yielded the best results. Few-shot prompts generally outperform zero-shot prompts in most writing tasks.

The advantages and limitations of current AI models coexist. The strengths are evident in their ability to process enormous linguistic patterns, generate fluent and diverse expressions and provide immediate interactive feedback (Wang et al., 2024). However, their limitations are also pronounced: their comprehension of specific contexts, writers’ creative intent and sociocultural backgrounds remains superficial (Zhou et al., 2023; Zeevy-Solovey, 2024). The generated feedback may contain factual inaccuracies, internal logical inconsistencies or potential biases (Chen & Lyu, 2024). Excessively direct and detailed corrective feedback may inhibit the development of students’ capacity for self-directed inquiry and metacognitive reflection (Sun et al., 2025). Their role in fostering higher-order competencies such as creative and critical thinking remains relatively limited (Liu et al., 2025).

3 Implications and Future Directions

Building upon the four thematic dimensions delineated in the preceding literature review, the following discussion synthesizes key insights into two interconnected points. The first point translates the empirical and theoretical findings into pedagogical implications for classroom practice, while the second identifies gaps in the current research landscape and proposes a forward-looking agenda for scholarly inquiry.

3.1 Implications for Teaching Practice

First, it is essential to actively advocate for and implement the “teacher-led human-machine collaboration” feedback model. Educational institutions should encourage teachers to apply generative AI as a regular pedagogical tool and provide them with systematic specialized training. Specific teaching processes can be designed, such as “student revision based on AI feedback, then teacher guidance on unresolved issues and deeper content,” thereby clarifying the division of labor between humans and AI (Han & Li, 2024).

Second, systematically cultivate “AI literacy” for both students and teachers. Core modules such as “ethics of AI use in academic settings,” and “critical evaluation of AI-generated content” should be organically integrated into the curriculum teaching system. Teacher professional development programs must strengthen specialized training in AI tool usage, pedagogical integration design, and AI output evaluation competencies.

Third, address differentiated instruction and the equity of technology application. It is crucial to fully recognize the individual differences in learners’ benefits from AI feedback and provide more scaffolded support for students with lower self-regulation ability or AI literacy, such as pre-designed prompt templates, step-by-step operational guidance, and specialized training in metacognitive strategies (Teng et al., 2025). Additionally, the low-cost and easily accessible nature of AI should be actively leveraged to enhance the quality and efficiency of writing instruction in educationally under-resourced regions and promote the realization of fairness in educational processes (Liu et al., 2025).

3.2 Directions for Future Research

First, deepen longitudinal effect and tracking studies. Most existing research consists of short-term intervention studies, and there is an urgent need for longitudinal tracking studies to thoroughly investigate the sustained impact of AI feedback on the long-term development of

students' writing skills, the development of autonomous learning habits, the improvement of thinking dispositions, and the shaping of perceptions of academic integrity.

Second, expand the diversity of research contexts and participants. Current studies often focus on specific genres (e.g., argumentative essays, narratives) and general education contexts at secondary or tertiary levels. Future research should further examine the effectiveness and suitability of AI applications across different genres (e.g., academic papers, creative writing), various educational stages (e.g., primary education, vocational training), and among learner groups from diverse linguistic and cultural backgrounds.

Third, explore the development of more intelligent and integrated learning systems. Future research could be dedicated to developing intelligent writing tutoring environments deeply integrated with generative AI. Such systems would be capable of constructing personalized learner profiles based on continuously collected learner data, dynamically adjusting feedback focus, instructional strategies and adaptive linguistic difficulty, and achieving seamless real-time data connectivity with teacher and administrative interfaces, thereby providing stronger support for scalable personalized writing learning.

Fourth, strengthen theoretical construction and interdisciplinary dialogue. Greater efforts should be made to deeply explain the nature and internal mechanisms of human-machine collaborative writing from theoretical perspectives such as sociocultural theory, activity theory, and complex dynamic systems theory (Chen & Lyu, 2024). Moreover, enhanced cross-disciplinary integration and academic dialogue with fields like educational technology, cognitive science, and computer science are necessary to jointly propel this research field towards more in-depth development.

This study has several limitations. Initially, the literature screening only covers the period from 2023 to 2025, excluding early foundational studies on ChatGPT when it was first launched in 2022, which may omit some key theoretical discussions. In addition, among the 21 included articles, most focus on higher education contexts, with insufficient coverage of basic education, vocational education, and other scenarios. Furthermore, the study does not explore differences in L2 learners' acceptance of AI feedback across different cultural backgrounds (e.g., Chinese vs. Western learners), which may affect the cross-cultural applicability of the conclusions. These limitations provide directions for further expansion in future research.

