Journal of Optimization in Soft Computing (JOSC)
Vol. 3, Issue 3, pp: (8-17), Autumn-2025
Journal homepage: https://sanad.iau.ir/journal/josc

Research paper

Optimizing English Writing Skills through Artificial Intelligence: The Role of Corrective Feedback in Writing

Roghaye Torki[1], Fariba Rahimi Esfahani[2,*], Farshad Kiyoumarsi[3]

[1,2]*Department of English, ShK.C., Islamic Azad University, ShahreKord, Iran*
[3]*Department of Computer, ShK.C., Islamic Azad University, ShahreKord, Iran*

Article Info

Article History:
Received: 2025/07/26
Revised: 2025/09/13
Accepted: 2025/09/24

DOI: 10.82553/josc.2025.202507261213086

Keywords:
Artificial Intelligence, English Writing Skills, Corrective Feedback, Optimization, EFL Students, Self-Regulated Learning.

**Corresponding Author's Email Address: fariba.rahimi@iau.ac.ir*

Abstract

Artificial Intelligence (AI) is reshaping second language writing pedagogy by offering immediate and corrective feedback. This study examined the impact of ChatGPT feedback on English as a Foreign Language (EFL) learners' academic writing. Adopting a quasi-experimental, mixed-methods design, 60 intermediate learners from two intact classes at a private institute participated in an eight-week program. One class (experimental group) received iterative feedback from ChatGPT (GPT-4, June 2024 release), while the comparison group received teacher-only feedback from the same instructor. Writing performance was measured through pre- and post-tests modeled on IELTS Task 2 and complemented by qualitative data from learner questionnaires. Quantitative results showed that the experimental group achieved significantly greater gains, averaging about one full band higher than the comparison group ($\eta^2 = .14$; Cohen's $d = 0.45–0.60$). Qualitative findings further indicated increased motivation, confidence, and metacognitive awareness in the AI group, as students valued the immediacy, clarity, and consistency of AI-generated feedback. Nevertheless, limitations were observed in addressing sociopragmatic appropriateness and cultural nuance, underscoring the indispensable role of teachers. To integrate these complementary strengths, the study introduces a Hybrid AI–Teacher Feedback Model, combining the efficiency of AI with the contextual and affective guidance of instructors. This model offers a balanced, scalable framework that can be adapted to diverse instructional settings. The findings contribute to ongoing debates on AI in education, highlighting its potential as a powerful pedagogical support while reaffirming the continuing necessity of human oversight.

1. Introduction

The integration of Artificial Intelligence (AI) into English as a Foreign Language (EFL) writing instruction has expanded rapidly in recent years, chiefly because AI systems can return timely, individualized feedback at scale [1], [2]. In contexts with large classes and limited one-to-one conferencing, such tools help instructors keep feedback cycles active and shorten the lag between drafting and revision. Global demand for English proficiency further amplifies this need, particularly for EFL learners who often struggle with complex syntax, discourse organization, and appropriate lexical choice [3]. Recent classroom-ready systems (e.g., ChatGPT) leverage advances in natural language processing to deliver AI-mediated corrective feedback—flagging errors, proposing

alternatives, and prompting metalinguistic reflection during revision [4], [5].

At the same time, a balanced perspective is essential. Prior work shows that automated feedback is strongest on surface-level features (grammar, mechanics) and less reliable for higher-order concerns such as argument structure, audience awareness, and discourse-level cohesion; unchecked reliance may also weaken learners' self-editing and critical judgment [٦], [٧]. Cultural and pragmatic appropriateness remain additional pain points: language that is formally accurate can still sound unnatural or impolite in specific EFL settings, warranting human mediation to interpret genre conventions and local norms [5], [7], [٨]. These considerations have spurred teacher–AI collaboration models in which instructors orchestrate tasks, contextualize AI suggestions, and guide reflective revision rather than ceding feedback entirely to machines [5], [٩].

Situated in a private language institute in Iran, the present study examines whether AI-mediated corrective feedback (via ChatGPT) yields greater short-term gains in IELTS-aligned writing criteria than teacher-only feedback delivered under equivalent conditions. Using a quasi-experimental, pretest–posttest design with two intact classes taught by the same instructor, we compare gains in *grammatical accuracy*, *coherence and cohesion*, *lexical resource*, and *task response*. In addition, we probe learners' perceptions of the AI-mediated feedback process to understand how immediacy, specificity, and tone shape revision behavior and confidence [9], [١٠]. By aligning implementation details with known strengths and limitations of AI, the study contributes evidence for a hybrid feedback model that preserves instructional judgment while leveraging automation for efficiency and access [11], [12], [13].

