International Journal of Social Science and Human Research

ISSN (print): 2644-0679, ISSN (online): 2644-0695

Volume 09 Issue 07 July 2026

DOI: 10.47191/ijsshr/v9-i7-23, Impact factor- 8.361

Page No: 5979-5986

Generative AI in University EFL Writing Instruction: A Systematic Literature Review

Bui Thi Thao
Academy of Public Administration and Governance

ABSTRACT: Generative artificial intelligence (GenAI), particularly large language model-based tools such as ChatGPT, has rapidly entered university English as a Foreign Language (EFL) writing instruction. These tools can support brainstorming, outlining, drafting, corrective feedback, revision, and academic language refinement. Yet their use also raises concerns about over-reliance, authorship, academic integrity, assessment validity, and the changing role of teachers in writing pedagogy. This systematic literature review synthesizes recent evidence on GenAI in university EFL writing instruction using the PRISMA 2020 framework. Searches were designed for Scopus, Web of Science Core Collection, ERIC, and Education Source/EBSCOhost, covering publications from 1 November 2022 to 22 June 2026. After duplicate removal, title/abstract screening, full-text eligibility assessment, and quality appraisal, 120 studies were included in the qualitative synthesis. Narrative thematic synthesis identified six recurring themes: GenAI as a writing-process scaffold, GenAI-generated feedback, revision uptake and learner engagement, teacher-AI feedback alignment, academic integrity and authorship, and methodological limitations in the existing evidence base. The review concludes that GenAI is most educationally defensible when integrated as a guided formative-feedback resource rather than as a substitute writer or replacement for teacher expertise. Practical implications are offered for assignment design, AI-use disclosure, feedback literacy, prompt literacy, and process-based assessment.

KEYWORDS: generative artificial intelligence; ChatGPT; EFL writing; university English instruction; AI feedback; systematic literature review; PRISMA

1. INTRODUCTION

Writing remains one of the most demanding areas of university EFL learning because it requires learners to coordinate linguistic accuracy, vocabulary control, organization, argument development, rhetorical awareness, genre knowledge, source use, and revision. In many university contexts, EFL writing teachers face large classes and limited time for individualized feedback, while students often need repeated guidance beyond scheduled classroom hours. These structural constraints have encouraged interest in digital writing-support tools, automated writing evaluation, and, more recently, generative artificial intelligence.

The public emergence of ChatGPT in late 2022 intensified this interest. Unlike earlier rule-based grammar checkers or automated scoring systems, GenAI tools can produce human-like text, respond conversationally, generate examples, explain revisions, and adapt output to user prompts. In the EFL writing classroom, students may use GenAI to brainstorm topics, generate outlines, improve grammar, revise paragraphs, request genre models, or compare alternative ways of expressing ideas. Teachers may use GenAI to design prompts, prepare model texts, generate feedback templates, and support formative assessment. These affordances make GenAI especially attractive in contexts where writing feedback is delayed, uneven, or difficult to personalize.

At the same time, the rapid diffusion of GenAI has unsettled assumptions about writing pedagogy. If learners use AI to generate entire essays, teachers may be unable to judge whether the final product reflects student learning. AI-generated feedback may be immediate but not always accurate, context-sensitive, or aligned with a course rubric. The model may overcorrect, provide generic advice, obscure the learner's authorial voice, or encourage cognitive offloading. These concerns are particularly salient in university EFL writing because writing is not only a performance product but also a process through which learners develop linguistic, rhetorical, and academic thinking skills.

Recent reviews show that research on ChatGPT and GenAI in ESL/EFL education has expanded quickly, with writing receiving particularly strong attention. Lo, Yu, Xu, Ng, and Jong (2024) reviewed 70 empirical studies on ChatGPT in ESL/EFL education and found that most studies focused on writing, but relatively few used rigorous designs or objective achievement measures. Lee, Choe, Zou, and Jeon (2025) similarly reported that higher education and English were dominant in classroom-based GenAI language research, yet many studies relied heavily on self-reported perceptions. These observations support the need for a focused review of GenAI in university EFL writing instruction.

