



Castledown

 OPEN ACCESS

Language Education & Assessment

ISSN 2209-3591

<https://www.castledown.com/journals/lea/>

Language Education & Assessment, 9, 104046 (2026)

<https://doi.org/10.29140/lea.2026.104046>

Generative AI as an IELTS Tutor: The Effects of AI-Generated Feedback on EFL Students' Writing



NAPAT JITPAISARNWATTANA 

Silpakorn University, Thailand

napat.jit@gmail.com

Abstract

This study examines the impact of AI-generated feedback on Thai university students' performance in IELTS writing. Additionally, it explores the students' perceptions of the feedback and their evaluations of its pedagogical value. Adopting a sequential mixed-methods research design, the study integrates quantitative and qualitative evidence collected from 52 participants enrolled in an English Preparation for IELTS at a public university in Thailand. The effectiveness of AI-generated feedback was evaluated through a comparative analysis of students' pre- and post-intervention writing performance, while students' perceptions of the feedback were explored using questionnaire data and semi-structured interviews. The findings indicate that engagement with AI-generated feedback led to improvements in students' post-test writing performance across both writing tasks. The gains were statistically significant in three assessment criteria, with the most notable gains observed in grammatical range and accuracy. Despite marginal improvements, however, the gains on content-related dimensions were limited. The students perceived AI-generated feedback as beneficial and useful for their writing development, while still feeling the need for human feedback. The study concludes that while AI-generated feedback can effectively support the development of specific writing features, meaningful instructional impact continues to depend on active teacher involvement in the feedback cycle.

Keywords: ChatGPT, AI-generated feedback, IELTS writing, academic writing, students' perceptions

Copyright: © 2026 Napat Jitpaisarnwattana. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within this paper.

Introduction

Recent developments in generative artificial intelligence (AI) have generated considerable scholarly interest in its capacity to support and enrich second language (L2) learning and instruction (Barrot, 2023; Hockly, 2023; Montakantiwong & Jitpaisarnwattana, 2026). One area in which generative AI is particularly well positioned to add pedagogical value is second language (L2) writing, where the cognitively demanding and time-intensive nature of feedback provision often results in delayed learner guidance and places substantial strain on teachers (Hwang et al., 2025). To this end, generative AI has emerged as a convenient and high-quality alternative for supporting L2 writing development (Mizumoto et al., 2024). Through interactive and adaptive assistance, such systems can support learners in elaborating ideas, refining lexical and grammatical accuracy, and polishing successive drafts (Barrot, 2023; Su et al., 2023).

However, findings from empirical studies on generative AI-assisted L2 writing suggest a more complex reality, with evidence pointing to both positive learning outcomes and emerging concerns regarding its use (Hwang et al., 2025). With regard to writing effectiveness, empirical studies have reported that generative AI-assisted interventions are associated with measurable improvements in learners' writing performance, particularly in terms of content development, organisational coherence, and grammatical accuracy (Jen & Salam, 2024; Jitpaisarnwattana, 2024; Mahapatra, 2024; Song & Song, 2023). However, studies examining learner perceptions reveal a more ambivalent stance, with L2 learners recognising motivational and instructional benefits alongside concerns regarding academic honesty and the trustworthiness of AI-generated outputs (Allen & Mizumoto, 2024; Yan, 2023).

As seen, existing research has provided valuable insights into how L2 learners, educators, and researchers conceptualise the pedagogical value of generative AI in supporting L2 writing, while also demonstrating its potential to yield meaningful improvements in learners' written production. Moreover, although the existing literature has largely focused on the effects of generative AI on overall writing development, the generative capabilities of these systems have also facilitated research targeting specific writing genres, including those associated with standardised assessments such as IELTS and TOEFL (Jitpaisarnwattana & Saville, 2025). Although a limited number of studies have explored the use of generative AI in supporting IELTS writing, these studies have largely focused on the generation of model responses and test tasks (Andriani, 2025; Sun, 2024). There remains a notable gap in the literature concerning the effects of generative AI use on learners' writing development within standardised test preparation contexts.

The present study therefore seeks to investigate the effects of ChatGPT-generated feedback on students' IELTS writing performance, as evaluated according to the four IELTS assessment criteria: task response (and task achievement), lexical resource, coherence and cohesion, and grammatical range and accuracy. In addition, the study examines students' perceptions of AI-generated feedback. Accordingly, the study is guided by three research questions:

- 1) What effects does ChatGPT-based feedback have on students' performance in IELTS writing as measured by the four official assessment criteria?
- 2) Do the effects of ChatGPT-generated feedback on IELTS writing performance differ between intermediate, upper-intermediate, and advanced learners?
- 3) What are the students' perceptions of ChatGPT-based feedback with regard to their writing development?

Literature Review

AI-Assisted Writing Development: Benefits and Concerns

Although consensus has yet to be reached among researchers and practitioners regarding the effectiveness of generative AI for writing development, both theoretical accounts and empirical evidence have underscored its potential pedagogical value. One of the principal advantages of generative AI systems lies in their time efficiency (Barrot, 2023; Chen & Chen, 2025), as they are capable of delivering immediate feedback on learners' written output, thereby substantially enhancing the feedback cycle for both teachers and learners (Bonner et al., 2023; Kasneci et al., 2023). Notably, generative AI is capable of producing context-sensitive and linguistically natural responses to unpredictable learner input (Kohnke et al., 2023). In contrast to earlier generations of AI-assisted writing tools, such as automated writing evaluation (AWE) systems, which primarily delivered surface-level feedback on grammar, spelling, and punctuation, generative AI, when guided by well-designed prompts, can offer more substantive, discourse-level feedback addressing content development, communicative achievement, organisational coherence, and language use (Jitpaisarnwattana & Saville, 2025 2025; Wang & Bin, 2024). Moreover, from an instructional perspective, generative AI can support teachers by facilitating the creation of writing prompts and the production of model texts for pedagogical use (Kasneci et al., 2023). Owing to its generative capabilities, such systems can also assist learners across multiple stages of the writing process, including planning, idea generation, drafting, and revision, thereby supporting writing development in a more process-oriented manner (Nguyen et al., 2024). In the context of IELTS, generative AI is particularly relevant to writing assessment, which evaluates task achievement/task response, coherence and cohesion, lexical resource, and grammatical range and accuracy. By providing immediate feedback on grammar, vocabulary, organisation, and content, generative AI supports key aspects of writing performance. Its feedback aligns closely with the IELTS assessment criteria, making it a valuable tool for drafting, revision, and test preparation.