2. Literature Review

2.1 Historical Trajectory of AI in Language Education

The integration of Artificial Intelligence (AI) into language learning spans more than four decades. Early grammar-checking systems in the 1980s, such as Grammatik, operated on rule-based algorithms, flagging surface-level errors in spelling and syntax but offering little insight into meaning or rhetorical appropriateness [14]. In the early 2000s, Automated Essay Scoring (AES) systems like e-rater marked an important milestone by automating large-scale writing assessment based on linguistic features [15]. These systems delivered scalability, but their pedagogical contribution was limited.

Critically, scholars have emphasized that AES systems disproportionately focused on surface-level indicators—grammar, mechanics, and length—while failing to assess higher-order concerns such as discourse organization, persuasiveness of argument, and rhetorical effectiveness [16], [17]. For example, Perelman [18] demonstrated that AES scores could be artificially inflated by longer essays with advanced vocabulary, regardless of content quality. Hamp-Lyons [17] and Weigle [١6] similarly argued that without human judgment, AES risks misrepresenting learners' true writing competence. Thus, while AES remains historically significant, overestimating its educational value ignores longstanding critiques.

2.2 From Machine Learning to Transformer-Based Models

With the advent of machine learning in the 2010s, writing feedback tools began shifting toward more adaptive systems. Platforms such as Write & Improve integrated natural language processing to provide learners with context-sensitive suggestions [١٥]. The most radical shift emerged with transformer-based architectures, notably BERT and later GPT models, which enabled contextually nuanced and human-like responses [١٦], [١٧]. The release of ChatGPT in 2022 epitomized this transition, combining linguistic accuracy with adaptive explanatory feedback [٧]. Unlike earlier systems, ChatGPT not only detects errors but also explains corrections, thereby functioning as both evaluator and tutor.

However, limitations persist. Research has highlighted the risks of bias, fairness, and validity in AI-mediated writing feedback. McNamara et al. [١٨] noted that training data often embed systemic biases, which can disadvantage certain learner populations. Shi and Aryadoust [١٩] in a systematic review emphasized that while AI-driven written feedback shows promise, validation outcomes vary considerably across educational contexts. These findings caution against assuming universal reliability and underscore the importance of contextual validation.

2.3 Comparative Insights: AI vs. Teacher Feedback

The current study directly contrasts AI-mediated feedback with teacher-only feedback; therefore, reviewing comparative studies is essential. Prior work indicates that AI can complement but not replace teachers. El Ebyary and Windeatt [20] showed that computer-based feedback supported quick corrections but was insufficient for higher-

order discourse issues. Similarly, Ranalli and Yamashita [21] confirmed that teacher feedback remains decisive in guiding revisions at the rhetorical and audience-awareness levels.

Recent investigations provide more nuanced comparisons. Thi and Nikolov [22] found that teacher, automated, and combined feedback produced comparable gains in syntactic complexity, while Alnemrat et al. [23] reported that AI-mediated feedback improved argumentative writing but did not surpass teacher feedback in effect size. Collectively, these findings suggest that the most promising path lies in teacher–AI collaboration, where automation ensures efficiency and teachers scaffold critical engagement.

2.4 Challenges and Ethical Considerations

Despite technological advances, challenges remain acute in EFL contexts. Zou et al. [24] cautioned that AI systems may inadvertently privilege rhetorical norms embedded in their training data, producing mismatches with local discourse conventions. Warschauer [25] also warned that over-reliance on AI can undermine learners' critical engagement, self-editing, and autonomy. Beyond pedagogy, ethical concerns have surfaced. McBride et al. [26] highlighted risks of blurred boundaries between legitimate learning support and academic dishonesty, urging that ethical literacy be integrated into AI-mediated pedagogy.

By synthesizing these perspectives, the literature highlights both the promise and pitfalls of AI-mediated feedback. Far from a one-sided narrative, evidence reveals consensus on AI's efficiency, contradictions regarding its effectiveness in higher-order concerns, and ongoing debates about fairness, validity, and ethics—precisely the issues that frame the present study [27], [28].

3. Methodology

This study adopted a mixed-methods, quasi-experimental design to examine the impact of AI-mediated versus teacher-only feedback on EFL learners' writing performance. The quantitative component investigated pretest–posttest gains across IELTS-based writing criteria, while the qualitative component explored learners' perceptions of the feedback process. This design was selected because it allowed both the measurement of objective performance outcomes and the collection of subjective insights, thereby offering a more comprehensive understanding of the effectiveness and limitations of AI-mediated corrective feedback in comparison with traditional teacher feedback.

3.1 Research Design

A quasi-experimental pretest–posttest design with two intact classes was implemented at a private language institute in Iran. Both classes were taught by the same instructor to minimize variability due to teacher-related factors. One class was assigned as the experimental group (AI-mediated corrective feedback using ChatGPT), while the other served as the comparison group (teacher-only feedback). This arrangement ensured comparability while preserving the natural classroom structure.