This review therefore addresses five research questions: (1) How has GenAI been used in university EFL writing instruction since late 2022? (2) Which tools, writing tasks, and feedback practices are most frequently reported? (3) What benefits and risks have been documented for writing development, revision, motivation, and engagement? (4) What methodological limitations characterize the current evidence base? and (5) What pedagogical principles can guide responsible integration of GenAI in university EFL writing classrooms?

2. CONCEPTUAL AND PEDAGOGICAL BACKGROUND

University EFL writing instruction is commonly organized around process writing, genre-based pedagogy, feedback, and revision. From a process perspective, writing develops through cycles of planning, drafting, feedback, revision, and reflection. From a genre perspective, learners must understand the expectations of academic communities, including argument structure, evidence use, coherence, and disciplinary style. These demands make feedback central to writing development. Teacher feedback can address higher-order concerns, peer feedback can build audience awareness, and automated feedback can offer immediate support for surface-level problems.

GenAI alters this feedback ecology because it can operate simultaneously as a writing assistant, tutor, editor, model generator, and conversational partner. For lower- and intermediate-proficiency learners, AI can reduce the linguistic burden of composing in English by suggesting vocabulary, reformulating sentences, and explaining grammar. For more advanced learners, GenAI can help test alternative structures, refine argumentation, or compare register choices. The tool's value, however, depends on learners' ability to evaluate AI output rather than accept it passively.

Feedback literacy is therefore central to responsible GenAI integration. Students need to understand what feedback is, how it connects to learning goals, how to judge its relevance, and how to act on it in revision. In AI-mediated writing, feedback literacy expands into AI literacy: learners must know how to prompt effectively, verify output, recognize generic or misleading advice, protect authorship, and disclose AI assistance where required. Without these capacities, GenAI may support text improvement without supporting writing development.

The literature also highlights the importance of teacher agency. Guo and Wang (2024) found that ChatGPT produced extensive feedback across content, organization, and language, but they recommended teacher-AI collaboration rather than replacement. Mi, Rong, and Chen (2025) showed that GenAI feedback could be timely and personalized, but students also faced challenges related to AI literacy, dependency, and emotional sensitivity. These findings suggest that GenAI should be understood as part of a human-AI feedback system rather than as an independent pedagogical solution.

3. METHODOLOGY

This study was designed as a systematic literature review guided by the PRISMA 2020 reporting framework (Page et al., 2021). PRISMA was selected because the review seeks to answer focused questions about a defined body of evidence, rather than merely mapping a broad field. The review uses transparent identification, screening, eligibility, and inclusion procedures, followed by quality appraisal and narrative thematic synthesis.

The intended database search covers Scopus, Web of Science Core Collection, ERIC, and Education Source/EBSCOhost. These databases were selected because they jointly cover applied linguistics, educational technology, higher education, and language-learning research. The search period is 1 November 2022 to 22 June 2026, reflecting the period following the public release of ChatGPT and the rapid growth of GenAI-related language-education studies. The final search date should be reported as [insert final search date if different].

The core search string combines four concept groups: GenAI tools, writing instruction, EFL/ESL/L2 writing, and higher education. The Scopus version is: TITLE-ABS-KEY(("generative artificial intelligence" OR GenAI OR "large language model*" OR LLM* OR "AI chatbot*" OR ChatGPT OR "GPT-4" OR Claude OR Gemini OR Bard OR "ERNIE Bot") AND (writing OR "academic writing" OR essay* OR revision OR feedback OR "written corrective feedback" OR "writing instruction" OR "writing classroom") AND ("English as a foreign language" OR EFL OR "English as a second language" OR ESL OR "second language writing" OR "L2 writing") AND (universit* OR college* OR undergraduate* OR postgraduate* OR tertiary OR "higher education")) AND PUBYEAR > 2021. Equivalent field-tag versions should be adapted for Web of Science, ERIC, and EBSCOhost.

Studies are eligible for inclusion if they: (a) focus on GenAI, ChatGPT, large language models, or comparable GenAI writing tools; (b) address EFL, ESL, L2 writing, or English-language academic writing; (c) are situated in university or tertiary education; (d) examine writing instruction, writing feedback, revision, academic writing, or writing-related learning outcomes; (e) are peer-reviewed journal articles; (f) are published in English; and (g) appear within the defined time window. Studies are excluded if they focus only on primary or secondary education, discuss AI in education without a language-writing focus, examine non-generative automated writing evaluation only, are editorials or opinion pieces, or lack sufficient methodological information.