Although generative AI has gained increasing prominence in writing instruction and assessment, a number of inherent and practical challenges persist. Chief among these are concerns related to academic integrity arising from the use of AI tools in academic writing. The content-generation capabilities of generative AI may produce inaccurate information and fabricated references that learners do not always detect (Abdelhalim, 2024). Another concern is learners' over-reliance on technology, which may encourage behaviours that undermine academic writing, self-editing, and authentic self-expression, while also constraining critical thinking, creativity, and higher-order writing development (Malik et al., 2023; Marzuki et al., 2023). Excessive dependence on AI has further been linked to risks of "cognitive decline," whereby rapid automated solutions may negatively affect learners' critical thinking and problem-solving abilities (Kasneci et al., 2023). Further concern relates to inconsistencies in AI-generated feedback, as variability in feedback quality may confuse lower-proficiency learners and impede the development of autonomous language learning (Long, 2024).

Empirical Studies on Generative AI and Writing Development

Although research on AI in L2 writing remains at a relatively early stage, scholarly interest in this area has grown rapidly, with particular attention directed towards ChatGPT among other platforms (Hwang et al., 2025). One prominent strand of this research has examined the pedagogical benefits of generative AI for the development of L2 learners' writing performance. A recurring finding across the literature is that generative AI can support improvements in L2 writing performance; however, the scope of these improvements appears uneven across different dimensions of writing. Studies conducted in

China and Malaysia, for example, reported gains across multiple aspects of writing quality, including content, organisation, communicative achievement, and language use (Jen & Salam, 2024; Song & Song, 2023). These findings suggest that generative AI may function as a multifaceted scaffold that supports both linguistic and rhetorical aspects of writing. For instance, Song and Song (2023) examined the effects of AI-assisted language learning on the writing proficiency and writing motivation of Chinese learners of English as a Foreign Language (EFL), reporting statistically significant gains in both domains among students who received AI-supported instruction. In addition, Jen and Salam (2024) examined the role of Google Bard in supporting essay writing among Malaysian learners and found that learners who engaged with the tool demonstrated significant gains in key aspects of writing quality, namely content, communicative achievement, organisation, and language use. However, these largely positive accounts warrant more cautious interpretation. Notably, the extent and nature of improvement appear to vary considerably depending on the specific linguistic features examined. Chen and Chen (2025), for instance, found that generative AI-supported instruction led to gains in syntactic complexity but did not significantly enhance other dimensions of writing performance, suggesting that AI tools may be more effective in supporting form-focused development than higher-order writing skills. This pattern is echoed in Yu (2025), where improvements were similarly confined to syntactic complexity among intermediate-to-lower proficiency IELTS learners. These findings point to an emerging divergence in literature: while some studies report broad-based improvements, others indicate more selective, linguistically bounded effects. The divergence in findings suggests that the effectiveness of generative AI may be highly context-dependent rather than universally beneficial. Studies reporting broad improvements typically involved structured pedagogical integration of AI within instructional activities, whereas studies showing more limited gains often focused on narrower linguistic outcomes or involved shorter interventions. Consequently, it remains unclear whether observed improvements can be attributed to the technology itself or to the pedagogical conditions under which it is implemented. Moreover, the benefits of AI-mediated feedback appear to be moderated by learner proficiency. Research on automated writing evaluation (AWE) systems has suggested that lower-proficiency learners tend to benefit more from automated feedback than their more proficient counterparts, particularly in relation to grammatical accuracy and language form (Ranalli, 2018; Stevenson & Phakiti, 2014). Similarly, Jitpaisarnwattana and Saville (2026) demonstrated that Write & Improve yielded the greatest gains among lower-proficiency learners (IELTS ≤ 5.5), with diminishing returns for higher-proficiency learners (IELTS ≥ 7.0). Collectively, these findings suggest that AI-assisted writing tools may be particularly effective in supporting foundational language development by helping learners identify and address linguistic errors that might otherwise go unnoticed. In contrast, more proficient learners may require feedback that targets higher-order aspects of writing, such as argumentation, content development, and rhetorical effectiveness, areas in which current AI systems may offer less pedagogical value.

Another line of research has looked at L2 learners' behavioral patterns of ChatGPT usage in authentic, real-world contexts. Research examining learners' actual use of ChatGPT has revealed a notable mismatch between the broad capabilities of generative AI and the relatively limited ways in which learners tend to employ the technology. Although ChatGPT is capable of supporting idea generation, argument development, audience awareness, and rhetorical organisation, learners predominantly use it for proof-reading, grammar correction, and language refinement (Koltovskaia et al., 2024; Nguyen et al., 2024). This pattern suggests that learners may not yet possess sufficient AI literacy to exploit the tool's full pedagogical potential, potentially restricting its impact on higher-order writing development. While these studies provide valuable insights into how learners operationalise ChatGPT in real-world tasks, they also reveal a relatively narrow pattern of use. In both cases, engagement appears to be largely oriented towards form-focused support, with learners using the tool primarily as a resource for refining linguistic accuracy and surface-level features of their writing. Taken together, these studies suggest that learners primarily use ChatGPT as a tool for form-focused assistance, with engagement largely

centred on linguistic accuracy and text improvement. However, this pattern also raises questions about whether such use supports the development of higher-order writing skills or merely reinforces reliance on AI for surface-level revision.

Notably, the majority of existing studies have examined how learners view generative AI as a pedagogical resource (Allen & Mizumoto, 2024; Utami & Winarni, 2023; Wang & Bin, 2024;). For instance, Allen and Mizumoto (2024), in a comparative investigation of Japanese tertiary-level learners of English as a second language, examined students' engagement in collaborative peer-based editing alongside individual interactions with ChatGPT for text revision. Their findings indicate a clear learner preference for the use of AI-mediated tools in editing and proofreading tasks, with participants highlighting the systems' capacity to deliver timely and targeted feedback, as well as their perceived effectiveness in enhancing textual clarity and cohesion. Additionally, Yuan et al. (2024) investigated students' motivations for using ChatGPT in academic writing and the perceived benefits of such use, finding that learners were primarily drawn to the tool for its convenience and needs-responsiveness, while reporting scaffolded improvements in writing fluency, grammar, content development, and domain knowledge.

In summary, existing research on AI-assisted writing suggests growing recognition of the pedagogical potential of generative AI in L2 writing. Despite growing evidence of the potential benefits of generative AI for L2 writing (Chen & Chen, 2025; Song & Song, 2023; Wang & Bin, 2024), three important gaps remain. First, much of the existing research has focused on general academic writing rather than high-stakes assessment contexts such as IELTS writing. Second, studies have produced inconsistent findings regarding whether AI supports broad improvements in writing quality or primarily facilitates gains in discrete linguistic features. Third, although learner perceptions have been widely investigated, relatively few studies have examined writing outcomes and learner perceptions concurrently, limiting understanding of how students experience and benefit from AI-generated feedback.

