

AI-Powered Writing Tools in ESL/EFL Contexts: A Systematic Review of Empirical Studies

Seyede Motahare Mousavi ¹ , Forough Amirjalili ^{1*} 

¹ Department of English Language and Literature, Yazd University, Yazd, Iran



10.22080/iselt.2026.31653.1189

Received

May 2, 2026

Accepted

July 21, 2026

Available online

August 1, 2026

Keywords:

Artificial Intelligence, ChatGPT, Academic Writing, Feedback, ESL/EFL.

Abstract

Recent advances in generative artificial intelligence (AI) and automated writing tools (e.g., ChatGPT, Grammarly, QuillBot, Poe, etc.) have accelerated AI integration into writing instruction. This PRISMA-informed systematic review examined empirical studies from Scopus and Web of Science (January 2019–May 2025) using explicit screening, selection, and appraisal criteria. From 312 sources, 46 peer-reviewed studies were synthesized narratively. Findings indicated that AI-based writing tools enhance linguistic fluency, cohesion, coherence, and accuracy. Several studies also reported higher student motivation and reduced writing anxiety. The quasi-experimental research showed that AI-generated corrective feedback such as Poe could lower anxiety and improve linguistic quality more effectively than the teacher feedback alone (Wang, 2024). Comparative studies confirmed the benefits of AI feedback for fluency, accuracy, and syntactic complexity. Learners generally expressed positive attitudes toward AI writing aids, appreciating their immediacy and support for autonomous learning. Nonetheless, concerns persisted about text homogenization, superficial engagement with automated suggestions, plagiarism risks, ethical dilemmas, and threats to assessment integrity. Overall, empirical evidence suggests that integrating AI into writing pedagogy offers measurable linguistic and affective gains, but requires careful pedagogical guidance and ethical consideration to ensure authentic learning outcomes and maintain academic integrity.

1. INTRODUCTION

Importance of Writing at an Academic Level

Academic writing is widely recognized as a crucial determinant of academic success and professional achievement, particularly for students writing in a second or foreign language (Mahapatra, 2024). It requires the integration of linguistic resources (e.g., grammar and lexis), discourse-level organization, argumentation, and higher-order cognitive skills such as critical

* **Corresponding Author:** Forough Amirjalili, Assistant Professor, Department of English Language and Literature, Yazd University, Yazd, Iran, **Email:** f.amirjalili@yazd.ac.ir



thinking and creativity (Graham, 2019; Hyland, 2019), and is often identified as the most challenging skill for ESL and EFL students because it demands both linguistic competence and familiarity with academic rhetorical conventions (Mahapatra, 2024) — making learner support throughout the writing process a central pedagogical concern in second language (L2) education.

The Growing Use of AI in Writing Education

To address these challenges in ESL and EFL writing instruction, artificial intelligence (AI) has become an increasingly influential technology for supporting teaching and learning writing, with generative AI (GenAI) systems and automated writing evaluation tools such as ChatGPT, Grammarly, and QuillBot increasingly explored for their potential to support idea generation, organization, and linguistic accuracy (Bozorgian & Rahimi, 2025). Empirical studies report that such tools support idea generation and text organization, improve linguistic accuracy and coherence, and foster self-directed learning across English-as-a-second/foreign-language (ESL/EFL) contexts (Ardiana et al., 2025; Khany & Barzan, 2025; Mahapatra, 2024; Talapngoen & Thuratham, 2025). However, a systematic review of university-level studies found that while AI-assisted writing tools improve linguistic quality and feedback, the evidence remains largely exploratory (Pérez et al., 2025).

Perspectives of Students on AI Use in Writing

Student perceptions are an important mediator in how AI-based tools are adopted in the learning environment. Across most reviewed studies, learners report generally positive attitudes toward AI-assisted writing tools, valuing their accessibility, immediacy, and support for idea generation, organization, and revision (Al-Raimi et al., 2024; Bristi et al., 2025). Both cognitive and affective attitudes significantly predict learners' intentions to use ChatGPT for academic writing, with affective attitudes often the stronger predictor of adoption (Bristi et al., 2025). Alongside these benefits, however, students also report concerns about overreliance, reduced human interaction, and the risk that AI-generated text may be viewed as academic cheating, underscoring calls for process-based assessment approaches that reduce incentives for unethical use (Nelson et al., 2025; Zakarneh et al., 2025). Learners also tend to engage with AI-generated feedback superficially, accepting suggested changes without critically evaluating or incorporating them. AI-produced responses, while efficient at the sentence level, often fall short emotionally — pointing to the importance of feedback literacy in AI-assisted writing instruction (Wang et al., 2024; Zhan & Yan, 2025).

Innovative and Generative Mindsets

The relationship between AI and creativity has become a focal theme in the literature, with researchers examining both AI's potential to enhance idea generation and its effects on originality and creative expression. Liu et al. (2024), for example, found that ChatGPT can facilitate short-term improvements in idea generation and creative task performance, but cautioned that prolonged reliance on AI may homogenize students' ideas and reduce originality over time. Conversely, Gao et al. (2025) found that AI-based brainstorming enhanced fluency, flexibility, and hypothesis generation among Early Childhood Education students, particularly in linguistic construction, but produced no significant gains in critical thinking or self-efficacy.

Specialized Educational Settings

AI-based writing systems have also been employed in specialized educational settings. [Fung et al. \(2026\)](#) for example, introduced the CHATTING system to support students with dyslexia, reporting improvements in behavioral, emotional, and cognitive engagement. In English for Specific Purposes (ESP) education, ChatGPT-based instruction significantly improved the writing skills of Saudi university students, with female learners showing greater gains than male counterparts for reasons that remain unexplored. The AI KAKU system developed by [Gayed et al. \(2022\)](#) similarly reduced students' cognitive load by automating lower-level writing processes, allowing them to concentrate more effectively on organization and revision.