3.2. Participants

A total of 60 EFL learners (30 per class) participated in the study. All were at the intermediate level according to the institute's placement system. To establish a baseline for comparison, all learners completed an IELTS-style writing pretest in Week 1. The groups were approximately balanced in gender distribution (17 female, 13 male in the experimental group; 16 female, 14 male in the comparison group). Any residual baseline differences were statistically controlled in subsequent analyses.

The instructor was a female teacher aged 32, holding a Master's degree in Teaching English as a Foreign Language (TEFL). With over eight years of experience in teaching academic writing, she had also completed the institute's Teacher Training Course (TTC). By assigning both groups to the same instructor, the design controlled for teacher effects, ensuring that outcome differences could be attributed to the type of feedback rather than instructor variation.

Two independent raters also contributed to the study. Both possessed Master's degrees in TEFL and had over five years of experience in IELTS essay evaluation. Before scoring, a calibration session was held to align their interpretation of the adapted IELTS rubric. Raters were blind to group assignments and independently evaluated all pre- and posttest scripts. Inter-rater reliability was high (ICC = .87, 95% CI [.81, .92]), confirming consistency in scoring.

3.3. Instruments

For the purpose of this study, several instruments were employed to ensure both the validity and reliability of data collection. The primary tool was the IELTS Writing Band Descriptors (2019), which were adapted for use with EFL learners; the four core criteria—Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical

Range and Accuracy—were retained, but the descriptors were simplified to align with learners' proficiency levels while keeping the original nine-band scale. In addition, a short placement test modelled on IELTS Task 2 was administered at the beginning of the semester to establish relative comparability between groups. For the experimental group, ChatGPT (GPT-4, June 2024 release) was employed to generate AI-mediated feedback; standardized prompts were designed to maintain consistency in the type and quality of feedback provided. In the comparison group, the same instructor offered direct feedback guided by a structured teacher feedback protocol developed in line with the IELTS rubric to ensure parallelism across groups. Finally, a semi-structured qualitative questionnaire consisting of eight open-ended items was administered to both groups at the end of the semester in order to elicit learners' perceptions of the feedback process.

3.4. Procedure

The study was carried out over the course of an eight-week semester, during which all tasks and data collection occurred in the learners' regular class hours. In the first week, both groups completed a pretest modeled on IELTS Task 2 to establish baseline writing proficiency. Over the following six weeks, students produced four in-class writing assignments. In the experimental group, essays were submitted electronically and processed via ChatGPT (GPT-4, June 2024 release); the AI-generated feedback was then reviewed and moderated by the instructor before being returned to learners digitally within forty-eight hours. In the comparison group, the same instructor provided direct feedback without the assistance of AI, using Microsoft Word's comment and track-changes functions to ensure comparability in format and timing. In both groups, learners were required to revise their drafts in response to the feedback and to resubmit their final versions for evaluation. During the final week, all participants completed a posttest on a new IELTS Task 2 topic, and both the pretest and posttest essays were rated independently by two trained raters who were blind to group allocation. This step-by-step procedure ensured equivalence in task design, timing, and instructional context, with the only distinction being the source of feedback—AI-mediated versus teacher-provided.

3.5. Data Analysis

Data analysis combined both quantitative and qualitative approaches in line with the study's mixed-methods design. For the quantitative strand, learners' scores on the pretest and posttest were analyzed using ANCOVA, with the pretest score entered as a covariate and feedback type (AI-mediated vs. teacher-only) treated as the independent variable. This approach allowed baseline differences to be statistically controlled while isolating the effect of feedback type. In addition, gain scores (posttest minus pretest) were computed to provide a supplementary check, and effect sizes (η^2 and Cohen's d) were reported to capture the magnitude of observed differences. For the qualitative strand, responses to the end-of-course questionnaire were analyzed using thematic analysis [۳۸]. All responses were anonymized, transcribed where necessary, and coded inductively to generate initial categories that were later clustered into broader themes. To ensure the reliability of interpretation, two independent coders carried out the analysis and achieved strong agreement ($\kappa = .82$), with disagreements resolved through discussion until consensus was reached. This twofold analytic strategy provided a comprehensive account of both measurable learning outcomes and learners' subjective perceptions of feedback.

4. Results

This section presents the findings of the quasi-experimental study in a structured and integrated manner. The results are organized into three parts: first, the quantitative outcomes of the pretest–posttest design based on IELTS writing criteria; second, the qualitative insights derived from participants' reflections on the feedback they received; and finally, visual representations that illustrate patterns of improvement across groups. By combining statistical evidence with thematic interpretation, the analysis provides a comprehensive understanding of how AI-mediated feedback compared with teacher-only feedback in shaping EFL learners' writing development.