Methodology

Research Contexts and Participants

The participants were English-major students enrolled in the *English Preparation for IELTS* course at a public university in Thailand. A total of 52 students were registered for the course, all of whom were invited to participate in the study. As shown in Table 1, questionnaire data indicated that the sample comprised 40 female and 12 male students aged between 20 and 24 years, the majority of whom had studied English for more than ten years. With regard to prior experience, most participants ($n = 42$) reported having previously used AI tools for writing feedback, whereas 10 students indicated no such experience. Participants' English proficiency levels were determined using scores from the Silpakorn

Table 1. *Demographic information of the participants*

Participants	Age	F	%	Gender	F	%	English Language Proficiency	F	%
	20	25	48.07	Male	12	23.08	A2	0	0.00
	21	22	42.30	Female	40	76.92	B1	15	28.85
	22	4	7.69				B2	27	51.92
	24	1	1.92				C1	10	19.23

Note: F = Frequency.

Test of English Proficiency (STEP; Sridhanyarat et al., 2021): 15 students were classified at the B1 level, 27 at B2, and 10 at C1. The STEP test is a CEFR-based English Proficiency test used to determine the English proficiency level of students at the university. The proficiency levels displayed in the STEP test have been mapped with descriptors in the CEFR framework. The test assesses four language skills: listening, reading, speaking, and writing.

The *English Preparation for IELTS* course is a 16-week elective offered to English major students at the university, with three hours of instruction per week. The course is designed to develop core competencies required for standardised language examinations. The curriculum addresses all four IELTS skills in a structured sequence: reading in Weeks 1–3, writing in Weeks 4–8, listening in Weeks 9–11, and speaking in Weeks 12–14. Week 15 is allocated to course reflection, followed by the final examination in Week 16.

The Intervention

In this study, ChatGPT was implemented as a learning intervention. Over a five-week period, 52 students engaged with the AI tool for independent writing practice. Prior to the intervention, an orientation session was conducted to familiarise participants with the use of the platform, the nature of the feedback provided, and the overall intervention schedule. The researcher delivered an introductory session during the final two hours of Week 8 to explain the intervention requirements. Students were provided with a structured set of prompts to guide their use of ChatGPT for writing feedback (see Appendix A). In addition, they were given a pool of ten Task 1 and ten Task 2 questions from which to choose. Participants were required to complete at least one Task 1 and one Task 2 writing task per week, although they were permitted to submit additional tasks if they wished. Participants were permitted to revise and resubmit their drafts as many times as they wished. The five-week intervention was implemented from Weeks 8 to 12. Students were also instructed to document their interactions with ChatGPT via screen capture and to keep records of both the number of writing tasks completed and the frequency of revisions made to each piece of writing.

Importantly, as the study lacked a control group, measures were taken to reduce threats to the study's validity. First, the intervention was implemented within a clearly bounded five-week period without additional targeted IELTS writing instruction. The students were also instructed not to use any other writing assistants during the intervention period. This limited external instructional influences, though not entirely. Moreover, the collection of student's engagement records with ChatGPT allowed the author to document students' interactions and ensure that the extent of their engagement with the intervention met the expected level of interaction with the program. Despite these mitigating strategies, the findings of this study should still be viewed with caution as other external factors might have influenced their performance in the post-test.

Data Collection and Analysis

A pre-test was administered in Week 9 of the course, with a corresponding post-test conducted in Week 13; both assessments were carried out in a face-to-face classroom setting. The researcher supervised the administration of both tests. Participants were required to complete two writing tasks (Task 1 and Task 2) within a 60-minute time limit. The participants completed both tests without AI assistance. All scripts were evaluated independently by two certified IELTS examiners in accordance with the official IELTS writing band descriptors. Scores were recorded for each assessment criterion—task achievement/response, lexical resource, grammatical range and accuracy, and coherence and cohesion. The data were analysed using SPSS version 31.0 (Statistical Package for the Social Sciences). Paired-samples *t*-tests were then performed to examine differences between pre-test and post-test scores across each criterion.

At the conclusion of the intervention in Week 13, the attitude questionnaire was administered. The instruments were distributed online using Google Forms, and all participants ($N = 52$) completed them in a classroom setting immediately after the post-test. Responses to the Likert-scale items were analysed using SPSS version 31.0, with descriptive statistics computed to summarise the results. The open-ended responses were deductively coded according to three predefined categories including writing improvement, feedback usefulness, and feedback quality. It is also important to note that the researcher also served as the instructor for the course in which the study was conducted. To minimise potential power imbalances and reduce the likelihood of socially desirable responses, participants were informed that their participation was entirely voluntary, that their decision to participate or withdraw would have no effect on their course grades or academic standing, and that all responses would be treated confidentially.

Seven participants were purposively selected for interviews based on their questionnaire responses. Participants who demonstrated positive, neutral, and negative attitudes were selected. The proficiency levels of the selected interviewees ranged from B1 to C1 (two B1 participants, three B2 participants, and two C1 participants). Individual semi-structured interviews were conducted online via Zoom (see Appendix B). To minimise the potential for miscommunication or misunderstanding, the interviews were conducted in Thai by the researcher. All interviews were transcribed and subsequently translated into English by the researcher. In line with the analysis of the open-ended questionnaire responses, the interview data were deductively coded according to the three predefined categories. To strengthen the reliability of the coding process, the researcher carried out the initial coding, and 50% of the data were independently coded by an English lecturer at Silpakorn University as part of a dual-coding procedure. It should be noted that the interview scripts presented in the subsequent sections were a representation of opinions gathered from the interviews.

Ethical Considerations

Ethical clearance was secured from the Silpakorn University Ethics Committee before the implementation of the study. Prior to participation, all students were provided with detailed information about the research and were required to give informed consent. They were explicitly informed that their involvement was entirely voluntary and that they retained the right to withdraw from the study at any point, including after the research had begun. Although participant identities were initially known to the researchers, anonymity was subsequently ensured through the use of pseudonyms during data handling and reporting. The researcher was also the instructor for the course. Despite the dual role, the role as a researcher only became prominent during questionnaire administration and semi-structured interviews.

Results

The Effects of AI-Generated Feedback on Students' Writing Performance

To determine whether statistically significant differences emerged between learners' pre-test and post-test writing performance, a series of paired-samples *t*-tests was employed. These analyses were conducted across eight analytic scoring dimensions. Both IELTS writing tasks share the same four core constructs: task achievement/response, lexical resources, grammatical range and accuracy, and coherence and cohesion.