4 Conclusion

In summary, generative artificial intelligence (GAI) has demonstrated significant application potential and practical value in the domain of second language (L2) writing feedback. It is not aimed at replacing teachers, but rather a powerful assistant and collaborator that empowers teaching and unleashes teachers' productivity. Its core value lies in freeing teachers from certain routine repetitive tasks, enabling them to focus more on the most humane, creative, and challenging core aspects of instruction. Existing research has begun to outline a practical roadmap for "human-machine collaboration with complementary strengths". The key to future development is to construct a harmonious, efficient and educationally wise human-machine collaborative ecosystem. This requires not only the continuous optimization of model performance, safety and interpretability by technology developers, but more importantly, demands that educational researchers and practitioners, in a spirit of critical thinking and creativity, actively explore evidence-based diverse models and best practices for the deep integration of AI with L2 writing pedagogy, ultimately achieving an overall leap forward in teaching quality and learning outcomes under technological empowerment.

References

- [1]Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research*, 63(2), 431-463.
- [2]Chen, M., & Lü, M. C. (2024). College English writing instruction in a ChatGPT context. *Contemporary Foreign Language Studies*, 1, 161-168.
- [3]Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 57.
- [4]Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435-8463.
- [5]Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, 126, 103502.
- [6]Huang, X. T., & Guo, L. T. (2024). The application of large language models in formative assessment: In the case of the scoring and feedback in English writing. *Education Research Monthly*, 7, 74-80.
- [7]Liu, M. J., Zou, W. L., Zong, M., & Yang, B. (2025). How generative artificial intelligence empowers teaching and learning: An empirical study on teaching English writing to middle school students. *China Educational Technology*, (6), 70-79.
- [8]Mao, Y. S., Wang, Y. H., & Xing, Y. R. (2024). An empirical study of ChatGPT assisting senior high school English writing feedback. *Educational Measurement and Evaluation*, (1), 3-13.
- [9]Ren, W., Liu, Y. B., & Xie, Y. (2024). A comparative study of ChatGPT and teacher feedback in English writing instruction. *Foreign Language Learning Theory and Practice*, 30-38.
- [10]Sun, P. J., Xu, J. Y., & Zhang, J. (2025). Exploring the value-added benefits of GAI feedback on English writing quality among Chinese university students. *Foreign Language Education Research Frontiers*, 8(4), 24-36.

- [11]Teng, L. , Yang, Y. X. , & Yang, J. (2025). The role of self-regulation capacity in multi-dimensional engagement with GAI writing feedback. *Foreign Language Education Research Frontiers*, 8(3), 85-96.
- [12]Wang, Y. B. , Xuan, Q. Y. , & Zhang, J. L. (2025). The impact of generative AI intervention on Chinese EFL students' engagement with writing feedback. *Foreign Languages in China*, 22(2), 71-78.
- [13]Wang, Y. H. , Mao, Y. S. , & Dong, Y. J. (2024). An empirical study of ChatGPT-assisted feedback on senior high school English reading-writing integrated continuation task. *Educational Measurement and Evaluation*, (5), 71-83.
- [14]Wei, S. , & Li, L. Y. (2023). Artificial intelligence-assisted feedback in second language writing: The case of ChatGPT. *Foreign Languages in China*, 20(3), 33-40.
- [15]Xu, C. G. (2025). Generative artificial intelligence-enabled human-AI collaborative writing engagement for foreign language learners. *Foreign Languages and Their Teaching*, (2), 61-73, 147.
- [16]Xu, L. L. , Hu, J. H. , & Su, Y. (2024). Learners' perceptions and behaviors in AI-assisted academic English writing. *Foreign Language World*, (3), 51-58.
- [17]Yan, D. , & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communications*, 11(1), 1-14.
- [18]Zeevy-Solovey, O. (2024). Comparing peer, ChatGPT, and teacher corrective feedback in EFL writing: Students' perceptions and preferences. *Technology in Language Teaching & Learning*, 6(3), 1482-1482.
- [19]Zhang, J. J. , Guo, Y. M. , & Zhu, S. H. (2025). Generative AI versus teacher feedback in middle school English writing instruction under AES technology: The case of Doubao AI. In *Proceedings of the 9th Symposium on Educational Examination and Evaluation (SEEE 2025)* (pp. 356 - 367).
- [20]Zhou, T. , Cao, S. , Zhou, S. , Zhang, Y. , & He, A. (2023). Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. *System*, 118, 103141.
- [21]Zou, S. , Guo, K. , Wang, J. , & Liu, Y. (2024). Investigating students' uptake of teacher-and ChatGPT-generated feedback in EFL writing: A comparison study. *Computer Assisted Language Learning*, 1-30.