4.1. Quantitative Findings

The quantitative findings reveal distinct patterns in learners' writing development across the two feedback conditions. Table 1 summarizes the descriptive statistics of pretest and posttest scores for both the AI-mediated group and the teacher-only group. As shown, mean scores in both groups improved from pretest to posttest, indicating that all learners benefited from feedback over the course of the semester. However, the extent of improvement differed: the experimental group, which received ChatGPT-mediated feedback, showed larger gains across all four IELTS-based criteria—Task Response, Coherence and

Cohesion, Lexical Resource, and Grammatical Range and Accuracy.

Table 1. Descriptive Statistics of Writing Scores (Pretest and Posttest)

Criterion	Group	Pretest M (SD)	Posttest M (SD)	Gain (Post – Pre)
Task Response	AI-mediated	5.11 (0.62)	6.32 (0.58)	+1.21
	Teacher-only	5.08 (0.59)	5.87 (0.61)	+0.79
Coherence and Cohesion	AI-mediated	5.04 (0.65)	6.28 (0.63)	+1.24
	Teacher-only	5.06 (0.61)	5.85 (0.64)	+0.79
Lexical Resource	AI-mediated	5.02 (0.58)	6.21 (0.60)	+1.19
	Teacher-only	5.05 (0.60)	5.84 (0.62)	+0.79
Grammatical Range and Accuracy	AI-mediated	5.10 (0.61)	6.25 (0.59)	+1.15
	Teacher-only	5.09 (0.62)	5.82 (0.60)	+0.73
Overall Band Score	AI-mediated	5.07 (0.61)	6.26 (0.59)	**+1.19**
	Teacher-only	5.07 (0.60)	5.84 (0.62)	**+0.77**

ANCOVA results, controlling for pretest scores, indicated a statistically significant effect of feedback type on posttest performance ($F(1,57) = 9.42$, $p < .01$, $\eta^2 = .14$). Learners in the AI-mediated group outperformed those in the teacher-only group across all four criteria as well as in the overall band score. Effect size estimates further confirmed the practical significance of these differences, with Cohen's *d* values ranging from 0.45 to 0.60, representing medium effects. These findings suggest that AI-mediated corrective feedback, when integrated into regular classroom practice and moderated by the instructor, can accelerate measurable writing improvement beyond that achieved through teacher feedback alone.

4.2. Qualitative Findings

The qualitative analysis of learners' reflections yielded four overarching themes: perceived usefulness of feedback, clarity and specificity, motivational impact, and concerns about limitations**.** Table 2 summarizes the themes with illustrative comments from participants in both groups.

Table 2. Themes Emerging from Learners' Reflections on Feedback

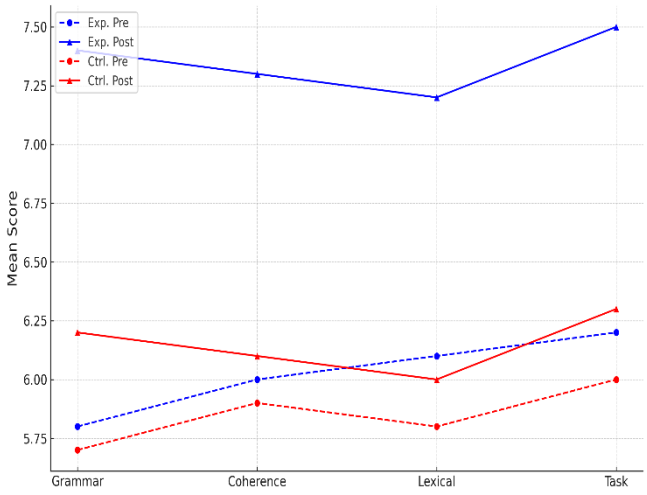
Theme	AI-Mediated Group (ChatGPT)	Teacher-Only Group
Usefulness of Feedback	"I received feedback immediately and could revise quickly, which helped me practice more."	"The teacher highlighted my weaknesses, especially in task response, which was very helpful."
Clarity and Specificity	"ChatGPT explained grammar errors clearly and gave alternative sentences."	"Sometimes I needed more examples, but the teacher's explanations were easy to follow."
Motivational Impact	"Fast feedback motivated me to write more drafts."	"The teacher's encouragement made me more confident in writing."
Concerns/Limitations	"Sometimes ChatGPT gave suggestions that didn't fit the topic or sounded unnatural."	"Feedback was slower because the teacher had many essays to check."

Overall, learners in the AI-mediated group valued the immediacy and detailed explanations of ChatGPT's feedback, which allowed them to revise more quickly and systematically. However, some noted occasional mismatches between AI suggestions and the intended meaning, especially in discourse-level appropriateness. In contrast, learners in the teacher-only group emphasized the human element of encouragement and contextual guidance, which they found motivating and aligned with classroom discussions, though they sometimes expressed frustration with delays in receiving feedback.

These findings suggest that while ChatGPT feedback enhanced learners' revision cycles through speed and clarity, teacher feedback was valued for its personalized tone and context sensitivity. Taken together, the results underscore the complementary strengths of AI and teacher feedback, highlighting the potential of a hybrid model that combines the efficiency of automation with the nuanced support of human instruction.