Ethical Considerations

Despite the positive role of AI in writing education, ethical concerns persist. Educators remain reluctant to endorse AI use for high-stakes academic assignments, citing risks of plagiarism and authorship issues ([Titko et al., 2023](#)). Consistent with the review noted earlier, [Pérez et al. \(2025\)](#) also cautioned that the evidence remains largely exploratory, raising concerns about academic integrity, data privacy, and AI's limits with complex writing errors. AI-generated text detection tools show high accuracy, but may unfairly flag non-native speakers' writing as AI-generated, since their linguistic characteristics can resemble AI-generated content ([Campino, 2024](#)). A scientometric review of ChatGPT-related research highlights clustering around ethical issues and educational innovations, with recurring attention to hallucinations, plagiarism, and authorship, calling for precautionary measures ([Lendvai, 2025](#)). [Kwon & Lee \(2025\)](#) further found that process-based approaches outperform product-based approaches in detecting AI-generated content.

Research Queries

Although prior research has examined individual AI tools, specific educational contexts, and isolated learning outcomes, this evidence remains fragmented across writing performance, learner engagement, creativity, and ethical concerns, with limited comprehensive synthesis specifically focusing on empirical ESL/EFL writing studies. A systematic review of this literature is, therefore, needed to clarify not only whether AI-assisted writing tools improve writing outcomes, but also under what pedagogical conditions their benefits and risks emerge.

Three research questions guide this systematic review:

1. What does the empirical literature reveal about the impact of AI-assisted writing tools on writing quality, learner performance, and writing assessment?
2. What do the reviewed empirical studies reveal about students' and teachers' engagement with and perceptions of AI-generated feedback in writing instruction?

What pedagogical, creative, and ethical tensions are reported in empirical studies on AI use in writing education?

2. METHODOLOGY AND DESIGN

Information sources and search strategy

A systematic literature search was conducted to identify peer-reviewed empirical studies investigating the use of AI-assisted writing tools in ESL/EFL writing instruction. The search was performed in Scopus and Web of Science Core Collection databases because they are widely

recognized for their comprehensive coverage of high-quality research in applied linguistics, language education, educational technology, and related disciplines.

The search strategy was developed by combining keywords related to AI and writing instruction using Boolean operators. The primary search terms included *artificial intelligence*, *generative AI*, *ChatGPT*, *Grammarly*, *QuillBot*, *automated writing evaluation*, *AI-assisted writing*, *writing instruction*, *academic writing*, *writing feedback*, *ESL*, *EFL*, *second language*, and *foreign language*. Database-specific search strings were adapted to the indexing requirements of each database.

The database searches were conducted on 15 May 2025. Results were limited to English-language, peer-reviewed journal articles published between 1 January 2019 and 15 May 2025. This time frame was selected because it captures the rapid development and widespread educational adoption of generative AI technologies and AI-assisted writing tools, particularly following the emergence of large language models and automated writing support systems in language education. Restricting the search to this period enabled the review to synthesize the most current empirical evidence reflecting contemporary AI technologies and pedagogical practices.

To ensure the comprehensiveness of the review, the reference lists of all included studies were manually examined to identify any additional eligible publications that were not retrieved through the database searches. Duplicate records were removed prior to the screening process, and all retrieved records were subsequently screened according to the predefined inclusion and exclusion criteria described in the following section.

Eligibility criteria

Eligible studies included empirical research employing experimental, quasi-experimental, mixed-methods, or well-defined observational designs. To be included, studies had to investigate the use of AI-assisted writing tools in writing instruction and report empirical findings on at least one outcome relevant to the research questions of the review. Moreover, the studies included were required to offer measurable evidence of the effects of using such AI-based instruments.

Non-empirical articles (e.g., conceptual papers, editorials) and studies not exploring writing instruction, not measuring writing outcomes, or being duplicates were excluded from further analysis.

Screening and selection

Titles and abstracts were independently screened by two reviewers (A and B) according to the predefined eligibility criteria. Reviewer A is a faculty member holding a PhD degree in Applied Linguistics with research experience and publications in AI and education. Reviewer B is a graduate researcher whose master's thesis focused on AI in education. The same two reviewers independently assessed the full-text articles for eligibility. Any disagreements were resolved through discussion and, when necessary, consultation with a third reviewer (C), a faculty member holding a PhD degree in Applied Linguistics with expertise in AI and education.

Differences were sorted out by discussions with another independent reviewer (C) when necessary. A PRISMA 2020 flowchart of the process is provided in [Figure 1](#).

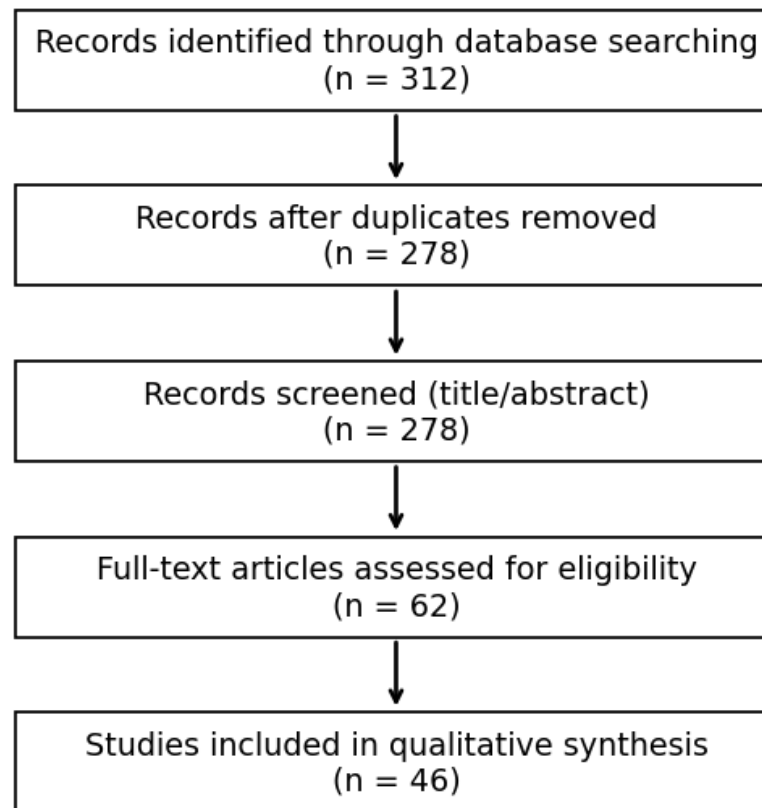


Figure 1: PRISMA 2020 flow diagram of the study selection process

Note. The diagram presents the identification, screening, eligibility, and inclusion of studies in this systematic review (N = 46). Records were retrieved from Scopus and Web of Science (n = 312), screened after duplicate removal (n = 278), assessed at full text (n = 62), and included in the final synthesis (n = 46).