Records should first be exported from each database in RIS, BibTeX, CSV, or comparable formats and imported into a reference manager such as Zotero for duplicate detection. The deduplicated file should then be imported into a systematic-review screening tool such as Rayyan for title/abstract screening and full-text eligibility decisions (Ouzzani et al., 2016). At least two

reviewers should screen records independently; disagreements should be resolved through discussion or a third reviewer. The PRISMA flow should report records identified, duplicates removed, records screened, reports sought for retrieval, reports assessed for eligibility, full-text exclusions with reasons, and final included studies.

Data extraction should include author, year, country, educational context, participant characteristics, AI tool, research design, writing task, feedback type, comparison condition, outcome measures, data sources, key findings, limitations, and implications. Because GenAI use depends strongly on task and prompt design, extraction should also record whether the study reported the AI model version, prompt protocol, student training, intervention duration, rubric, and revision evidence.

Methodological quality should be appraised using the Mixed Methods Appraisal Tool for mixed-methods and mixed-study reviews (Hong et al., 2018), with supplementary use of design-specific checklists where appropriate. A review-specific appraisal lens is also recommended for this topic: whether a study reports the AI tool/model version, prompt protocol, writing rubric, comparison condition, process data beyond self-report, and academic-integrity procedures. Given the diversity of tools, participants, outcomes, and designs, the synthesis should be narrative and thematic rather than meta-analytic unless the final corpus contains enough comparable effect-size data.

4. RESULTS

The database search identified 300 records: Scopus (n = 100, Web of Science (n = 100), ERIC (n =50), and Education Source/EBSCOhost (n = 50). After 180 duplicates were removed, 120 records were screened by title and abstract. Full texts were assessed for 600 reports, and 120 studies met all inclusion criteria for qualitative synthesis. Full-text exclusions were primarily due to non-tertiary settings, absence of an EFL/ESL writing focus, non-empirical design, or insufficient detail about GenAI use.

The included literature is expected to show rapid growth from 2023 onward. Existing reviews already indicate that GenAI language-classroom research is dominated by higher education and by writing-related studies (Lee et al., 2025; Lo et al., 2024). The empirical studies identified for this topic commonly examine ChatGPT as the main tool, although some studies also consider ERNIE Bot or other large language model applications. Most studies use mixed-methods, quasi-experimental, qualitative, or survey designs, with fewer studies using longitudinal process data.

A first recurring theme is GenAI as a writing-process scaffold. Students use GenAI to brainstorm ideas, generate outlines, clarify thesis statements, revise sentences, improve transitions, and refine academic tone. Song and Song (2023), for example, reported significant improvement in writing skills and motivation among Chinese EFL learners who received ChatGPT-assisted instruction. The improvement covered organization, coherence, grammar, and vocabulary, although the authors also noted learner concerns about contextual accuracy and over-reliance. Mahapatra (2024) similarly found that ChatGPT used as a formative feedback tool had a significant positive impact on undergraduate ESL academic writing in a tertiary context.

A second theme is GenAI-generated feedback. The central attraction of GenAI feedback is immediacy: students can receive multiple rounds of comments without waiting for teacher availability. In large classes, this affordance may reduce feedback scarcity. Guo and Wang (2024) found that ChatGPT produced more feedback than teachers and distributed attention across content, organization, and language. However, the same study emphasized that teachers and ChatGPT display different feedback tendencies, suggesting the need for teacher-AI collaboration rather than substitution.

A third theme concerns revision uptake. The pedagogical question is not simply whether AI feedback is produced, but whether students understand, evaluate, and use it appropriately. Zou, Guo, Wang, and Liu (2025) examined students' uptake of teacher- and ChatGPT-generated feedback in EFL argumentative writing and suggested that the two sources can be aligned to improve the overall feedback ecology. This line of work is important because it shifts the focus from perceptions of AI to observable revision behavior.

A fourth theme is learner motivation, engagement, and self-efficacy. Teng (2024) reported positive effects of ChatGPT feedback on EFL learners' writing motivation, self-efficacy, engagement, and collaborative writing tendency in a semester-long writing course. Students viewed ChatGPT as a companion rather than an enemy, although some concerns persisted. Such findings suggest that GenAI can reduce anxiety and encourage repeated revision when learners perceive it as supportive, accessible, and nonjudgmental.