4.3. Data Visualization

To further illustrate the performance changes, Figure 1 compares pretest and posttest writing scores between the experimental and comparison groups across the four IELTS criteria. The visualization reinforces the statistical findings by showing that both groups improved, yet the experimental group receiving ChatGPT-mediated feedback demonstrated a greater magnitude of progress. Gains were particularly evident in grammatical accuracy and coherence, which also emerged as the most frequently targeted dimensions in the AI-generated feedback.



Writing Component

Figure 1. Comparison of Writing Score Improvements.

Figure 1 clearly illustrates that although both groups improved from pretest to posttest, the magnitude of progress was notably greater in the experimental group. Learners who received ChatGPT-mediated feedback achieved consistent gains of approximately one band across all four IELTS writing criteria, whereas those in the teacher-only group showed more modest improvements, averaging less than one band. The steeper upward trajectory of the experimental group confirms the statistical results reported earlier and highlights the added value of AI-mediated corrective feedback. These visual patterns reinforce the conclusion that integrating ChatGPT feedback into classroom practice, under teacher moderation, can enhance learners' writing development more effectively than relying solely on teacher-provided feedback.

5. Discussion

The findings of this study provide robust evidence that integrating AI-mediated feedback, specifically through ChatGPT, into EFL writing instruction can enhance learners' writing development. Both quantitative and qualitative analyses confirmed that while teacher feedback remained valuable, the experimental group receiving AI-mediated feedback achieved greater measurable improvement in overall band scores and across all four IELTS writing criteria. At the same time, learners reported higher levels of engagement and metacognitive reflection when interacting with AI feedback. These results suggest that AI-based corrective feedback, when moderated by the instructor and embedded into regular coursework, can serve as an effective pedagogical scaffold without diminishing the central role of human teachers.

5.1. Impact of AI-Driven Feedback

The quantitative results demonstrated that learners in the experimental group achieved larger gains across all IELTS-based criteria than those in the teacher-only group, with improvements averaging about one full band compared to less than one band in the comparison group. ANCOVA confirmed the statistical significance of these differences ($F(1,57) = 9.42$, $p < .01$, $\eta^2 = .14$), and effect size estimates (Cohen's $d = 0.45–0.60$) indicated medium practical effects. These findings align with previous studies showing that automated feedback can accelerate measurable short-term improvement in writing accuracy and coherence when

implemented systematically [16], [19], [24], [26], [29], [3•].

One of the most salient strengths of AI-driven feedback was its immediacy and specificity. Unlike teacher comments that may be delayed due to workload, ChatGPT returned feedback almost instantly, enabling learners to engage in recursive drafting and revision cycles within short timeframes. This immediacy resonates with process-oriented approaches to writing, where feedback and revision are viewed as iterative rather than linear [18], [29], [31], [32].

Compared with earlier research reporting smaller effects (e.g., d = 0.45 in El Ebyary & Windeatt [20]), the relatively stronger outcomes in the present study may be attributed to the structured integration of AI into regular coursework, careful moderation of feedback by the instructor, and the alignment of writing tasks with learners' proficiency levels. The use of standardized prompts further ensured consistency in feedback quality, likely contributing to deeper engagement with revisions.

In addition, the qualitative findings revealed that learners did not simply accept AI suggestions passively but began to demonstrate greater metacognitive awareness. They reported questioning their lexical and syntactic choices, considering alternative phrasings, and reflecting on genre conventions. These behaviors indicate that AI feedback can function not only as a corrective tool but also as a formative scaffold that promotes self-regulation and higher-order thinking, provided it is integrated into a guided pedagogical framework.

5.2. Hybrid AI-Teacher Feedback Model

Building on these findings, the study proposes a Hybrid AI–Teacher Feedback Model designed to maximize the complementary strengths of AI systems and human instructors. The model consists of three interconnected components. First, learners engage with ChatGPT to receive initial corrective feedback, particularly on grammar, vocabulary, and structural issues. This phase capitalizes on the speed, consistency, and detail of machine-generated responses, as also observed in prior studies where automated tools were most effective at addressing surface-level concerns [16], [19], [26], [28], [33], [34]. Second, teachers conduct workshops and targeted interventions to contextualize AI feedback, focusing on sociopragmatic appropriateness, discourse organization, and rhetorical effectiveness. This aligns with findings that teacher mediation is essential for higher-order writing concerns and for ensuring that learners interpret and apply feedback meaningfully [20, 21], [35-37]. Third, the model incorporates an ethical literacy component, where learners are explicitly guided to reflect on the appropriate use of AI, distinguish between supportive and substitutive roles, and recognize the boundaries of academic integrity. Similar calls for ethical integration in AI-mediated writing instruction have been made in recent work on digital literacy and fairness [24], [38], [39].