As shown in Table 2, although a marginal increase in Task Achievement scores was observed in the post-test ($M = 6.05$, $SD = 0.849$) compared to the pre-test ($M = 6.02$, $SD = 0.83$), the difference was not statistically significant at the .05 level, $t(51) = -0.81$, $p = .420$. These results indicate that, for

Table 2. Results from a paired-sample *t*-test on the scores for the four criteria of writing task 1

Criteria	Tests	<i>N</i>	<i>M</i>	<i>SD</i>	<i>T</i>	<i>df</i>	Sig. (Two-sided <i>p</i>)	Cohen's <i>d</i>
Task Achievement	Pre-test	52	6.02	.83	-0.814	51	.420	0.11
	Post-test	52	6.05	.85				
Lexical Resources	Pre-test	52	6.26	.74	-4.201	51	.001	0.58
	Post-test	52	6.63	.86				
Grammatical Range and Accuracy	Pre-test	52	6.01	.73	-7.714	51	.001	1.07
	Post-test	52	6.55	.98				
Coherence and Cohesion	Pre-test	52	5.98	.89	-6.648	51	.001	0.92
	Post-test	52	6.48	1.09				

Note: *N* = number of subjects, *M* = mean of scores, *SD* = Standard deviation, *t* = *t* scores, *df* = degrees of freedom, sig. = significance.

this cohort, exposure to AI-generated feedback did not yield a statistically meaningful improvement in Task Achievement performance. In contrast, participants demonstrated a statistically significant improvement in Lexical Resource scores from the pre-test ($M = 6.26$, $SD = 0.74$) to the post-test ($M = 6.63$, $SD = 0.86$). The paired-samples *t*-test revealed that this increase was significant at the .05 level, $t(51) = -4.20$, $p < .001$ and a medium effect size ($d = 0.58$). This finding suggests that the AI-mediated intervention may be associated with a meaningful enhancement in participants' lexical resource performance in the post-test.

Similarly, the results indicate that participants achieved substantially higher scores in Grammatical Range and Accuracy in the post-test ($M = 6.55$, $SD = 0.98$) than in the pre-test ($M = 6.01$, $SD = 0.73$), with the paired-samples *t*-test confirming that this difference was statistically significant the .05 level, $t(51) = -7.71$, $p < .001$ and a large effect size ($d = 1.07$). These findings suggest that engagement with ChatGPT was associated with marked improvements in both grammatical accuracy and the breadth of syntactic structures employed in the post-test writing tasks. On top of that, participants obtained significantly higher scores for Coherence and Cohesion in the post-test ($M = 6.48$, $SD = 1.09$) than in the pre-test ($M = 5.98$, $SD = 0.89$). The paired-samples *t*-test confirmed that this improvement was statistically significant at the .05 level, $t(51) = -6.65$, $p < .001$ and a large effect size ($d = 0.92$). This finding indicates that engagement with ChatGPT may be associated with enhanced organisational clarity and more effective use of cohesive devices in participants' post-test writing.

As seen in Table 3, although participants recorded a marginal increase in Task Response scores in the post-test ($M = 6.34$, $SD = 0.93$) relative to the pre-test ($M = 6.30$, $SD = 0.93$), this difference did not attain statistical significance at the .05 level, $t(51) = -0.44$, $p = .330$. These results suggest that, while some degree of improvement was observed following engagement with ChatGPT, the gains in Task Response were not sufficiently robust to be considered statistically meaningful. The participants, however, achieved higher Lexical Resource scores in the post-test ($M = 6.82$, $SD = 0.73$) than in the pre-test ($M = 6.42$, $SD = 0.63$). The paired-samples *t*-test indicated that this difference was statistically significant at the .05 level, $t(51) = -5.12$, $p < .001$ and was accompanied by a medium effect size ($d = 0.71$). Taken together, these findings suggest that engagement with ChatGPT-generated feedback was associated, though not entirely, with a meaningful improvement in participants' lexical resource performance in the post-test.

Table 3. Results from a paired-sample *t*-test on the scores for the four criteria of writing task 2

Criteria	Tests	<i>N</i>	<i>M</i>	<i>SD</i>	<i>T</i>	<i>df</i>	Sig. (Two-sided <i>p</i>)	Cohen's <i>d</i>
Task Response	Pre-test	52	6.30	.93	−0.444	51	.330	0.06
	Post-test	52	6.34	.93				
Lexical Resources	Pre-test	52	6.42	.63	−5.117	51	.001	0.71
	Post-test	52	6.82	.73				
Grammatical Range and Accuracy	Pre-test	52	6.13	.71	−6.959	51	.001	0.96
	Post-test	52	6.76	.98				
Coherence and Cohesion	Pre-test	52	6.27	.68	−8.339	51	.001	1.16
	Post-test	52	6.72	.74				

Note: *N* = number of subjects, *M* = mean of scores, *SD* = Standard deviation, *t* = *t* scores, *df* = degrees of freedom, sig. = significance.

Moreover, the participants exhibited markedly higher levels of Grammatical Range and Accuracy in the post-test ($M = 6.76$, $SD = 0.98$) than in the pre-test ($M = 6.13$, $SD = 0.71$). The paired-samples *t*-test confirmed that this difference was statistically significant at the .05 level, $t(51) = -6.96$, $p < .001$ and was associated with a large effect size ($d = 0.96$). Collectively, these findings indicate that engagement with ChatGPT-based feedback over the five-week intervention period was linked to substantive gains in participants' grammatical range and accuracy. Finally, the participants' post-test scores ($M = 6.72$, $SD = 0.74$) were substantially higher than their pre-test scores ($M = 6.27$, $SD = 0.68$). The paired-samples *t*-test revealed that this increase was statistically significant at the .05 probability level, $t(51) = -8.33$, $p < .001$ and was accompanied by a large effect size ($d = 1.16$). These results indicate that engagement with ChatGPT-based feedback was associated, though not entirely, with meaningful improvements in the coherence and cohesion of participants' writing performance.

The Effects of AI-Generated Feedback on Different Groups of Students

To investigate the differential effects of AI-generated feedback across proficiency levels, participants were stratified into three groups according to their pre-test performance: intermediate (≤ 5.5), upper-intermediate (6.0–6.5), and advanced (≥ 7.0). Descriptive statistical analyses were subsequently employed to examine the distribution of students within each proficiency group who demonstrated gains in their post-test scores.

As shown in Table 4, the vast majority of students in the ≤ 5.5 proficiency group (80%) demonstrated score gains in the post-test. A similarly high proportion of improvement was observed among students in the 6.0–6.5 group, with 81% achieving higher overall scores in the post-test. In contrast, only half

Table 4. Score improvements by different groups of students (task 1)

Group	<i>N</i> (52)	Increased	Unchanged	Decreased
5.5 and below	15	12 (80%)	3 (20%)	–
6.0–6.5	27	22 (81%)	5 (19%)	–
7.0 and above	10	5 (50%)	4 (40%)	1 (10%)

of the students in the ≥ 7.0 group obtained higher post-test scores, while the remaining participants showed no improvement or a decline in performance. These patterns suggest that, for Task 1, learners at intermediate and upper-intermediate proficiency levels (≤ 5.5 and $6.0\text{--}6.5$) derived the greatest benefit from engagement with ChatGPT-based feedback, whereas the effects were less pronounced among higher-proficiency students.