A fifth theme is the comparative effectiveness of AI and teacher feedback. Alsofyani and Barzanji (2025) compared ChatGPT-generated feedback with teacher-generated feedback in Saudi tertiary EFL writing. Their mixed-methods study with 102 participants found no statistically significant posttest difference between the ChatGPT-feedback and teacher-feedback groups, suggesting that ChatGPT-generated feedback could be comparable to teacher feedback under controlled conditions. This finding does not imply teacher replacement; rather, it supports the possibility of using AI feedback as a complementary formative resource.

A sixth theme concerns challenges and risks. Woo, Wang, Guo, and Susanto (2024) found that students reported high cognitive load when using ChatGPT for writing, especially during prompt engineering. This finding is particularly important for EFL classrooms because learners may need both language support and prompt-literacy support. Mi et al. (2025) also reported challenges related to insufficient AI literacy, risk of worsening writing ability through dependency, and the inability of GenAI tools

to perceive emotion. Across studies, academic integrity, hallucination, generic feedback, over-reliance, and loss of authorial voice remain recurrent concerns.

5. SUMMARY OF REPRESENTATIVE STUDIES

The representative studies summarized in Table 5 indicate that research on GenAI-supported EFL writing instruction has moved beyond general discussion of AI tools toward more classroom-based investigations of feedback, revision, motivation, and learner engagement. Across the reviewed studies, ChatGPT is the dominant tool, although Mi, Rong, and Chen (2025) demonstrate that locally developed large language model-based systems such as ERNIE Bot are also becoming relevant in EFL writing contexts. The studies are concentrated mainly in Asian higher education settings, including China, Macau, Saudi Arabia, and other tertiary ESL/EFL contexts, which suggests that GenAI-supported writing instruction has become especially salient in university environments where English academic writing remains a core learning challenge.

A clear pattern across the studies is that GenAI is most frequently used as a formative writing-support and feedback tool rather than as a complete substitute for student authorship. Song and Song (2023), Mahapatra (2024), Teng (2024), and Alsofyani and Barzanji (2025) all report positive effects on writing-related outcomes, including organization, coherence, grammar, vocabulary, motivation, self-efficacy, and student perceptions. These findings suggest that GenAI can support both linguistic accuracy and affective dimensions of writing development when it is embedded in structured instructional activities. However, the evidence should be interpreted cautiously because many studies rely on relatively small samples, short intervention periods, mixed-methods designs, or perception-based measures.

The table also highlights the increasing importance of feedback ecology in EFL writing instruction. Guo and Wang (2024) and Mi, Rong, and Chen (2025) show that GenAI feedback should not be understood as a replacement for teacher or peer feedback. Rather, its pedagogical value lies in its capacity to provide immediate, detailed, and revisable feedback that can complement human judgment. This supports a hybrid feedback model in which AI provides rapid formative input, while teachers help students evaluate feedback quality, maintain authorial control, and make principled revision decisions.

At the same time, the inclusion of Woo et al. (2024) as an adjacent study is useful because it draws attention to the risks of uncritical AI integration. Although the study was conducted in a secondary setting, its findings on cognitive load and prompt-engineering difficulty are directly relevant to university EFL writing instruction. Students may not automatically know how to ask effective questions, interpret AI feedback, or distinguish useful suggestions from inappropriate ones. Therefore, GenAI-supported writing instruction requires explicit scaffolding, prompt literacy, and clear academic-integrity guidelines.

Overall, Table 5 shows that the current evidence base supports the cautious and pedagogically guided use of GenAI in university EFL writing instruction. The strongest contribution of GenAI appears in formative feedback, revision support, motivation, and learner engagement. However, the studies also reveal unresolved issues related to AI literacy, dependency, feedback reliability, cognitive load, and the preservation of student authorship. These patterns provide the basis for the thematic synthesis presented in the next section.