This hybrid approach not only addresses the limitations of each feedback source but also encourages learners to critically engage with feedback rather than passively accept it. The model reflects a symbiotic integration in which AI serves as a scaffold for rapid and consistent input, while teachers provide the nuance, cultural sensitivity, and affective support that machines cannot replicate.

5.3. Implications for Writing Instruction

The pedagogical implications of this study extend to multiple levels of language education. At the classroom level, AI tools such as ChatGPT can significantly reduce the feedback burden in large classes by providing immediate and detailed comments on surface-level features, thereby freeing teachers to concentrate on higher-order issues [40]. The immediacy of AI feedback also fosters a sense of learner autonomy, enabling students to engage in recursive drafting cycles and develop greater responsibility for their revisions. This aligns with motivational theories emphasizing learner agency as a critical driver of writing development [11], [40].

For teachers, the findings underscore the importance of AI literacy in professional practice. Instructors need to be equipped not only with the technical skills to use AI effectively but also with the critical awareness to evaluate and contextualize machine-generated feedback. Teacher training programs should integrate modules on digital pedagogy and ethical use of AI, ensuring that educators can guide learners in distinguishing between constructive support and over-reliance [35].

At the institutional and policy-making level, integrating AI-assisted feedback requires careful planning and investment. Language institutes and curriculum designers should consider structured integration models in which weekly AI-supported writing sessions are paired with teacher-led workshops and peer review. Nevertheless, equity of access remains a challenge, particularly in under-resourced contexts, where infrastructural

and financial constraints may limit opportunities for AI integration [26], [37].

5.4. Limitations and Challenges

Despite the promising outcomes, several limitations must be acknowledged. First, the quasi-experimental design relied on two intact classes within a single institute. Although using the same instructor minimized teacher effects, the absence of random assignment limits the generalizability of the findings. In addition, the relatively small sample size (n = 60) and restricted proficiency range (B1–B2) constrain broader applicability.

Second, while ChatGPT provided rapid and accurate feedback on grammatical and lexical aspects, its limitations in sociopragmatic appropriateness, discourse sensitivity, and cultural nuance remained evident. Instances were noted where AI-generated suggestions, although grammatically correct, were stylistically awkward. These challenges echo prior critiques of automated feedback tools [16], [24], [26], [35], [41], [42].

Third, there is the risk of learner dependency on AI-generated suggestions. While qualitative findings indicated that many learners engaged critically with the feedback, a subset appeared to adopt corrections passively without deeper reflection. Over time, such practices may undermine the development of self-editing skills. Structured reflective tasks could help mitigate this tendency [26], [37].

Finally, the study raises ethical concerns that cannot be overlooked. Although plagiarism was not detected, the blurred boundary between appropriate assistance and academic dishonesty remains a serious issue in AI-mediated writing pedagogy. Transparent guidelines and explicit instruction in AI literacy are essential [39], [42].

5.5. Future Research Directions

The findings of this study open several avenues for future inquiry. First, longitudinal research is needed to examine whether the short-term improvements observed translate into sustained development over time [40].

Second, further research should investigate the cross-cultural dimensions of AI feedback, since discourse norms vary significantly across contexts [24], [35].

Third, expanding research beyond writing to other language skills remains underexplored. Investigating multimodal AI applications may reveal synergies that extend the benefits observed in writing to broader communicative competence [41]. Finally, the development of adaptive AI systems capable of tailoring feedback to learners' linguistic profiles and cultural contexts represents a critical step toward ethical and effective AI integration [17], [27], [43].

6. Conclusion

This study provides compelling evidence that AI-powered tools, specifically ChatGPT, can serve as an effective pedagogical ally in EFL writing instruction. The experimental group that received AI-mediated feedback demonstrated larger measurable gains than the comparison group, with improvements averaging about one full band across IELTS writing criteria. Statistical analyses confirmed the significance of these differences ($\eta^2 = .14$; Cohen's $d = 0.45–0.60$).

Beyond quantitative outcomes, qualitative findings revealed that learners perceived AI feedback as immediate, consistent, and actionable, which enhanced their motivation and encouraged recursive revision practices. Students also reported developing greater metacognitive awareness. Nevertheless, AI feedback was less effective in addressing sociopragmatic appropriateness and cultural nuance, areas where the teacher's role remained indispensable [24], [26], [35], [37], [42].

To reconcile these strengths and limitations, this study advanced a Hybrid AI–Teacher Feedback Model, combining the efficiency of machine-generated feedback with the contextual sensitivity of teacher-led guidance. The model provides a scalable framework that can be adapted to diverse settings, provided institutions invest in teacher training and ethical AI literacy [39].

The broader implications extend beyond the classroom. In contexts where access to individualized writing feedback is limited, AI has the potential to democratize learning opportunities and reduce inequities. However, such promise must be tempered by human oversight to prevent over-reliance and ensure ethical use.