As presented in Table 5, a substantial majority of students in the ≤ 5.5 proficiency group (83.34%) achieved higher scores in the post-test following five weeks of engagement with ChatGPT-based feedback. Consistent with the findings for Task 1, approximately two thirds of students in the $6.0\text{--}6.5$ group also demonstrated score gains in the post-test. In contrast, fewer than half of the students in the advanced group (≥ 7.0) showed improvement, with only 41.67% attaining higher post-test scores, while a larger proportion (58.33%) exhibited no change or a decline in performance. These results further indicate that, although ChatGPT feedback was beneficial for learners at intermediate (≤ 5.5) and upper-intermediate ($6.0\text{--}6.5$) proficiency levels, its impact was comparatively less pronounced for advanced learners (≥ 7.0).

Table 5. Score improvements by different groups of students (task 2)

Group	N (52)	Increased	Unchanged	Decreased
5.5 and below	12	10 (83.34%)	1 (8.33%)	1 (8.33%)
6.0–6.5	28	20 (71.43)	7 (25%)	1 (3.57%)
7.0 and above	12	5 (41.67%)	6 (50%)	1 (8.33%)

Students' Perceptions of AI-Generated Feedback

The questionnaire comprised four items designed to measure participants' perceptions of the extent to which engagement with ChatGPT contributed to improvements in their writing. Overall, the results revealed a high level of agreement across all four statements. A substantial majority of participants (82.70%) perceived that AI-generated feedback was beneficial for enhancing their overall writing

Table 6. ChatGPT and writing improvement

		SD	D	N	A	SA	Mean	STD
Q1	F	0	1	8	30	13	4.05	.697
	Percentage	0.0	1.92	15.38	57.70	25		
Q2	F	0	1	10	26	15	4.05	.751
	Percentage	0.0	1.92	19.23	50	28.85		
Q3	F	0	2	6	23	21	4.21	.800
	Percentage	0.0	3.85	11.54	44.23	40.38		
Q4	F	0	0	8	32	12	4.07	.621
	Percentage	0.0	0.0	15.38	61.54	23.08		

Note: F = frequency, SD = strongly disagree, D = disagree, N = neither agree nor disagree, A = agree, SA = strongly agree, STD = standard deviation.

Q1: ChatGPT helped me improve my writing skill.

Q2: ChatGPT helped me improve my vocabulary in writing.

Q3: ChatGPT helped me improve my grammar and sentence structure.

Q4: ChatGPT helped me improve the coherence of my writing.

performance. Comparable levels of agreement were observed for specific writing components, with 84.61% of respondents indicating improvements in sentence structure and grammatical accuracy, 78.85% reporting perceived gains in vocabulary use, and 84.62% perceiving enhanced coherence in their writing.

Responses to the open-ended questionnaire items and the interview data revealed a recurring theme, whereby many students reported that engagement with AI-generated feedback was particularly beneficial for improving their grammar and sentence structure. For instance, one student remarked:

I feel that my greatest improvement has been in grammar and sentence structure. As ChatGPT not only corrected my errors, but also provided explanations as to why they were wrong, I became more conscious of the mistakes I was making and made a conscious effort not to repeat them. Moreover, it provides feedback in bullet points, which makes it easy to read and understand. As a result, I now have a much clearer understanding of my grammatical weaknesses. (Interview, student 23, Upper-intermediate group).

In addition, a number of participants reported that ChatGPT-based feedback contributed to the expansion of their lexical repertoire, particularly when they were able to effectively interpret and apply the AI's suggestions, as illustrated by the following participant account:

I think using ChatGPT can really help me improve my vocabulary, especially for IELTS writing. ChatGPT not only provided scores on my lexical resource criteria, but also suggested words that I could use to improve my score in this area. I used the suggested words in my revisions. Also, because I used the words in my revised drafts, I remember the words as well, not all of them, but some of them. (Interview, student 2, Upper-intermediate group).

Four items in the questionnaire asked about the usefulness of feedback from ChatGPT. The students generally thought that the feedback was useful, especially on improving their overall writing and grammar. The majority of the students (80.76%) felt that the feedback from ChatGPT was useful for

Table 7. ChatGPT and usefulness of feedback

		SD	D	N	A	SA	Mean	STD
Q5	F	0	2	8	21	21	4.17	.833
	Percentage	0.0	3.85	15.39	40.38	40.38		
Q6	F	0	4	15	19	14	3.82	.982
	Percentage	0.0	7.69	28.85	36.54	26.92		
Q7	F	1	3	9	20	19	4.01	.708
	Percentage	1.92	5.77	17.31	38.46	36.54		
Q8	F	0	2	15	23	12	3.86	.980
	Percentage	0.0	3.85	28.85	44.23	23.07		

Note: F = frequency, SD = strongly disagree, D = disagree, N = neither agree nor disagree, A = agree, SA = strongly agree, STD = standard deviation.

Q5: Feedback from ChatGPT was useful for my writing.

Q6: Feedback from ChatGPT was useful for improving my coherence.

Q7: Feedback from ChatGPT was useful for improving my grammar.

Q8: Feedback from ChatGPT was useful for improving my vocabulary.

their overall writing. Also, a large percentage of the students (75%) thought that the feedback was particularly useful for improving their grammar. Nevertheless, a less strong level of agreement was evident in the usefulness of feedback on coherence with 63.46% agreeing with the statement, whereas 28.85% feeling unsure and 7.69% expressing disagreement, respectively.

The qualitative findings revealed that students perceived the feedback generated by ChatGPT as pedagogically valuable for their writing development overall, with particular emphasis on its contribution to grammatical improvement. This perception was evidenced in students' accounts, as illustrated by the following excerpts:

I think that the feedback from ChatGPT was useful for improving my writing in general. In particular, I think it did a great job at spotting my grammatical mistakes both at word and sentence levels. (Interview, student 35, Intermediate group).

What I liked about feedback from ChatGPT was that it did not only identify mistakes, but it also suggested some possible sentences I could use to improve my writing. For me, I know all the grammar rules, but sometimes I do not know how to form complex structures, so having examples to follow is really useful. (Open-ended response, student 18, Upper-intermediate group).

Nevertheless, some students did not find ChatGPT feedback useful in improving other writing components, mainly because of the nature of the feedback and their lack of skills to refine prompts.

