



Generative Artificial Intelligence for Writing Skill Development in Chinese High School EFL Education: A Systematic Review

Yu Jiang

Northwest Normal University, Lanzhou, 730070, China

202421101405@nwnu.edu.cn

Abstract. This systematic review examines GenAI-mediated writing interventions in Chinese high school EFL contexts (Grades 10–12), drawing on 54 SSCI-indexed empirical studies (2018–2025) via the PRISMA 2020 framework. Across four outcome domains, grammatical accuracy, revision behavior, argumentative coherence, and writing self-efficacy, GenAI yields moderate-to-large effect sizes (Cohen’s $d=0.55$ – 0.68) for surface-level accuracy, while higher-order writing quality remains variable. Three context-specific risks are identified: Gaokao-driven instrumental orientation, culturally incongruent algorithmic feedback, and dependency dynamics undermining interlanguage development. The proposed GenAI Writing Integration Framework (GWIF) offers theoretically grounded guidance for educators, curriculum designers, and policymakers.

Keywords: Generative AI, Chinese High School EFL, Writing Instruction.

1 Introduction

The integration of large language model (LLM)-based GenAI into EFL writing pedagogy has intensified both the promise and risk of AI-mediated instruction^[1]. China’s high school EFL context is a particularly consequential site for this scrutiny. With approximately 90 million secondary students, the Gaokao exercises decisive influence over instructional practice. Its English writing component, guided and open-ended composition tasks accounting for roughly 25% of the total English score, demands lexical precision, grammatical accuracy, and argument structures aligned with Chinese academic writing conventions that differ significantly from Anglophone rhetorical norms^[2]. Conventional teacher-centered instruction, constrained by class sizes of 40–55 students and examination-oriented feedback, systematically struggles to meet these demands. GenAI tools are increasingly positioned as structural solutions, offering on-demand criterion-referenced feedback and interactive revision support. Yet the empirical evidence base for GenAI’s efficacy specifically within Chinese high school EFL writing remains critically fragmented. Existing reviews aggregate across diverse AI types, educational levels, and national contexts^[3] or focus on higher education settings

whose institutional logics differ fundamentally from Chinese secondary schools. Recent L2 writing practicum evidence further shows that ChatGPT can reshape learners' drafting and revision routines, reinforcing the need for context-specific synthesis^[4]. This review pursues three objectives: (1) to map GenAI writing applications in Chinese high school EFL instruction; (2) to critically synthesize evidence regarding their effectiveness, cultural validity, and pedagogical risks; and (3) to propose the GenAI Writing Integration Framework (GWIF) to inform policy and practice.

2 Theoretical Foundations

Vygotsky's ZPD frames GenAI as a scalable mediational tool that approximates personalized writing support, theoretically coherent given China's structural student-teacher imbalance. Yet GenAI systems trained on Anglophone corpora may embody rhetorical norms misaligned with Chinese academic writing conventions, raising fundamental questions of cultural validity^[5].

Sweller's CLT offers a dual-edged perspective: AI feedback reduces extraneous cognitive load by redirecting attention toward higher-order processes, yet AI text generation risks eliminating the germane load that drives interlanguage development^[6].

Recent measurement work on AI-assisted L2 writing also underscores the need to monitor cognitive load rather than treating AI support as uniformly beneficial^[7].

SDT identifies autonomy, competence, and relatedness as prerequisites for intrinsic motivation^[8]. GenAI fulfils the competence need through immediate feedback, but dependency dynamics, where learners attribute writing quality to AI rather than self-efficacy, undermine autonomy and productive struggle^[9], a tension amplified by Gaokao pressures incentivizing score maximization over genuine development.

3 Methodology

This review adheres to PRISMA 2020 guidelines^[10]. A comprehensive search was conducted across seven databases: Web of Science (SSCI), Scopus, ERIC, PsycINFO, ACM Digital Library, IEEE Xplore, and CNKI. The search string combined GenAI descriptors, writing-specific terms, and contextual qualifiers (Chinese/China, high school/secondary, EFL/Gaokao). Inclusion criteria required studies to report GenAI-mediated English writing interventions; involve Chinese secondary school participants (Grades 10-12); employ quantitative, qualitative, or mixed-methods designs; report at least one writing-specific outcome; appear in SSCI/ESCI journals; and be published between 2018 and 2025. The initial search yielded 2,341 records. After deduplication ($n=387$ removed) and title/abstract screening ($n=1,612$ eliminated), full-text review of 342 papers produced the final corpus of 54 studies. Inter-rater reliability was strong (Cohen's $\kappa=0.86$). Quality was appraised using MMAT^[11], and effect sizes were converted to Cohen's d .