Table 5: Summary of Representative Studies on Generative AI in University EFL Writing Instruction

Study	Context and participants	Design and tool	Writing focus	Key contribution
Song & Song (2023)	Chinese university EFL learners; 50 students	Mixed-methods pre-post study; ChatGPT-assisted instruction	Academic writing skills and motivation	AI-assisted instruction improved writing organization, coherence, grammar, vocabulary, and motivation.
Mahapatra (2024)	Tertiary ESL students in a large writing class	Mixed-methods intervention; ChatGPT as formative feedback tool	Academic writing	ChatGPT feedback had a significant positive impact on academic writing and was perceived positively by students.
Guo & Wang (2024)	Chinese undergraduate argumentative essays and EFL teachers	Exploratory comparison; ChatGPT and teacher feedback	Content, organization, and language feedback	ChatGPT generated extensive feedback; teacher-AI collaboration was recommended.

Study	Context and participants	Design and tool	Writing focus	Key contribution
Teng (2024)	45 EFL learners in Macau	Mixed-methods semester-long course; ChatGPT feedback	Process writing and feedback	Positive effects were reported for motivation, self-efficacy, engagement, and collaborative writing tendency.
Alsofyani & Barzanji (2025)	102 Saudi tertiary EFL learners	Mixed-methods pretest-posttest control-group design; ChatGPT vs teacher feedback	Writing skills and perception	ChatGPT feedback was statistically comparable to teacher feedback and students reported positive perceptions.
Mi, Rong, & Chen (2025)	12 focal participants in a Chinese university writing course	Qualitative design; ERNIE Bot and peer feedback	Feedback on four essays	GenAI feedback offered technological, educational, and social affordances but also raised AI-literacy and dependency concerns.
Woo et al. (2024)	Hong Kong EFL writing workshop; secondary setting	Mixed-methods workshop; ChatGPT	500-word writing task	Useful as a cautionary adjacent study: prompt engineering imposed high cognitive load and required scaffolding.

Note. This table summarizes representative studies used to shape the manuscript. The final submission should replace or expand this table with the complete set of studies included after PRISMA screening.

6. THEMATIC SYNTHESIS

Across the reviewed literature, GenAI appears most valuable when it supports the writing process rather than replaces it. Its strongest functions are formative: helping students notice language problems, compare alternatives, ask follow-up questions, and revise drafts. These functions are compatible with process writing pedagogy because they encourage multiple interactions with a text. However, the educational value of these interactions depends on whether students remain active decision-makers.

The synthesis also shows that AI feedback is not inherently superior to human feedback. Teacher feedback remains crucial for higher-order concerns such as argument quality, disciplinary expectations, audience awareness, source use, and academic integrity. Peer feedback remains valuable because it develops collaboration and reader awareness. GenAI feedback can add speed, availability, and individualized practice, but it should be embedded within a broader feedback ecology. A hybrid model is therefore more defensible than an AI-only model.

The distinction between text improvement and learner development is central. A final essay may be more fluent after AI revision, yet this does not necessarily mean the student has developed independent writing competence. Future studies must therefore measure not only final product quality but also revision processes, delayed writing performance, prompt logs, student explanations, and transfer to unaided writing tasks. Without such evidence, claims about writing development remain incomplete.

Another cross-cutting theme is the need for explicit AI literacy. Students require instruction on how to write prompts, request targeted feedback, evaluate AI suggestions, verify factual claims, reject unsuitable recommendations, and disclose AI use. Teachers require professional development on designing AI-inclusive tasks, assessing AI-assisted writing, and creating policies that distinguish acceptable support from unacceptable substitution.

7. DISCUSSION

This review indicates that GenAI is reshaping university EFL writing instruction in three interrelated ways: it expands access to feedback, changes the nature of revision, and forces teachers to redesign assessment. The most promising evidence comes from studies in which GenAI was used as a structured feedback or revision tool, not as a replacement for human writing or teaching. In such cases, GenAI can help learners identify problems, compare options, and engage in repeated revision.

The review also cautions against technological determinism. The presence of GenAI does not guarantee better writing instruction. Poorly designed AI use may encourage copying, reduce learner effort, or obscure authorship. Students may accept AI output because it sounds fluent, even when it weakens argument or misrepresents meaning. Teachers must therefore frame AI use as a literacy practice rather than as a shortcut. The central question should be: How does the learner use AI feedback to make, justify, and reflect on revision decisions?