I do not think that my coherence improved that much in the five weeks. Yes, ChatGPT gave me scores and provided me feedback on how to improve my coherence, but I still found it difficult to use it in my writing. Maybe, I need to use more prompts to get more concrete feedback. (Interview, student 2, Advanced group).

Four questionnaire items were designed to elicit students' perceptions of the nature and quality of ChatGPT-based feedback. Descriptive statistical analysis indicated that a majority of participants (73.08%) agreed that AI-generated feedback was mechanistic rather than humanistic, although approximately one fifth of the respondents reported uncertainty regarding this characterisation. This perception

Table 8. ChatGPT and feedback quality

		SD	D	N	A	SA	Mean	STD
Q9	F	0	3	11	18	20	4.05	.916
	Percentage	0.0	5.77	21.15	34.62	38.46		
Q10	F	0	2	14	22	14	3.92	.836
	Percentage	0.0	3.85	26.92	42.31	26.92		
Q11	F	0	4	7	22	19	4.07	.904
	Percentage	0.0	7.69	13.46	42.31	36.54		
Q12	F	1	1	12	27	11	3.88	.832
	Percentage	1.92	1.92	23.08	51.93	21.15		

Note: F = frequency, SD = strongly disagree, D = disagree, N = neither agree nor disagree, A = agree, SA = strongly agree, STD = standard deviation.

Q9: I think the feedback from ChatGPT is mechanistic, not humanistic.

Q10: I think ChatGPT is interactive because the feedback is instant.

Q11: I think the feedback from ChatGPT is not comparable to human interaction.

Q12: I think the feedback from ChatGPT can highlight my weaknesses very well.

was further reinforced by a high level of agreement (78.85%) with the statement that ChatGPT-based feedback could not be equated with human interaction. In contrast, a substantial proportion of students acknowledged the pedagogical affordances of AI feedback, particularly its immediacy and diagnostic value. Specifically, 69.23% of participants agreed that the instantaneous nature of AI-generated feedback enhanced interactivity, while 73.08% reported that it effectively helped them identify weaknesses in their writing.

The qualitative findings corroborate the quantitative results, indicating that students perceived ChatGPT-generated feedback as largely mechanistic and not directly comparable to feedback provided by human teachers, particularly with respect to higher-level aspects of writing such as content development. For example, one student mentioned:

If I compare this with the feedback I previously received from teachers, I would say that teacher feedback was more helpful, particularly in addressing content-related issues. I feel that I need greater support with developing ideas and content rather than focusing mainly on grammar, and overall, I found human feedback to be more effective than AI-generated feedback. (Interview, student 7, Advanced group).

Moreover, many students reported that the immediacy of AWE feedback enabled more effective engagement with the feedback process and allowed them to identify their writing weaknesses more clearly, as reflected in the following students' comment:

The best benefit is the ability to provide quick feedback on my writing. To get feedback right away after I have just finished writing means that I can still remember how and why I write certain things very well. I can relate more with the feedback I received. (Open-ended responses, student 12, Upper-intermediate group).

What I liked about my interaction with the program was that it could effectively identify my writing problems very well. Also, not only did it spot my weaknesses, but it also provided suggestions on how I could improve my writing. (Interview, student 17, Advanced group).

Discussion

The observed post-test gains across all four assessment criteria in both IELTS Task 1 and Task 2 provide initial empirical evidence that AI-generated feedback may have had a positive influence on learners' writing performance. With respect to Task 1, inferential statistical analyses indicated statistically significant improvements across the three language-focused dimensions of writing quality with moderate to large effect sizes. The most pronounced gains were identified in Grammatical Range and Accuracy and Coherence and Cohesion, suggesting that AI-mediated feedback may be particularly useful in supporting learners' control over linguistic form and textual organisation. However, despite a marginal increase in Task Achievement, this improvement did not reach statistical significance. With regard to Task 2 writing performance, the statistical analyses revealed significant post-test improvements in three of the four assessment criteria. Although a modest increase was observed in Task Response (a gain of 0.04 points), this change did not attain statistical significance. In line with the pattern identified for Task 1, the largest gains were recorded in Grammatical Range and Accuracy (0.63-point increase) and Coherence and Cohesion (0.45-point increase), suggesting that the AI-generated feedback was more likely to enhance learners' linguistic accuracy and discourse-level organisation rather than task-specific content fulfilment. From an empirical standpoint, given that no explicit writing instruction was provided, it may be reasonable to suggest that the observed improvements

in post-test writing performance were associated with learners' engagement with ChatGPT-mediated feedback, albeit not attributable to this factor alone.

These findings give rise to several important observations. Most notably, the largest score gains were evident in Grammatical Range and Accuracy, suggesting that AI-generated feedback, in its current iteration, may be particularly well suited to addressing grammar- and structure-related aspects of writing. This finding is consistent with prior research (e.g. Liu et al., 2024; Wang & Bin, 2024) highlighting the significant role of generative AI tools in enhancing writing quality, particularly with respect to grammatical accuracy and structural refinement. Secondly, the comparatively limited gains observed in Task Achievement (Task 1) and Task Response (Task 2) were not unexpected. Given the nature of AI-generated feedback, its capacity to evaluate the extent to which a written response fully addresses a task prompt may be inherently constrained. Such judgements arguably demand a more human-centred, interpretive approach, rather than the predominantly form-focused feedback typically generated by current AI systems. Moreover, it is plausible that content-oriented dimensions of writing are less amenable to measurable improvement within a relatively short instructional timeframe, such as the five-week intervention examined in this study. It is also possible that students might not be able to interpret and apply content-related feedback effectively in their revised drafts, requiring further scaffolding from teachers. The present findings appear to partially corroborate those reported by Jen and Salam (2024), in that statistically significant gains were observed in grammatical accuracy, coherence, and lexical resource, whereas improvements in content-related dimensions were not as evident. Furthermore, the pronounced improvements in grammar and structural control appear to be consistent with the findings of Zhang et al. (2025), who similarly reported that generative AI-mediated feedback enhanced learners' writing performance, particularly with respect to grammatical accuracy and sentence-level variety. Theoretically, this pattern may be interpreted through the lens of the Noticing Hypothesis (Schmidt; 1990; Truscott, 1998), which posits that learners are more likely to benefit from feedback that is salient, explicit, and immediately noticeable. Grammatical corrections provided by AI are typically localised and form-focused, thereby increasing their perceptual salience and potentially facilitating uptake. In contrast, feedback on content and coherence often operates at a more global, discourse-level, requiring deeper cognitive processing and integration across the text, which may reduce immediate noticing and uptake.