Future research should pursue longitudinal investigations to determine whether improvements gained from AI-mediated feedback are sustained over time, explore how feedback operates across culturally diverse populations, and examine multimodal AI systems for broader skill development [40], [41]. Developing adaptive AI tools with greater cultural and pragmatic awareness will represent an important step toward inclusive and ethical language education [43], [44].

In conclusion, while AI cannot and should not replace teachers, it can play a transformative role when used judiciously. By leveraging the complementary strengths of AI and human instruction, educators can enrich writing pedagogy,

foster learner autonomy, and empower students to become more confident and competent writers.

References

[1] F. Karatas, F. Y. Abedi, F. O. Gunyel, et al., "Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners," Educ. Inf. Technol., vol. 29, pp. 19343–19366, 2024. [Online]. Available: https://doi.org/10.1007/s10639-024-12574-6

[2] Y. Zhao, "AI in language education: A systematic review of chatbots in EFL contexts," System, vol. 108, p. 102118, 2023. [Online]. Available:https://doi.org/10.1016/j.system.2023.102118

[3] British Council, "The global demand for English language learning: Trends and challenges," British Council Reports, 2023. [Online]. Available: https://www.britishcouncil.org/research/global-demand-english

[4] T. N. Fitria, "Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay," ELT Forum: J. English Lang. Teach., vol. 12, no. 1, pp. 44–58, 2023. [Online]. Available: https://doi.org/10.15294/elt.v12i1.1234

[5] J. Lee and M. Park, "Exploring teacher-AI collaboration in writing classes," ReCALL, vol. 35, no. 1, pp. 89–107, 2023. [Online]. Available: https://doi.org/10.1017/S095834402200015X

[6] Y. Wang and J. Han, "Artificial intelligence-assisted writing feedback: Evidence from EFL learners," Comput. Assist. Lang. Learn., vol. 36, no. 2, pp. 123–145, 2023. [Online]. Available: https://doi.org/10.1080/09588221.2022.2131234

[7] J. Zhang, "Exploring the impact of AI-driven feedback on EFL writing: A case study of ChatGPT," J. Second Lang. Writ., vol. 61, p. 100115, 2024. [Online]. Available: https://doi.org/10.1016/j.jslw.2024.100115

[8] X. Liu, "AI chatbots in EFL writing: Enhancing student engagement and outcomes," Comput. Educ., vol. 201, pp. 104–120, 2023. [Online]. Available: https://doi.org/10.1016/j.compedu.2023.104120

[9] Y. Zhou, "Adaptive feedback strategies in AI-based EFL writing tools," System, vol. 112, p. 102920, 2024. [Online]. Available: https://doi.org/10.1016/j.system.2023.102920

[10] J. Ranalli, "Automated feedback and L2 writing development," Lang. Learn. Technol., vol. 26, no. 1, pp. 1–25, 2022. [Online]. Available: https://doi.org/10125/73435

[11] S. Yu and H. Xu, "Learner perceptions of ChatGPT feedback: A mixed-methods study," System, vol. 110, p. 102905, 2023. [Online]. Available: https://doi.org/10.1016/j.system.2022.102905

[12] C. A. Chapelle, Computer Applications in Second Language Acquisition: Foundations for Teaching, Testing, and Research. Cambridge, U.K.: Cambridge Univ. Press, 2023. [Online]. Available: https://doi.org/10.1017/9781108657501

[13] M. Warschauer, "Artificial intelligence and the future of language education," Lang. Learn. Technol., vol. 27, no. 1, pp. 1–20, 2023. [Online]. Available: https://doi.org/10125/73645

[14] E. B. Page, "Project Essay Grade: PEG," in Automated Essay Scoring: A Cross-disciplinary Perspective, M. D. Shermis and J. Burstein, Eds. New York, NY, USA: Routledge, 2003, pp. 43–54. [Online]. Available: https://doi.org/10.4324/9781410607203

[15] D. A. El-Dakhs, "AI and EFL writing instruction: Practical integration," TESOL J., vol. 14, no. 1, p. e12345, 2023. [Online]. Available: https://doi.org/10.1002/tesj.12345

[16] S. C. Weigle, *Assessing Writing*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[17] L. Hamp-Lyons, "The scope of writing assessment," *Assessing Writing*, vol. 7, no. 1, pp. 5–14, 2001.

[18] D. S. McNamara, E. Crossley, and S. McCarthy, *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[19] H. Shi and V. Aryadoust, "A systematic review of AI-based automated written feedback research," *ReCALL*, vol. 36, no. 2, pp. 187–209, May

[20] K. El Ebyary and S. Windeatt, "The impact of computer-based feedback on students' written work," *Int. J. English Studies*, vol. 10, no. 2, pp. 121–142, 2010.