Data Extraction

Study-level data were extracted using a uniform template, which included authors, year, country, sample size (n), participants' characteristics, study design, control group, AI-assisted writing tools used, intervention details, outcomes/measurement instruments, numerical data on outcomes (mean, standard deviation (SD), test statistic, *p*-value), and funding/COI. Two independent reviewers extracted data from studies, and disagreements were solved through discussion and a third reviewer.

Quality assessment

The methodological quality of the included studies was assessed using the 2018 version of the Mixed Methods Appraisal Tool (MMAT). Items in the MMAT domains relevant to the specific design used in the studies were rated for every study under review. MMAT gives an assessment of the item-level judgments of the study along with an overall judgment category (high, medium, or low) without aggregating into one numerical score. Inter-rater reliability for the MMAT assessments was evaluated by calculating Cohen's kappa coefficient based on the independent ratings of the two reviewers. The resulting kappa coefficient was 0.78, indicating substantial agreement.

Data Synthesis

Considering the substantial heterogeneity across studies in terms of research design, participant characteristics, AI-assisted writing tools, outcome measures, and reporting practices, the findings were synthesized using a narrative synthesis approach following the guidance of [Popay et al. \(2006\)](#). Extracted data were coded and grouped into thematic categories corresponding to the review research questions, including writing performance, learner engagement and affective outcomes, creativity, feedback models, and ethical considerations. Descriptive quantitative summaries (e.g., frequencies and percentages of studies reporting particular outcomes) were used to support the narrative synthesis where appropriate. No formal meta-analysis was conducted because the included studies lacked sufficient methodological and statistical comparability for reliable quantitative pooling.

3. RESULTS

AI Writing Assistants and Feedback Tools

From all the reviewed literature, it is clear that there is a recurring trend of improvement in writing abilities due to the use of AI writing aids for students. ChatGPT was found to be one of the most analyzed platforms, followed by Grammarly and QuillBot. The qualitative analysis conducted with pre-service teachers indicates that ChatGPT can provide an interactive experience for feedback generation; however, there might be inconsistencies or sensitive prompts, indicating the need for teacher training before using such technology-based auto-feedback tools ([Kurt & Kurt, 2024](#)). Compared to control groups receiving conventional teaching techniques, quasi-experiments and experiments showed statistically significant differences in favor of AI-enhanced writing outcomes post-test scores among students ([Khany & Barzan, 2025](#)).

While drafting and revising texts, students could detect grammar mistakes, choose better vocabulary, and improve sentence structures due to instant and personalized feedback provided by AI-assisted writing tools (several sources mentioned). Case-study evidence indicates ChatGPT can support students' development of formal register knowledge in writing tasks, although its usefulness appears more limited for informal or neutral register learning ([Punar Özçelik & Yangın Ekşi, 2024](#)). Similarly, Grammarly can assist learners in becoming more aware of grammatical and stylistic conventions, thereby improving the accuracy and complexity of their writing. QuillBot has likewise been found to support paraphrasing skills and promote learner autonomy, particularly among university-level EFL learners ([Talapngoen & Thuratham, 2025](#)).

In spite of these findings, however, several studies have also cautioned against relying on AI feedback. Classroom evidence indicates that ChatGPT's feedback can be helpful, but is sometimes inconsistent and tends to focus on accuracy over message-level or affective feedback—so AI feedback is pedagogically useful only with teacher monitoring ([Alghannam, 2024](#)). In particular, when students relied on AI feedback, they tended to focus on surface features of writing such as grammar and vocabulary and neglect other essential aspects of writing such as argumentation and coherence. The empirical data supports the idea that the use of combined feedback approaches, which combine AI with the feedback of teachers, is the most pedagogically effective. Design-based practices involving the use of Writing Workshop theory alongside AI-assisted writing tools such as Grammarly and ChatGPT have proven to be scalable and effective in class settings, resulting in better grammar, cohesion, organization, and self-assurance, especially in projects ([Jaramillo et al., 2025](#)).

Table 1: Statistical Tendencies of AI-Assisted Writing Outcomes

Outcome Variable	% of Studies Reporting Positive Effect	Typical <i>p</i> -value Range	Reported Effect Magnitude (Approx.)
Grammatical accuracy	~78% (36/46)	$p < .05$ to $p < .001$	Small–Moderate ($d \approx 0.30$ – 0.65)
Textual coherence	~70% (32/46)	$p < .05$	Small–Moderate
Organization	~67% (31/46)	$p < .05$	Small
Lexical complexity	~59% (27/46)	$p < .10$ to $p < .05$	Small
Writing fluency	~63% (29/46)	$p < .05$	Small–Moderate

Note. Percentages use $N = 46$ (the number of included empirical studies) as the denominator unless otherwise noted. “Typical *p*-value range” summarizes the range of inferential statistical significance reported across the original studies. Where original papers did not report *p*-values but provided effect-size estimates, the effect-size column was used for classification. Reported effect-size ranges (e.g., $d \approx 0.30$ – 0.65) are approximate and were derived from original-study statistics where available. When exact effect sizes were not reported, they were estimated from reported test statistics (*t*, *F*) using standard conversion formulas. These aggregated estimates should be interpreted with caution because of heterogeneity in outcome measures, study designs, and educational contexts.

In many experimental and quasi-experimental research studies, most researchers found statistically significant improvements in learner outcomes after using AI-assisted writing tools as opposed to control groups (i.e., no AI). The *p*-values were also found to be most commonly between .05 and .001, with effect sizes typically being small to moderate. What these studies show is that all researchers (with few exceptions) have shown that AI-supported writing tools have a positive effect on learners overall, and that the magnitude and direction of these effects vary greatly depending on the context in which they are used and the instructional design.