Analysis of AI-generated feedback effects across proficiency bands indicated that, for both Task 1 and Task 2, the AI-generated feedback appeared to be most effective in improving overall scores among intermediate learners (≤ 5.5), followed by the upper-intermediate group (6.0–6.5). By contrast, improvements were less pronounced among advanced learners (7.0+), with approximately half of the participants in this group demonstrating measurable post-test improvement. This pattern may be partly explained by the nature of AI-generated feedback (reflected in the significant gains observed across three language-related criteria) that primarily targets linguistic accuracy, surface-level mechanics, and lexical refinement, areas that are particularly salient for learners at the intermediate to upper-intermediate proficiency levels. Nevertheless, learners at higher proficiency levels may have already consolidated these linguistic features and therefore require more targeted feedback addressing content development and higher-order rhetorical dimensions of writing. The results are consistent with prior research on both automated writing evaluation (AWE) and AI-generated feedback, which has shown that such feedback tends to be more effective for intermediate and upper-intermediate learners than for those at advanced proficiency levels (Chen & Chen, 2025; Lavolette, 2015; Yu, 2025; Zhang et al., 2025). Moreover, as Steiss et al. (2024) suggest, ChatGPT may demonstrate greater effectiveness in delivering feedback on lower-quality writing, whereas its effectiveness tends to decline when applied to higher-quality texts.

With respect to learners' perceptions, participants expressed generally positive views of AI-generated feedback in relation to their IELTS writing development. In particular, students reported

marked improvement in grammatical knowledge, with a high level of agreement across perception measures. This outcome may be attributed to the explicit identification and explanation of grammatical errors, which may have facilitated learners' ability to notice, understand, and subsequently revise problematic language forms in their writing. Moreover, the students also thought that AI-generated feedback was beneficial for their vocabulary and coherence, albeit to a lesser degree. Notably, these positive perceptions of grammatical, lexical, and organisational development extended beyond perceptions and appeared to be reflected in corresponding performance gains in Grammatical Range and Accuracy, Lexical Resource, and Coherence and Cohesion across both tasks.

In terms of the usefulness of feedback, the students found AI-generated feedback most useful in improving their grammar and overall writing, and less so for vocabulary and coherence. This could be due to the explicit nature of grammatical feedback, making it more effective to incorporate into subsequent revisions. On the other hand, although the feedback on vocabulary and coherence might have been of good quality, it is likely more challenging to assimilate and may require a longer period for its benefits to become evident. These findings, both in terms of writing improvement and perceived feedback usefulness, echoed results from previous studies on students' perceptions of AI-generated feedback, which has consistently reported positive evaluations of AI support for grammar and cohesion (Allen & Mizumoto, 2024; Utami & Winarni, 2023; Wang & Bin, 2024; Yuan et al., 2024). However, feedback targeting cohesion was perceived as comparatively less useful in the present study.

Regarding feedback quality, although students perceived ChatGPT-generated feedback as effective in identifying linguistic weaknesses, they also regarded it as predominantly mechanistic and not comparable to teacher feedback. This finding is perhaps unsurprising, as students used pre-formulated prompts designed primarily to generate feedback; alternative prompting strategies aimed at eliciting more teacher-like responses may have produced different perceptions. Qualitative findings further revealed that only advanced learners (Band 7.0+) considered AI-generated feedback insufficient, suggesting that this group may require more human-centred guidance, particularly for content development, stylistic nuance, and rhetorical organisation. These findings are consistent with Zhang et al. (2025), who found that while generative AI-generated feedback adequately supports basic writing skills, hybrid feedback models combining teacher and AI input may be more effective for developing more complex academic writing abilities.

Pedagogical Implications and Conclusion

The findings indicate that while generative AI-generated feedback can enhance IELTS-style writing through improvements in grammatical accuracy, sentence-level variety, lexical choice, and coherence, its effects on higher-order writing skills remain limited. This suggests that a hybrid feedback model (teacher + AI) might be essential for comprehensive writing development. In this regard, teachers can use generative AI tools to provide immediate, automated feedback on lower-order skills (such as grammar and mechanics), while they focus more on providing feedback related to content and task fulfillment. Implementing a hybrid feedback model can both reduce teachers' workload and enhance feedback efficiency and overall quality. Furthermore, writing teachers should teach students how to interpret and apply AI-generated feedback effectively, especially content-related feedback. Targeted training can be organised to ensure understanding of how to integrate generative AI feedback in the writing process and foster students' feedback literacy. This scaffolded approach is crucial if generative AI tools are to be adopted as independent learning tools for writing development (Chiu & Chai, 2020).

There are several limitations in this study that should be acknowledged. First, although the students were instructed not to use other writing support tools during the intervention, it is possible that they may have received additional instructional support from external sources. Therefore, the observed

score gains cannot be attributed exclusively to ChatGPT-generated feedback. Second, due to administrative constraints, the study did not include a control group, preventing direct comparison between students receiving AI-generated feedback and those following a business-as-usual instructional approach. Future research incorporating both experimental and control groups is recommended. Moreover, the extent to which these findings generalise to other AI models remains uncertain, as the study examined only a single system (*ChatGPT Version 4.1*). Finally, the five-week duration of the study may have been insufficient to capture substantial language development. Hence, future research should adopt longitudinal designs spanning a full semester to better examine the long-term effects of AI-generated feedback on IELTS writing.

Use of Generative AI

This manuscript made use of ChatGPT and Gemini to assist in paraphrasing and improving the clarity of text. No AI tools were used for the generation of research hypotheses or the writing of conclusions. The final manuscript was authored and reviewed by the listed authors without additional AI input.

References

- Abdelhalim, S. M. (2024). Using ChatGPT to promote research competency: English as a Foreign Language undergraduates' perceptions and practices across varied metacognitive awareness levels. *Journal of Computer Assisted Learning*, 40(3), 1261–1275. <https://doi.org/10.1111/jcal.12948>
- Allen, T.J. & Mizumoto, A. (2024). ChatGPT over my friends: Japanese English-as-a-Foreign-Language learners' preferences for editing and proofreading strategies. *RELC Journal*, 57(1), 1–18 <https://doi.org/10.1177/00336882241262533>.
- Andriani, E. (2025). AI for IELTS writing preparation: Reviewing ChatGPT-generated answer mock-ups and its text evaluation. *ELTR Journal*, 9(2), 86–112. <https://doi.org/10.37147/eltr.v9i2.260>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *Teaching English with Technology*, 23(1), 23–41. <https://doi.org/10.56297/BKAM1691/WIEO1749>
- Chen, Y., & Chen, J. (2025). Generative AI in EFL writing instruction for lower-intermediate learners: Do perceived affective gains translate to improved performance? *Language Teaching Research*, 29(7), 1–26. <https://doi.org/10.1177/13621688251397724>
- Chiu, T. K. F., & Chai, C. S. (2020). Sustainable curriculum planning for artificial intelligence education: A Self-Determination Theory perspective. *Sustainability*, 12(14), Article 5568. <https://doi.org/10.3390/su12145568>.
- Hockly, N. (2023). Artificial intelligence in English language teaching: The good, the bad and the ugly. *RELC Journal*, 54, 445–451. <https://doi.org/10.1177/00336882231168504>
- Hwang, H., Chang, X., & Sun, J. (2025). Generative AI is useful for L2 writing, but when, why, and for how long do learners use it? *Journal of Second Language Writing*, 69, Article 101230. <https://doi.org/10.1016/j.jslw.2025.101230>
- Jitpaisarnwattana, N. (2024). Idea sharing: Using ChatGPT for diagnostic writing assessment. *PASAA*, 69, 561–574. <https://doi.org/10.58837/CHULA.PASAA.69.18>
- Jitpaisarnwattana, N., & Saville, N. (2025). Autonomous learning and students' perceptions of automated writing evaluation as a tool to improve writing skills. *Technology in Language Teaching & Learning*, 7(3), Article 102826. <https://doi.org/10.29140/tltl.v7n3.102826>