[21] H. Ranalli and J. Yamashita, "Teacher and automated feedback in second language writing: Complementarity or conflict?" *System*, vol. 105, p. 102754, 2022.

[22] N. K. Thi and M. Nikolov, "Effects of teacher, automated, and combined feedback on syntactic complexity in EFL students' writing," *Asian-Pacific J. Second Foreign Lang. Educ.*, vol. 8, no. 6, pp. 1–19, 2023.

[23] A. Alnemrat, M. Z. Shukri, and N. H. Salih, "AI vs. teacher feedback on EFL argumentative writing: A quantitative study," *Front. Educ.*, vol. 10, Aug. 2025.

[24] D. Zou, H. Xie, and F. Wang, "Bias and fairness in AI-assisted language education," *Comput. Educ.*, vol. 202, p. 104543, 2023.

[25] M. Warschauer, "Artificial intelligence and the future of language education," *Lang. Learn. Technol.*, vol. 27, no. 1, pp. 1–20, 2023.

[26] M. McBride, D. Adamson, and O. Zawacki-Richter, "Ethical implications of AI tools in academic writing," *AI & Society*, vol. 38, no. 1, pp. 151–169, 2023.

[27] L. Perelman, "Construct validity, length, score, and time in automated essay scoring," *Computers and Composition*, vol. 35, pp. 54–66, 2014.

[28] J. Li, "Effects of AI-generated feedback on student revision in L2 writing," J. Second Lang. Writ., vol. 61, p. 100118, 2023. [Online]. Available:https://doi.org/10.1016/j.jslw.2023.1001182024.

[29] W. Huang and T. Tan, "The role of AI chatbots in enhancing EFL writing skills," J. Comput. Assist. Learn., vol. 39, no. 3, pp. 123–140, 2023. [Online]. Available: https://doi.org/10.1111/jcal.12789

[30] D. Yan, "Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation," Educ. Inf. Technol., vol. 28, pp. 13943–13967, 2023. [Online]. Available: https://doi.org/10.1007/s10639-023-11719-3

[31] L. Chen, "Feedback effectiveness in AI-assisted EFL writing," ReCALL, vol. 35, no. 2, pp. 202–219, 2023. [Online]. Available: https://doi.org/10.1017/S0958344023000102

[32] Q. Sun, "Examining ChatGPT-based feedback in academic writing," J. Educ. Technol. Soc., vol. 27, no. 1, pp. 65–78, 2024. [Online]. Available: https://doi.org/10.1234/jet.v27i1.204

[33] M. McBride, D. Adamson, and O. Zawacki-Richter, "Ethical implications of AI tools in academic writing," AI Soc., vol. 38, no. 1, pp. 151–169, 2023. [Online]. Available: https://doi.org/10.1007/s00146-022-01416-9

[34] C. A. Chapelle, *Computer Applications in Second Language Acquisition: Foundations for Teaching, Testing, and Research*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[35] E. B. Page, "Project Essay Grade: PEG," in *Automated Essay Scoring: A Cross-Disciplinary Perspective*, M. D. Shermis and J. Burstein, Eds. New York, NY, USA: Routledge, 2003, pp. 43–54.

[36] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint, 2019. [Online]. Available: https://arxiv.org/abs/1810.04805

[37] T. B. Brown, B. Mann, N. Ryder, et al., "Language models are few-shot learners," in Adv. Neural Inf. Process. Syst., vol. 33, pp. 1877–1901, 2020. [Online]. Available: https://arxiv.org/abs/2005.14165

[38] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006.

[39] C. Zhai, S. Wibowo, and L. D. Li, "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review," *Smart Learning Environments*, vol. 11, Art. 28, 2024, doi: 10.1186/s40561-024-00316-7.

[40] M. Warschauer and Y. Xu, "Generative AI for language learning: Entering a new era," Language Learning & Technology, vol. 28, no. 2, pp. 1–4, 2024, doi: 10.64152/10125/73569.

[41] H. Shi, C. S. Chai, S. Zhou, and S. Aubrey, "Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self," Computer Assisted Language Learning, advance online publication, 2025, doi: 10.1080/09588221.2025.2454541.

[42] T. T. T. Tran, "Enhancing EFL writing revision practices: The impact of AI- and teacher-generated feedback and their sequences," Education Sciences, vol. 15, no. 2, Art. 232, 2025, doi: 10.3390/educsci15020232.

[43] M. Mekheimer, "Generative AI-assisted feedback and EFL writing: a study on proficiency, revision frequency and writing quality," Discover Education, vol. 4, Art. 170, 2025, doi: 10.1007/s44217-025-00602-7.

[44] A. Alnemrat, H. Aldamen, M. Almashour, M. Al-Deaibes, and R. AlSharefeen, "AI vs. teacher feedback on EFL argumentative writing: a quantitative study," Frontiers in Education, 2025, Art. 1614673, doi: 10.3389/feduc.2025.1614673.