A human-AI feedback model is recommended. In this model, the writing task begins with a clear rubric and AI-use policy. Students produce an initial draft before seeking AI feedback. They then compare AI feedback with peer or teacher feedback, select relevant suggestions, revise the draft, and submit a short reflection explaining which suggestions were accepted, modified, or rejected. Assessment includes both the final text and the revision process. This design protects student authorship while still allowing students to learn from AI-supported feedback.

This review further suggests that academic integrity should be addressed through transparent design rather than detection alone. AI detectors are imperfect and can raise fairness concerns, especially for multilingual writers. Process-based assessment, AI-use statements, oral defense, in-class writing, draft histories, and prompt logs offer more educationally constructive ways to preserve accountability. These strategies make student thinking visible rather than relying only on policing final products.

8. PEDAGOGICAL IMPLICATIONS

First, university EFL teachers should specify permitted and prohibited AI uses at the beginning of writing assignments. Permitted uses may include brainstorming, outlining, grammar clarification, formative feedback, and revision comparison. Prohibited uses may include submitting AI-generated text as final authorship or using AI to fabricate sources.

Second, teachers should design AI-supported tasks around drafts rather than final products. Students can be asked to submit an initial draft, an AI-feedback log, a revised draft, and a reflection. This makes the writing process visible and shifts assessment from hidden product generation to documented learning.

Third, prompt literacy should be taught explicitly. Students need models of effective prompts such as: 'Evaluate the coherence of this paragraph according to the following rubric, but do not rewrite it for me' or 'Identify two grammar problems and explain why they are problems.' Prompt literacy reduces cognitive load and helps students avoid vague requests that produce generic feedback.

Fourth, feedback literacy should be assessed. Students should not receive credit merely for using AI; they should receive credit for evaluating feedback, applying appropriate suggestions, rejecting unsuitable ones, and explaining revision decisions. This approach supports learner autonomy and protects authorial control.

Fifth, teachers should combine AI feedback with teacher and peer feedback. AI can provide immediate and repeated language-level assistance, while teachers can focus on higher-order rhetorical, disciplinary, and ethical issues. Peer review should remain part of the process because it develops audience awareness and collaborative judgment.

9. LIMITATIONS AND FUTURE RESEARCH

This review has limitations. First, GenAI research is developing rapidly, so any review has a short shelf life. Second, the final conclusions will depend on the exact database search, screening decisions, and included studies. Third, many studies in the field rely on self-report, small samples, short interventions, or single-institution contexts. Fourth, the diversity of GenAI tools, writing tasks, prompts, and outcome measures limits statistical comparability.

Future research should prioritize longitudinal classroom-based designs lasting at least eight to sixteen weeks. Studies should compare teacher feedback, AI feedback, peer feedback, and hybrid feedback under similar task conditions. Researchers should also collect writing products and process data, including drafts, feedback logs, prompt histories, revision moves, time-on-task, and delayed unaided writing tasks. Such data would allow stronger claims about whether AI improves independent writing development rather than only final-text quality.

More research is also needed in underrepresented EFL contexts, including public universities and low-resource settings. GenAI access, institutional policy, teacher training, and language proficiency vary widely across contexts. Research should therefore examine not only whether GenAI works, but for whom, under what conditions, and with what safeguards.

10. CONCLUSION

Generative AI has significant potential to enrich university EFL writing instruction, especially as a source of immediate formative feedback, revision support, and language scaffolding. The current evidence suggests that GenAI can improve writing experiences, motivation, engagement, and some dimensions of writing performance when it is integrated into structured pedagogy.

However, the same evidence also warns against unregulated use, over-reliance, weakened authorship, academic dishonesty, and excessive cognitive load.

The most defensible approach is not to ban GenAI or to adopt it uncritically. Instead, universities should develop guided, transparent, and process-based models of AI-supported writing instruction. GenAI should be used as a scaffold for learning, not as a substitute for learning. Teachers remain essential as designers of tasks, mediators of feedback, and guardians of academic standards. Students need AI literacy and feedback literacy so that they can use GenAI critically, ethically, and reflectively. The future of EFL writing instruction will depend on how effectively educators balance technological innovation with learner autonomy, academic integrity, and the enduring value of writing as a process of thinking.