- Jitpaisarnwattana, N., & Saville, N. (2026) Using automated writing evaluation to improve EFL students' IELTS writing: Effects across proficiency levels and learning behaviours. *LEARN Journal: Language Education and Acquisition Research Network*, 19(2).
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, S., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kohnke, L., Moorhouse, B. L., & Zou, D. D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence*, 5, Article 100156. <https://doi.org/10.1016/j.caeai.2023.100156>.
- Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of ChatGPT for academic text revision: Behavioral, cognitive, and affective engagement. *Journal of Second Language Writing*, 65, Article 101130. <https://doi.org/10.1016/j.jslw.2024.101130>
- Jen, S. L., & Salam, A. R. H. (2024). Using Google Bard to improve secondary school students' essay writing performance. *Journal of Creative Practices in Language Learning and Teaching*, 12, 91–112. <https://doi.org/10.24191/cplt.v12i1.24999>
- Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language, Learning & Technology*, 19(2), 50–68. <https://doi.org/10.64152/10125/44417>
- Liu, Z. M., Hwang, G. J., Chen, C. Q., Chen, X. D., & Ye, X. D. (2026). Integrating large language models into EFL writing instruction: Effects on performance, self-regulated learning strategies, and motivation. *Computer Assisted Language Learning*, 39(3), 466–490. <https://doi.org/10.1080/09588221.2024.2389923>
- Long, H. S. (2024). Exploring the use of ChatGPT as a tool for written corrective feedback in an EFL classroom. *The Journal of Asia TEFL*, 21(2), 397–412. <https://doi.org/10.18823/asiatefl.2024.21.2.8.397>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, Article 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki. (2023). Exploring artificial intelligence in academic essay: Higher education student's perspective. *International Journal of Educational Research Open*, 5, Article 100296. <https://doi.org/10.1016/j.ijedro.2023.100296>
- Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2), Article 2236469. <https://doi.org/10.1080/2331186X.2023.2236469>
- Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*, 3, Article 100116. <https://doi.org/10.1016/j.rmal.2024.100116>
- Montakantiwong, A., & Jitpaisarnwattana, N. (2026). Humanising language teaching in the age of AI. *PASAA Journal*, 72, 1–18. <https://doi.org/10.66947/pasaa.v72i1.3279>
- Nguyen, L. Q., Le, H. V., & Nguyen, P. T. (2024). A mixed-methods study on the use of chatgpt in the pre-writing stage: EFL learners' utilization patterns, affective engagement, and writing performance. *Education and Information Technologies*, 30(8), 10511–10534. <https://doi.org/10.1007/s10639-024-13231-8>.
- Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Computer Assisted Language Learning*, 31(7), 653–674. <https://doi.org/10.1080/09588221.2018.1428994>

- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, Article 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>.
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, Article 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Stevenson, M. & Phakiti, A. (2014) The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51–65. <https://doi.org/10.1016/j.asw.2013.11.007>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, Article 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Sun, Q. (2024). Exploring human-generative AI interaction in L2 learners' source use practices: Issues, trials, and critical reflections. *Journal of Academic Writing*, 14(1), 24–42. <https://doi.org/10.18552/joaw.v14i1.1055>
- Truscott, J. (1998). Noticing in second language acquisition: A critical review. *SLA Research* 14, 103–135. <https://doi.org/10.1191/026765898674803209>
- Utami, S. P. T., & Winarni, R. (2023). Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive? *Contemporary Educational Technology*, 15(4), Article 450. <https://doi.org/10.30935/cedtech/13419>
- Wang, J. & Bin, L. (2024). An empirical study on continuation writing in senior high school under the assessment and feedback of ChatGPT. *Education Study*, 7(1), 70–73. <https://doi.org/10.35534/es.0701014>
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28, 13943–13967. <https://doi.org/10.1007/s10639-023-11742-4>
- Yu, Y. (2025). The role of generative artificial intelligence in IELTS writing. *International Journal of Multilingual Education*, 26, 88–103. <https://doi.org/10.22333/ijme.2025.10426>
- Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: a comparative study. *Innovation in Language Learning and Teaching*, 1–19. <https://doi.org/10.1080/17501229.2025.2503890>

Appendix A

Prompts for Students

Please use the following prompts when asking ChatGPT to provide feedback on your writing

Task 1:

Prompt 1: Here is the IELTS Writing Band Descriptor for IELTS writing Task 1 (Upload the provided file).

Prompt 2: Please provide feedback, corrections and band scores on this IELTS task 1 writing based on the following criteria 1) Task Achievement, 2) Lexical Resources, 3) Coherence and Cohesion and 4) Grammatical Range and Accuracy.

Prompt 3: Upload the question file from the list of task 1 questions provided (ChatGPT has to read the data).

Task 2:

Prompt 1: Here is the IELTS Writing Band Descriptor for IELTS writing Task 2 (Upload the provided file).

Prompt: Please provide feedback and corrections, and band scores on this IELTS task 2 writing based on the following criteria 1) Task Response, 2) Lexical Resources, 3) Coherence and Cohesion and 4) Grammatical Range and Accuracy.

Prompt 3: The question is—insert your question—(Copy and paste from the tests).

Appendix B

Semi-Structured Interview Questions

1. What do you think about the feedback you received from ChatGPT?
2. Do you think that using ChatGPT helped you improve your writing skills? If so, how?
3. Do you think the feedback from ChatGPT was useful for developing your writing skills?
4. How would you compare the feedback you received from ChatGPT with the feedback you previously received from your teachers?
5. Questions emerged from students' responses to the questionnaires.