Students' engagement and attitudes towards using AI

As demonstrated across diverse educational settings, learners have generally reported positive attitudes toward the use of AI-assisted writing tools and have actively engaged with these technologies to support their writing development (Barrett & Pack, 2023; Mahapatra, 2024; Syafei & Nuraeningsih, 2025). Several studies also suggest that AI-assisted writing tools can enhance learners' motivation, reduce writing anxiety, and strengthen their confidence as writers by providing timely, personalized support and opportunities for continuous revision (Apriani et al., 2025; Syafei & Nuraeningsih, 2025; Zhou et al., 2024). In addition, many learners have found the timeliness and accessibility of feedback from AI to be beneficial, especially when teaching large classes or when time is of the essence. Evidence from the classroom indicates that many students participate in multiple prompt/revise cycles with ChatGPT, using its template generation and sentence-level assistance, but many students commented that the output sounded robotic, and that learners need to be encouraged to add their own voice and emotion to AI-generated drafts (Wang et al., 2024). In addition, the cross-national qualitative findings provide additional support that learners perceive AI as having a positive impact on the development of grammar, vocabulary, paraphrasing, and idea generation; however, learners also express concerns about over-reliance, reduction in creativity, plagiarism, and authenticity in writing instruction, thereby highlighting the duality of the benefits and drawbacks of AI in writing instruction as found by Rafida et al. (2024).

The study conducted in the Middle East and Asian regions found that using AI increased student engagement and autonomy. Mixed-methods evidence from Indonesian English language teaching (ELT) students reported significant pre–post improvements in writing proficiency and

overwhelmingly positive attitudes towards real-time vocabulary and grammar support of AI, indicating strong classroom benefits when AI is combined with teacher scaffolding (Syafei & Nuraeningsih, 2025). The use of AI-assisted writing tools has been associated with improvements in writing outcomes and also enhanced students’ feedback literacy and engagement, as learners combined interactive practice, peer feedback, and community-building strategies to maximize the benefits of AI (Oktarin et al., 2024). However, the engagement levels depended on the teaching approach. For example, if using AI was not explicitly taught, students tended to engage in a passive manner. A small-scale pilot study further found that learners perceived both ChatGPT and teacher written corrective feedback as effective, but tended to prefer teacher-only feedback or a teacher-plus-ChatGPT combination rather than peer-only feedback, suggesting that blended teacher–AI models may better support student revision practices (Zeevy-Solovey, 2024). For example, students using AI-assisted writing tools tended to accept suggestions without critical evaluation.

On the other hand, studies that involved teaching students about feedback literacy showed that students had more meaningful engagement. For example, according to Zhan and Yan (2025), when students engaged in comprehension, evaluation, and application of feedback received from AI-assisted writing tools, their learning outcomes were more pronounced.

The above findings indicate that the mere use of AI cannot lead to improved learning outcomes. It is critical to ensure that pedagogical methods encourage interaction between the learners and AI-assisted writing tools.

Table 2: Affective and Engagement Outcomes

Affective Variable	% of Studies Reporting Improvement	Typical Statistical Reporting
Writing anxiety	~61% (28/46)	$p < .05$
Writing motivation	~65% (30/46)	$p < .05$
Self-efficacy	~54% (25/46)	Mixed ($p < .10$ to $p < .05$)
Engagement depth	~42% (19/46)	Often qualitative

Note. Unless otherwise indicated, the denominator is N = 46 (total number of included empirical studies). Values are presented either as counts of studies (k) or as percentages of the total sample of included studies. Where percentages are reported, they are calculated as $(k / 46) \times 100$. If a subset of studies contributed data to a specific construct (e.g., learner motivation), the reduced denominator is explicitly indicated in parentheses (e.g., $k = 12/46$). "Typical p-value range" summarizes the range of inferential statistics reported across the contributing studies; no new statistical tests were conducted at the review level for these entries. Percentages should not be interpreted as effect magnitude, but as frequency of reporting across studies.

Quantitative findings related to affective variables showed that reductions in writing anxiety and increases in motivation were reported in over half of the reviewed studies. Mixed-methods evidence further indicates that AI-generated feedback can improve learners’ motivation, foreign-language peace of mind, and trait emotional intelligence while concurrently supporting measurable gains in writing proficiency (Mohammed & Khalid, 2025). However, engagement depth was less consistently operationalized and was frequently assessed through qualitative or self-report measures rather than inferential statistics.

AI and Creativity in Writing

The study findings on using AI in teaching creativity and generative thinking were inconclusive. For example, several studies found that using AI improved students’ idea generation, fluency, and lexical diversity in short-term studies. A recent intervention with secondary school students

demonstrated that combining prompt-engineering pedagogy and AI-supported tasks (e.g., collaborative scriptwriting and AI-enhanced storytelling) produced measurable gains in creativity, critical thinking, and learner confidence while reducing writing anxiety, suggesting that explicit prompt design and cooperative AI activities can make creativity benefits more reliable in classroom settings (Kabeer et al., 2025). For example, Gao et al. (2025) found that using AI in brainstorming improved students' fluency and flexibility in early childhood education students' writing tasks.

However, longitudinal and comparative studies have also sparked concerns about creativity homogenization. For instance, Liu et al. (2024) found that over-reliance on ChatGPT can lead to a homogenization of ideas and reduced creativity, particularly when students use the generated content as the sole input rather than as a supplementary tool. Moreover, Wang et al. (2024) observed that although the texts generated by AI were syntactically advanced, they often lacked personal voice, emotional content, and reflection. Thus, it can be inferred that creativity results are closely related to the level of learner agency and the use of AI as a tool for teaching writing.

Table 3: Creativity and Generativity Outcomes

Creativity Dimension	Short-Term Effect	Long-Term Effect	Statistical Pattern
Idea fluency	Positive	Mixed	$p < .05$ (short-term)
Lexical diversity	Positive	Neutral	Small effects
Originality	Mixed	Negative	Non-significant
Voice / authorship	Neutral	Negative	Qualitative decline

Note. The denominator for percentage calculations is $N = 46$, unless a construct was examined by only a subset of studies, in which case the exact denominator is specified (e.g., $k = 8/46$). Entries are reported as counts of studies (k) and corresponding percentages. Where effect-size ranges are provided (e.g., $d \approx 0.20-0.50$), these reflect values reported in original studies or calculated from reported statistics using standard conversions. Because of the variability in how the measures were defined, percentages show the proportion of results that are favorable, unfavorable, or both instead of the aggregated quantification. Research that involved the effects of AI on creativity found statistically significant short-term benefits for idea generation and vocabulary range (usually $p < .05$). On the other hand, studies that examined longer-term or comparative impacts of AI on originality and personal expression produced results that were either non-significant or adverse.