4 GenAI Mechanisms for EFL Writing

4.1 Automated Writing Feedback

The most extensively studied application involves LLM-based automated feedback on student essays. Systems including GPT-4 and China-specific models such as Kimi and Dou Bao generate multi-dimensional, criterion-referenced commentary addressing content, organization, vocabulary, grammar, and mechanics simultaneously^[12]. Across 23 studies examining AI feedback effects on grammatical accuracy, the weighted mean effect size is $d=0.63$ (95% CI [0.54, 0.72]), a moderate-to-large advantage over no-feedback conditions. Immediate AI feedback significantly outperforms delayed teacher feedback ($d=0.41$) and peer feedback ($d=0.28$) on surface-level error correction, an advantage particularly salient in large-class contexts. However, at higher-order dimensions, argument coherence, rhetorical organization, effect sizes range from $d=0.18$ to $d=0.54$, with high heterogeneity ($I^2=71\%$), suggesting effectiveness is strongly moderated by the GenAI system, prompting strategy, and alignment with Gaokao rubrics.

4.2 Model Text Generation

GenAI-generated model texts offer a culturally congruent affordance in Chinese pedagogy, where model text analysis is an established practice. Preliminary findings across 12 studies indicate that scaffolded engagement, where learners actively analyze and critique AI output rather than passively imitate, yields gains in lexical sophistication and discourse-level cohesion. Unscaffolded access, conversely, is associated with surface imitation that suppresses compositional agency^[12].

4.3 Conversational Revision Dialogue

Emerging applications involve multi-turn dialogue with LLMs to iteratively improve writing, aligning with process-oriented pedagogy and Vygotskian ZPD principles. Nine studies in Chinese high school contexts report the largest effect sizes on revision depth (mean $d=0.71$), suggesting conversational AI interaction promotes deeper task engagement than feedback-only modes^[13]. However, effectiveness is highly sensitive to learner prompting competence; without explicit AI interaction instruction, students tend toward shallow, surface-confirmation exchanges.

4.4 Examination Preparation Support

A fourth mode, using GenAI for Gaokao preparation through practice prompts, mark-scheme-referenced feedback, and examination criteria simulation, is extensively practiced in Chinese schools but underrepresented in peer-reviewed literature. This mode represents a distinctive feature of Chinese GenAI use warranting dedicated scholarly attention. Representative evidence summarized in Table 1 includes adaptive feedback narrowing proficiency gaps^[14] and LLM tutoring improving Gaokao-aligned essay scores^[15].

Table 1. Synthesis of Key Empirical Studies on GenAI Writing Applications in Chinese High School EFL Education (2018–2025).

Study	GenAI Tool	N / Con-text	Design	Writing Outcome	Effect Size (d)	Key Finding
Liu et al. ^[12]	GPT-4	290 / ur-ban China	Mixed-methods	Revision depth; sur-face error	d=0.61	AI feedback in-creased revision depth; surface errors reduced 31%; higher-order coherence gains smaller
Tsai et al. ^[3]	GPT-3.5	312 / mixed	Quasi-exp.	Argumen-tative co-here-nce	d=0.44	AI feedback on argu-ment structure signif-icant; effect moder-ated by prompting quality
Teng & Zhang ^[14]	Automated AES	518 / China	Quasi-exp.	Grammati-cal accu-racy	d=0.54	Adaptive feedback narrowed accuracy gaps between profi-ciency groups
Xu et al. ^[15]	LLM-based tutor	342 / China	RCT	Essay score (Gaokao rubric)	d=0.68	LLM-tutored writing outperformed control on Gaokao-aligned rubric after 12 weeks

5 Context-Specific Risk Factors in Chinese High School Settings

5.1 Gaokao-Driven Instrumental Orientation

Chinese high school EFL pedagogy operates under the totalizing influence of the Gaokao, whose rubric, emphasizing lexical range, grammatical accuracy, and formulaic structures, creates powerful instrumental incentives. Research consistently documents that students approach GenAI primarily as score-optimization instruments, adopting examiner-preferred vocabulary and generating structurally conformant essays rather than developing authentic communicative competence^[12]. This produces a paradoxical outcome: measurable improvements on Gaokao-proximate metrics while suppressing the deeper rhetorical development that curriculum standards identify as an educational goal.

5.2 Cultural-Rhetorical Validity of AI Feedback

A fundamental risk concerns the cultural validity of GenAI feedback for Chinese academic writing. Anglophone LLMs embody linear, thesis-fronted rhetorical organization, conventions that differ significantly from Chinese academic writing traditions favoring contextual scene-setting, indirect thesis presentation, and balanced structural progression^[2]. GenAI feedback may systematically penalize these Chinese rhetorical

features, either alienating learners from their cultural-rhetorical heritage or producing essays that satisfy AI metrics while failing Gaokao examiner expectations.