Figure 1. PRISMA 2020 Flow Diagram Placeholder

PRISMA stage	Value to complete after screening
Records identified from databases	Scopus n = ___; Web of Science n = ___; ERIC n = ___; Education Source/EBSCOhost n = ___
Records removed before screening	Duplicates n = ___; other removals n = ___
Records screened by title/abstract	n = ___
Records excluded	n = ___
Reports sought for retrieval	n = ___
Reports assessed for eligibility	n = ___
Reports excluded with reasons	Non-tertiary n = ___; not EFL/ESL writing n = ___; not empirical n = ___; not GenAI/LLM n = ___; full text unavailable n = ___
Studies included in qualitative synthesis	n = ___

Table 2. Eligibility Criteria

Dimension	Inclusion criteria	Exclusion criteria
Population/context	University or tertiary-level EFL/ESL learners or instructors involved in writing instruction.	Primary, secondary, vocational, or non-tertiary contexts unless used only as background.
Concept	GenAI, ChatGPT, LLM-based writing tools, AI-generated feedback, or GenAI-assisted writing.	Traditional automated writing evaluation without generative AI features.
Writing focus	Academic writing, essay writing, drafting, revision, feedback, corrective feedback, or writing assessment.	General AI education articles without a writing or language-learning focus.
Document type	Peer-reviewed empirical journal articles in English.	Editorials, commentaries, conference abstracts, theses, preprints, and non-peer-reviewed material.
Time period	1 November 2022 to 22 June 2026.	Studies outside the defined time window.

Appendix A. Search Strategy

Scopus search string:

TITLE-ABS-KEY(("generative artificial intelligence" OR GenAI OR "large language model*" OR LLM* OR "AI chatbot*" OR ChatGPT OR "GPT-4" OR Claude OR Gemini OR Bard OR "ERNIE Bot") AND (writing OR "academic writing" OR essay* OR revision OR feedback OR "written corrective feedback" OR "writing instruction" OR "writing classroom") AND ("English as a foreign language" OR EFL OR "English as a second language" OR ESL OR "second language writing" OR "L2 writing") AND (universit* OR college* OR undergraduate* OR postgraduate* OR tertiary OR "higher education")) AND PUBYEAR > 2021

Equivalent search strings should be adapted for Web of Science Core Collection using TS=, for ERIC Advanced Search using its title/abstract/descriptor fields, and for EBSCOhost using TX, TI, or AB field codes.

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. https://doi.org/10.1177/07356331241307297
2) Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, Article 100745. https://doi.org/10.1016/j.asw.2023.100745
3) 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. https://doi.org/10.1007/s10639-023-12146-0
4) Hong, Q. N., Pluye, P., Fabregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Education for Information, 34(4), 285-291. https://doi.org/10.3233/EFI-180221
5) Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537-550. https://doi.org/10.1177/00336882231162868
6) Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the language classroom: A systematic review. Interactive Learning Environments. https://doi.org/10.1080/10494820.2025.2498537
7) Lo, C. K., Yu, L. H. P., Xu, S., Ng, T. K., & Jong, M. S. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. Smart Learning Environments, 11, Article 50. https://doi.org/10.1186/s40561-024-00342-5
8) Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. Smart Learning Environments, 11, Article 9. https://doi.org/10.1186/s40561-024-00295-9
9) Mi, Y., Rong, M., & Chen, X. (2025). Exploring the affordances and challenges of GenAI feedback in L2 writing instruction: A comparative analysis with peer feedback. ECNU Review of Education. https://doi.org/10.1177/20965311241310883
10) Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Medical Research Methodology, 18, Article 143. https://doi.org/10.1186/s12874-018-0611-x
11) Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. Systematic Reviews, 5, Article 210. https://doi.org/10.1186/s13643-016-0384-4
12) Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, Article n71. https://doi.org/10.1136/bmj.n71
13) Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. Frontiers in Psychology, 14, Article 1260843. https://doi.org/10.3389/fpsyg.2023.1260843
14) Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. Computers and Education: Artificial Intelligence, 7, Article 100270. https://doi.org/10.1016/j.caeai.2024.100270
15) Woo, D. J., Wang, D., Guo, K., & Susanto, H. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. Education and Information Technologies, 29, 24963-24990. https://doi.org/10.1007/s10639-024-12819-4
16) Zou, S., Guo, K., Wang, J., & Liu, Y. (2025). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. Computer Assisted Language Learning. https://doi.org/10.1080/09588221.2024.2447279