AI Use in Specialized Educational Contexts

AI-assisted writing tools have also demonstrated good potential for use in specialized and inclusive educational settings. For instance, Fung et al. (2026) whose study focused on learners with dyslexia, found that the CHATTING system has a positive impact on the behavioral, emotional, and cognitive engagement of learners with dyslexia. Thus, it can be inferred that AI-assisted writing tools can be used to create a more inclusive teaching environment. For ESP, Jmaiel et al. (2025) found that AI-assisted writing tools have good potential for use in ESP settings, as they can help learners with different first languages to overcome linguistic barriers and difficulties.

Table 4: Feedback Models and Comparative Effectiveness

Feedback Model	% Reporting Superior Outcomes	Typical Statistical Trend
AI-only feedback	~48%	Mixed
Teacher-only feedback	~55%	Stable
Hybrid (AI + teacher)	~76%	$p < .01$ (most studies)

Note. Percentages use N = 46 as the denominator unless otherwise specified. All entries are counts of studies (k) reporting the respective concern or implication. Where percentages are shown, they are derived from $(k / 46) \times 100$. Because many ethical themes were discussed qualitatively, some entries reflect narrative reporting rather than statistically tested outcomes. These frequencies indicate prevalence of discussion within the included empirical literature and should not be interpreted as weighted estimates of real-world prevalence.

Across comparative studies, hybrid feedback models combining AI-generated and teacher-provided feedback most consistently demonstrated superior learning outcomes, with a higher proportion of statistically significant effects compared to single-source feedback models. Recent mixed-method work suggests that integrating generative AI with teacher feedback can meaningfully improve students’ cognitive and affective engagement with English for Academic Purposes (EAP) feedback — though behavioral uptake may lag without structured scaffolding (Bozorgian & Rahimi, 2025; Zhang & Wang, 2025). The most recent meta-analytical synthesis was able to show that there is virtually no difference between learning performance for AI-generated feedback as compared to human feedback (with an overall pooled small effect size and thus, non-significant); therefore, that lends weight to the emergence of hybrid methods which utilize the scalable capabilities of AI and the contextualized nature of the human feedback (Kaliisa et al., 2026).

4. DISCUSSION

This systematic review synthesized the existing empirical literature on the use of AI-assisted writing tools and critically evaluated the methodological quality of the available evidence. Overall, the findings indicate that AI-assisted writing tools can support writing development by improving multiple dimensions of writing performance, including grammatical accuracy, fluency, writing quality, task completion efficiency, and learners' motivation. At the same time, the review identified several recurring challenges, including concerns about creativity, superficial engagement with AI-generated feedback, and ethical issues related to the responsible use of AI in writing instruction. These findings suggest that the educational value of AI-assisted writing tools depends not only on their technological capabilities, but also on how they are integrated into instructional practices.

Pedagogical Value of AI in Writing Instruction

The findings of this review indicate that AI-assisted writing tools can serve as effective instructional supports, particularly when they provide timely, individualized feedback that complements classroom instruction. Across the reviewed studies, improvements were consistently reported in grammatical accuracy, writing fluency, organization, and overall writing quality, suggesting that AI can facilitate writing development across diverse educational contexts. These benefits may be attributed to the capacity of AI in reducing learners' cognitive demands during lower-level writing processes, thereby allowing greater attention to higher-order processes such as planning, revision, and idea development. However, the evidence also indicates that these positive

outcomes are contingent on pedagogically informed implementation. Rather than serving as substitutes for teacher guidance, AI-assisted writing tools appear to be most effective when integrated into instructional practices that foster critical engagement, learner autonomy, and reflective use of AI-generated feedback.

However, it also suggests that these benefits are not automatic and that learners using AI feedback in isolation are likely to engage in surface rather than deep revision processes. This mirrors some of the concerns highlighted in feedback research that emphasize the contingencies surrounding feedback's effectiveness in relation to learners' capacity for interpretation and meaningful engagement. Therefore, the research provides strong support for the notion of hybrid feedback approaches that incorporate AI feedback and teacher feedback along with guided discussion.

Learner Engagement and Feedback Literacy

Learner engagement features as an important moderating variable in relation to the overall effectiveness of AI-assisted writing instruction. Despite learners' positive attitudes regarding the utilization of AI resources, their involvement is mostly mechanical in nature and limited to scenarios where the learning process involves direct instructions to interact with the feedback provided. In other words, learners rarely consider whether the feedback generated by AI is relevant and appropriate for their communication purposes.

More positive results, such as increased engagement with revisions and long-term improvement in writing quality, were reported in studies that included feedback literacy training. Qualitative case-study evidence suggests that ChatGPT, when used alongside tools such as Grammarly and QuillBot, can complement several dimensions of learners' feedback literacy, although its influence on feedback processing depends on students' feedback-seeking behavior — the authors, therefore, propose an Automated Writing Evaluation (AWE) integration framework for educators (Gozali et al., 2024). These results emphasize how crucial it is to instruct students in maintaining authorial control, negotiating meaning, and critically assessing AI feedback. AI-assisted writing tools run the risk of encouraging passive learning behaviors rather than encouraging the development of autonomous writing skills in the absence of such pedagogical scaffolding.

AI, Creativity, and Learner Agency

The relationship between the use of AI and creative writing is one of the most contentious conclusions of the review. Longer-term data points to a risk of creativity homogenization, despite the fact that multiple studies documented short-term improvements in idea generation and fluency, especially during brainstorming stages. Learners who heavily relied on AI-generated content produced texts that were well-written, but lacked a unique voice and were conceptually predictable.

The results suggest that the influence of AI on creativity is dependent on both how much power the learner has in determining what to create (learner agency) and the way that AI is incorporated into the educational environment (instructional framing). When learners were able to use AI-assisted writing tools in addition to, rather than in place of, their own idea generation, it allowed them to create and maintain control over the development of their own content. The findings of this study suggest that when students use AI-generated ideas instead of their own, the novelty of the piece will be diminished, and thus, AI-generated ideas should be being integrated into creative education as tools to foster creativeness among students versus underlying the fact

that students are the primary creators of the content they are developing when they substitute their own, student-driven ideas for AI-generated ideas.