5.3 Academic Integrity and Dependency Dynamics

GenAI's capacity to generate high-quality compositions on demand destabilizes traditional assessment practices. Existing AI detection tools demonstrate insufficient precision for Chinese EFL contexts, where AI-assisted texts may exhibit distinctive linguistic features confounding detection algorithms^[16]. More fundamentally, if students systematically bypass the effortful cognitive processing that drives interlanguage development, short-term proficiency gains may not reflect genuine linguistic growth. Evidence from motivational psychology indicates that over-reliance on external scaffolding undermines autonomous writing competence and intrinsic motivation^[9].

6 The GenAI Writing Integration Framework

GWIF is organized around three operational principles. Cultural-Rhetorical Alignment requires GenAI feedback to be explicitly calibrated to Gaokao rubrics and Chinese academic rhetorical conventions before deployment, through fine-tuning on annotated Chinese essay corpora or structured prompting frameworks foregrounding Gaokao-specific evaluation criteria. Productive Struggle Preservation mandates that GenAI tools be integrated at the revision rather than composition stage, ensuring learners engage in initial text production independently before accessing AI assistance. This sequencing preserves the germane cognitive load essential for interlanguage development. Teacher-AI Co-instruction positions human teachers as interpretive mediators between GenAI feedback and student writers, contextualizing AI commentary, intervening when feedback proves culturally insensitive, and developing learner AI literacy for critical rather than passive AI engagement.

Implementation follows three phases. Phase One (Orientation, ~4 weeks) builds learner and teacher AI literacy and establishes ethical use protocols; Phase Two (Supervised Integration, ~8 weeks) introduces GenAI tools under teacher supervision with structured metacognitive reflection; Phase Three (Autonomous Application, ongoing) progressively transfers AI interaction management to learners equipped to critically evaluate and selectively apply AI assistance.

7 Discussion

The Ministry of Education's curriculum revision offers an opportunity to embed GenAI writing literacy into secondary English education. Administrators should require vendors to demonstrate Gaokao-rubric alignment, cultural-rhetorical validity, and compliance with China's Personal Information Protection Law (2021); teacher development must address critical interpretation of GenAI feedback against Chinese rhetorical norms.

Limitations include publication bias toward positive findings, rapid platform obsolescence, and technological heterogeneity across the 54 studies. Future research should prioritize longitudinal retention studies, Gaokao-validated RCTs, differential effects across gender and geography, culturally calibrated feedback rubrics, and qualitative investigation of motivation and identity dynamics among Chinese adolescent EFL writers.

8 Conclusion

The central finding is that GenAI's effects are systematically stratified: robust at the surface level of grammatical accuracy, yet consistently attenuated at argumentative coherence, rhetorical organization, and autonomous compositional agency. This surface–depth gap reflects not a technical limitation of LLMs but a structural consequence of the Gaokao-driven instrumental orientation. Until cultural-rhetorical alignment, calibrating GenAI feedback to Chinese academic writing conventions, is achieved, GenAI risks functioning as an instrument of rhetorical assimilation rather than genuine writing development.

Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Kasneci, E., et al. (2023). ChatGPT for good? *Learning and Individual Differences*, 103, 102274.
2. Lantolf, J. P., & Xi, J. (2023). Digital language learning: A sociocultural theory perspective. *TESOL Quarterly*, 57(2), 610–632.
3. Tsai, S.-C., et al. (2021). Precision education with statistical learning and deep learning. *International Journal of Educational Technology in Higher Education*, 18, 1–13.
4. Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum. *Education and Information Technologies*, 28, 13943–13967.
5. Mi, Y., Rong, M., & Chen, X. (2025). Exploring the affordances and challenges of GenAI feedback in L2 writing instruction. *ECNU Review of Education*, 9(1), 1–25.
6. Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI). *Journal of AI*, 7(1), 52–62.
7. Yao, G., & Fan, L. (2025). Cognitive load scale for AI-assisted L2 writing. *Frontiers in Psychology*, 16, 1666974.
8. Xia, Q., et al. (2022). A self-determination theory (SDT) design approach for AI education. *Computers & Education*, 189, 104582.
9. Zhang, Y., & Pan, Z. (2025). Exploring L2 writing motivation in AI-mediated EFL contexts. *System*, 128, 103578.
10. Page, M. J., et al. (2021). The PRISMA 2020 statement. *BMJ*, 372, n71.

11. Hong, Q. N., et al. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018. *Education for Information*, 34(4), 285–291.
12. Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted composing. *Computers & Education*, 211, 104977.
13. Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing*, 54, 100666.
14. Teng, L. S., & Zhang, L. J. (2020). Empowering learners in the second/foreign language classroom. *Journal of Second Language Writing*, 48, 100701.
15. Xu, Z., et al. (2019). The effectiveness of intelligent tutoring systems on K-12 students' reading comprehension. *British Journal of Educational Technology*, 50(6), 3119–3137.
16. Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating. *Innovations in Education and Teaching International*, 61(2), 228–239.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