AI in Specialized and Inclusive Contexts

AI-assisted writing tools also show considerable potential in specialized educational settings. School-developed AI-assisted writing tools demonstrate that educational institutions can design context-specific tools that provide personalized feedback, enhance student motivation, and reduce teachers' workload, thereby extending AI applications beyond commercially available platforms (Zhou et al., 2024). For students with dyslexia (a learning disability resulting in difficulty reading and writing) and for students participating in ESP programs, there is a need for additional support to successfully engage in writing. These findings further suggest that AI-assisted writing tools can contribute to more inclusive writing instruction by providing personalized support that helps reduce barriers for learners with diverse educational needs (Fung et al., 2026; Zhou et al., 2024).

Learners can use AI-assisted writing tools to automate lower-level writing processes and use the time that would have been spent automating a lower-level activity on higher-level thinking skills such as organizing and coherently presenting their ideas. Experimental work with non-native medical students found measurable gains in structure, logic, and language after ChatGPT-assisted revision, with manual and AI scoring producing comparable improvements and most participants reporting enhanced writing quality (Li et al., 2024).

The variation in outcome across the sample population suggested that the effectiveness of AI was contingent upon the user context (proficiency level, curriculum goals, sociocultural backgrounds). Consequently, these results indicate that the user context will play a role in the successful use of AI within writing instruction (there cannot be a one-size-fits-all approach to integrating AI into writing instruction).

Ethical and Assessment-Related Implications

There are ongoing ethical issues associated with the use of AI to write in educational contexts, which have not been remedied. Even though detection technology has developed quickly, the studies indicate that there are serious problems of fairness and bias related to using the technology, especially with regard to L2 writers. Through product-based detection methods, using legitimate strategies to develop L2 language may be penalized and complex processes involved with writing L2 may not be taken into consideration.

Some authors have advocated for more process-based assessment through processes that are more aligned with formative assessment, such as establishing transparency, incorporating revision chronology and student reflection. Recent review evidence suggests that pairing AI-based feedback systems with adaptive rubrics can enhance the reliability and formative value of writing assessment, offering a potentially transformative approach to constructive, process-oriented evaluation (Alawad, 2025). In line with calls for process-oriented assessment, some scholars have proposed shifting part of assessment towards reflective and process-focused tasks (e.g., a reflect–rewrite model) so that students' writing journeys, not only final products, are foregrounded in academic literacy courses (Schutte & van Zyl, 2023). A shift toward a more process-based assessment model would allow for less reliance on the use of punitive and productive detection measures. According to the findings of this review, the ethical implementation of AI necessitates technological solutions as well as institution-based approaches that foster academic integrity, equality, and digital literacy. The qualitative interviews conducted among teachers from the Arab region further confirm these findings by noting the ability of AI to provide customized learning

support for L2 language acquisition and improve administration, although this needs to be balanced with adequate training for educators, equitable integration, and protection against academic dishonesty (Aljabr et al., 2025).

Summary of Key Insights

In general, outcomes from this systematic review indicate that AI-assisted writing tools are not, by their nature, either good or bad; instead, their pedagogical benefits will be qualitatively different depending on how they are implemented and whether learners are effectively supported, as well as the degree to which ethical issues are addressed in their use. When used in the context of a well-conceived instructional design that has a strong foundation in pedagogy, AI-assisted writing tools can positively impact the writing development process and increase learner engagement – when used in uncritical or unsupervised ways, however, it will likely result in less in-depth learning, reduced levels of creativity, and increased ethical dilemmas.

5. CONCLUSION

This systematic review synthesized empirical research from 2019 to 2025 on the use of AI in writing education. It specifically focused on learning outcomes, student and teacher perceptions, creativity, and ethical issues to understand AI-assisted writing tools' role in current instruction and their overall implications for the field. The findings of the review indicate that AI-assisted writing tools have become a significant development in modern writing education, providing clear benefits to learners while also introducing new ethical and pedagogical concerns.

Overall, the evidence synthesized in this review indicates that AI-assisted writing tools can enhance multiple aspects of writing performance, including grammatical accuracy, organization, fluency, and learner motivation. At the same time, the reviewed studies highlight important challenges related to creativity, learner engagement, and the ethical use of AI, underscoring the need for pedagogically informed and responsible integration of these technologies into writing instruction.

Pedagogical Implications

In terms of pedagogical implications, the studies reviewed in this section have shown that the use of AI-assisted writing tools in the field of writing should be integrated in such a way that it supports the work of the teacher, rather than replacing the work of the teacher. In this respect, the hybrid models of feedback have been noted to have the most positive effect, allowing learners to benefit from the use of AI-generated feedback while still allowing the teacher to provide the critical guidance needed by the learners.

In addition, it has been noted that the use of AI-assisted writing tools in the field of writing has implications for the role of the teacher, such as the need to provide learners with guidelines on the use of AI-assisted writing tools, particularly in terms of the importance of learners being aware of the use of such tools in an ethical manner. Survey comparisons of students and teachers indicate only modest disagreement about appropriate GenAI uses but a shared perception of institutional unpreparedness, underscoring the urgent need for explicit classroom- and institution-level guidelines and targeted teacher professional development (Barrett & Pack, 2023). Survey evidence from EFL teachers echoes these mixed outcomes: while teachers note ChatGPT's potential for personalization and collaborative tasks, they also report concerns about hallucinations, curricular mismatch, student dependence, and consequently call for clear integrity standards and targeted professional development (Almahboob, 2025).

Implications for Assessment and Policy

The growing involvement of AI technology in the teaching of writing holds significant implications for the assessment process and the policies adopted by the academic institution. The overdependence on the product-oriented AI assessment tools could create a disadvantage for non-native writers and create a lack of understanding of the complex process involved in the acquisition of L2 writing skills. The results obtained from the review align with the views of the proponents of the process-oriented assessment tools, such as the evaluation of the drafts, revision history, and commentaries.

Limitations of the Review

Some drawbacks are found in this literature review as well. First of all, the search was conducted only among databases of Scopus and Web of Science, and only those sources written in English were considered. Thus, some other studies, particularly empirical ones, written in other languages or presented at conferences, have been missed due to the language of search strategy and database restrictions. Consequently, relevant studies not indexed in Scopus or Web of Science, or those published in languages other than English, have been omitted from this review. Second, the included studies showed significant heterogeneity in terms of design, measurement of outcomes, and quality, which could limit comparisons and further meta-analysis. Furthermore, numeric data on outcomes were not always provided by the authors and had to be recalculated from some test statistic provided, which adds extra uncertainty. Finally, because of the dynamically developing sphere of research, results obtained in studies before May 15, 2025 could be reconsidered based on more recent empirical data.

Directions for Future Research

Future research should focus on longitudinal studies, especially process-oriented ones aimed at studying how students interact with AI-assisted writing tools and what effects such interactions might produce on writing, creativity, and learner autonomy. Another promising perspective for future research could be comparative analysis to determine the effectiveness of various AI-assisted writing tools. Another imperative for future studies is to investigate issues of equity, bias, and accessibility of AI-based writing instruction, especially in underrepresented educational contexts. It is imperative that future studies integrate AI studies with theories of second language acquisition, writing, and educational psychology. By so doing, future studies could help ensure that AI is used effectively and responsibly.

References

- Alawad, E. A. (2025). Transformative assessment in academic writing: AI-based feedback systems with adaptive rubrics. *The Journal of Teaching English for Specific and Academic Purposes*, 13(1), 141–149. <https://doi.org/10.22190/JTESAP250113012A>
- Alghannam, M. S. M. (2024). Artificial intelligence as a provider of feedback on EFL student compositions. *World Journal of English Language*, 15(2), 161. <https://doi.org/10.5430/wjel.v15n2p161>
- Aljabr, F., Zakarneh, B., Annamalai, N., & Al Said, N. (2025). Integrating AI: Challenges and opportunities in teaching English writing skills. *World Journal of English Language*, 15(5), 371. <https://doi.org/10.5430/wjel.v15n5p371>

- Almahboob, T. (2025). Exploring the impact of AI technologies on EFL writing proficiency. *Forum for Linguistic Studies*, 7(7), 270–282. <https://doi.org/10.30564/fls.v7i7.9952>
- Al-Raimi, M., Mudhsh, B. A., Al-Yafaei, Y., & Al-Maashani, S. (2024). Utilizing artificial intelligence tools for improving writing skills: Exploring Omani EFL learners' perspectives. *Forum for Linguistic Studies*, 6(2), 1177. <https://doi.org/10.59400/fls.v6i2.1177>
- Apriani, E., Daulay, S. H., Aprilia, F., Marzuki, A. G., Warsah, I., & Supardan, D. (2025). A mixed-method study on the effectiveness of using ChatGPT in academic writing and students' perceived experiences. *Journal of Language and Education*, 11(1), 26–45. <https://doi.org/10.17323/jle.2025.17913>
- Ardiana, A., Rum, E. P., Syatriana, E., & Khodjiev, Z. (2025). From feedback to artificial intelligence: A bibliometric mapping analysis of the thematic evolution of EFL writing assessment research trends (2014–2024). *International Journal of Learning, Teaching and Educational Research*, 24(9), 893–921. <https://doi.org/10.26803/ijlter.24.9.43>
- Barrett, A., & Pack, A. (2023). Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, 20(1), 59. <https://doi.org/10.1186/s41239-023-00427-0>
- Bozorgian, H., & Rahimi, H. (2025). Peer e-feedback and ChatGPT-4o in EFL writing: A cognitive-interpersonal comparison based on EFL students. *Journal of Language and Education*, 11(4), 54–65. <https://doi.org/10.17323/jle.2025.27195>
- Bristi, N. L., Shafie, L. A., Hajimia, H. B., & Hasan, M. K. (2025). Decoding EFL learners' intention to use ChatGPT for academic writing: Cognitive and emotional drivers. *International Journal of Learning, Teaching and Educational Research*, 24(3), 1–28. <https://doi.org/10.26803/ijlter.24.3.1>
- Campino, J. (2024). Unleashing the transformers: NLP models detect AI writing in education. *Journal of Computers in Education*, 12(2), 645–673. <https://doi.org/10.1007/s10639-024-12345>
- Fung, K. Y., Fung, K. C., Lee, L. H., Leung Lui, R. T., Qu, H., Song, S., & Sin, K. F. (2026). A study on using ChatGPT to help students with dyslexia learn Chinese and English writing. *ECNU Review of Education*, 9(1), <https://doi.org/10.1177/20965311251358269>
- Gao, J., Zhang, J., & Li, Y. (2025). Do AI chatbot-integrated writing tasks influence writing self-efficacy and critical thinking ability? A quasi-experimental study. *Computers and Education: Artificial Intelligence*, 9, 100472. <https://doi.org/10.1016/j.caeai.2025.100472>
- Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, 3, 100055. <https://doi.org/10.1016/j.caeai.2022.100055>
- Gozali, I., Wijaya, A. R. T., Lie, A., Cahyono, B. Y., & Suryati, N. (2024). ChatGPT as an Automated Writing Evaluation (AWE) tool: Feedback literacy development and AWE tools' integration framework. *JALT CALL Journal*, 20(1), n1. <https://doi.org/10.29140/jaltcall.v20n1.1200>
- Graham, S. (2019). Changing how writing is taught. *Review of Research in Education*, 43(1), 277–303. <https://doi.org/10.3102/0091732X18821125>
- Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press.
- Jaramillo, J. J., Chiappe, A., & Delgado, F. S. (2025). From struggle to mastery: AI-powered writing skills in ESL education. *Applied Sciences*, 15(14), 8079. <https://doi.org/10.3390/app15148079>

- Jmaiel, H. A., Abukhait, R. O., Mohamed, A. M., Shaaban, T. S., Al-khresheh, M. H., & AL-Qadri, A. H. (2025). The role of ChatGPT in enhancing EFL students' ESP writing skills: An experimental study of gender and major differences. *Discover Education*, 4(1), 240. <https://doi.org/10.1007/s44217-025-00700-6>
- Kabeer, A., Bhat, R. A., Antony, S., & Trambo, I. A. (2025). Enhancing creative writing skills in secondary school students through prompt engineering and artificial intelligence. *Forum for Linguistic Studies*, 7(3), 800–815. <https://doi.org/10.30564/fls.v7i3.8511>
- Kaliisa, R., Misiejuk, K., López-Pernas, S., & Saqr, M. (2026). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. *Educational Psychology*, 46(1), 80–111. <https://doi.org/10.1080/01443410.2025.2553639>
- Khany, R., & Barzan, P. (2025). Leveraging AI-powered writing assistants to enhance L2 writing proficiency: A mixed-methods study. *AI and Tech in Behavioral and Social Sciences*, 3(4), 1–14. <https://doi.org/10.61838/kman.aitech.4287>
- Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education: Research*, 23, 24. <https://doi.org/10.28945/5370>
- Kwon, J., & Lee, Y. (2025). Tracking editorial footprints: A process-oriented analysis of ChatGPT reliance in student writing. *Forum for Linguistic Studies*, 7(6), 941–954. <https://doi.org/10.30564/fls.v7i6.9985>
- Lendvai, G. F. (2025). ChatGPT in academic writing: A scientometric analysis of literature published between 2022 and 2023. *Journal of Empirical Research on Human Research Ethics*, 20(3), 131–148. <https://doi.org/10.1177/15562646251350203>
- Li, J., Zong, H., Wu, E., Wu, R., Peng, Z., Zhao, J., ... & Shen, B. (2024). Exploring the potential of artificial intelligence to enhance the writing of English academic papers by non-native English-speaking medical students: The educational application of ChatGPT. *BMC Medical Education*, 24(1), 736. <https://doi.org/10.1186/s12909-024-05738-y>
- Liu, Q., Zhou, Y., Huang, J., & Li, G. (2024). When ChatGPT is gone: Creativity reverts and homogeneity persists. *arXiv preprint arXiv:2401.06816*. <https://doi.org/10.48550/arXiv.2401.06816>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(1), 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Mohammed, S. J., & Khalid, M. W. (2025). Under the world of AI-generated feedback on writing: Mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia*, 15(1), 7. <https://doi.org/10.1186/s40468-025-00343-2>
- Nelson, A. S., Santamaría, P. V., Javens, J. S., & Ricaurte, M. (2025). Students' perceptions of generative artificial intelligence (GenAI) use in academic writing in English as a foreign language. *Education Sciences*, 15(5), 611. <https://doi.org/10.3390/educsci15050611>
- Oktarin, I. B., Saputri, M. E. E., Magdalena, B., Hastomo, T., & Maximilian, A. (2024). Leveraging ChatGPT to enhance students' writing skills, engagement, and feedback literacy. *Edelweiss Applied Science and Technology*, 8(4), 2306–2319. <https://doi.org/10.55214/25768484.v8i4.1600>
- Pérez, A., McClain, S. K., Roa, A. F., Rosado-Mendinueta, N., Trigos-Carrillo, L., Robles, H., & Campo, O. (2025). Artificial intelligence applications in college academic writing and

- composition: A systematic review. *Íkala, Revista de Lenguaje y Cultura*, 30(1), 12. <https://doi.org/10.17533/udea.ikala.355878>
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *A product from the ESRC methods programme Version, 1*(1), b92. <https://doi.org/10.13140/2.1.1018.4643>
- Punar Özçelik, N., & Yangın Ekşi, G. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, 11(1), 10. <https://doi.org/10.1186/s40561-024-00296-8>
- Rafida, T., Suwandi, S., & Ananda, R. (2024). EFL students' perception in Indonesia and Taiwan on using artificial intelligence to enhance writing skills. *Jurnal Ilmiah Peuradeun*, 12(3), 987–1016. <https://doi.org/10.26811/peuradeun.v12i3.1520>
- Schutte, N., & van Zyl, Z. (2023). Course correction: The role of reflective writing assessment in a post-pandemic academic literacy course. *Perspectives in Education*, 41(3), 77–94. <https://doi.org/10.38140/pie.v41i3.7291>
- Syafei, M., & Nuraeningsih, N. (2025). Leveraging artificial intelligence in writing: ELT students' perspectives and experiences. *International Journal of Learning, Teaching and Educational Research*, 24(5), 416–430. <https://doi.org/10.26803/ijlter.24.5.22>
- Talapngoen, S., & Thuratham, W. (2025). The effect of QuillBot utilization on the development of university students' writing skills. *Studies in Self-Access Learning Journal*, 16(2). <https://doi.org/10.37237/160206>
- Titko, J., Steinbergs, K., Achieng, M., & Uzule, K. (2023). Artificial intelligence for education and research: Pilot study on perception of academic staff. *Virtual Economics*, 6(3), 7–19. [https://doi.org/10.34021/ve.2023.06.03\(1\)](https://doi.org/10.34021/ve.2023.06.03(1))
- Wang, D. (2024). Teacher-versus AI-generated (Poe application) corrective feedback and language learners' writing anxiety, complexity, fluency, and accuracy. *International Review of Research in Open and Distributed Learning*, 25(3), 37–56. <https://doi.org/10.19173/irrodl.v25i3.7646>
- Wang, C., Aguilar, S. J., Bankard, J. S., Bui, E., & Nye, B. (2024). Writing with AI: What college students learned from utilizing ChatGPT for a writing assignment. *Education Sciences*, 14(9), 976. <https://doi.org/10.3390/educsci14090976>
- Zakarneh, B. I., Aljabr, F., Al Said, N., & Jlassi, M. (2025). Assessing pedagogical strategies integrating ChatGPT in English language teaching: A structural equation modelling-based study. *World Journal of English Language*, 15(3). <https://doi.org/10.5430/wjel.v15n3p364>
- 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. <https://doi.org/10.29140/tlrl.v6n3.1482>
- Zhan, Y., & Yan, Z. (2025). Students' engagement with ChatGPT feedback: Implications for student feedback literacy in the context of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 1-14. <https://doi.org/10.1080/02602938.2025.2471821>
- Zhang, J., & Wang, J. (2025). Student perceptions of hybrid feedback: Using Gen-AI to enhance engagement with EAP writing feedback. *The JALT CALL Journal*, 21(2), 2175–2175. <https://doi.org/10.29140/jaltcall.v21n2.2175>
- Zhou, Y., Qi, J., Zeng, Q., & Liu, N. (2024). Research on the application of WAgent in English writing teaching based on AI models. In *Proceedings of the 2024 9th International Conference on Distance Education and Learning* (pp. 10–15). <https://doi.org/10.1145/3675812.3675821>