Enhancing IELTS writing task 2 performance through AI-generated feedback: A mixed-methods study on learner improvement and perceptions

Sisila Fitriany Damanik[1*], Ariatna[2], Indra Hartoyo[2], Nazlah Syahaf Nasution[2]
[1]English Literature Study Program, Faculty of Languages and Arts, Universitas Negeri Medan, Kabupaten Deli Serdang, Sumatera Utara 20221
[2]English Education Study Program, Faculty of Languages and Arts, Universitas Negeri Medan, Kabupaten Deli Serdang, Sumatera Utara 20221

ABSTRACT
This study investigates the impact of AI-generated feedback on the performance of English as a Foreign Language (EFL) learners in IELTS Writing Task 2 and on their perceptions of the feedback process. Employing a convergent mixed-method design, the research involved 18 undergraduate students who completed four iterative writing tasks using official IELTS prompts. After each submission, participants received immediate feedback from a custom-developed AI tool powered by GPT-4, aligned with IELTS writing and descriptors. Quantitative data were collected through pre- and post-feedback essay scores across four assessment criteria: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy, rated by certified IELTS examiners. Qualitative data were gathered through open-ended questionnaire responses. Findings revealed significant improvements across all four writing criteria during a 4-week intervention period, particularly in grammatical range and lexical resources. Learners reported high levels of satisfaction with the AI feedback, citing clarity, immediacy, and a non-judgmental tone as motivating factors. Thematic analysis also indicated increased learner autonomy and deeper engagement with revision practices. These results suggest that AI-generated feedback can be an effective pedagogical tool in high-stakes test-preparation contexts. Implications for writing instruction, learner training, and the integration of AI in the EFL classroom are discussed.

Keywords: AI-generated feedback; automated writing evaluation; IELTS Writing Task 2; L2 writing

INTRODUCTION

English language proficiency remains a critical asset for academic access, international employment, and global mobility. Among the various standardized assessments used to measure English competence, the International English Language Testing System (IELTS) remains one of the most widely recognized and trusted frameworks worldwide. Accepted by over 11,000 institutions in more than 140 countries, the IELTS test evaluates four core language skills—listening, reading, speaking, and writing—and is frequently used as a gateway requirement for higher education and immigration (IELTS.org, 2023). Within the IELTS assessment, Writing Task 2 presents significant cognitive and linguistic challenges. This task requires test-takers to complete the exercise within 40 minutes. The task assesses the candidate's ability to formulate a position, develop ideas with support, organize arguments coherently, and employ a range of vocabulary and grammatical structures. These demands are operationalized through four band descriptors: Task Response,

*Corresponding author
Email: sisila@unimed.ac.id

Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy.

Despite the relative familiarity of this task type, research consistently shows that EFL learners struggle to meet these standards. Studies have reported frequent weaknesses in developing well-structured arguments, linking ideas cohesively, and achieving lexical variety and grammatical precision (Banerjee et al., 2007; Nguyen & Le, 2022; Todd et al., 2007). These challenges are further amplified by the pressure of high-stakes testing and the limited opportunities many learners have to receive individualized, criterion-aligned feedback during preparation.

Feedback plays a critical role in the development of L2 writing proficiency. It provides learners with information that bridges the gap between their current performance and expected outcomes, guiding both surface-level corrections and higher-order improvements such as content development, coherence, and rhetorical organization (Biber et al., 2011; Ferris, 2004). In the context of IELTS preparation, where writing success is determined by strict, rubric-based criteria, timely and specific feedback can serve as a powerful pedagogical intervention. Traditionally, input in writing instruction has been delivered by teachers or peers. Teacher feedback, while potentially accurate and detailed, often suffers from limitations in availability, turnaround time, and consistency—especially in large EFL classrooms (Bijami et al., 2013; Man et al, 2018). Peer feedback may foster collaborative learning and self-reflection, but it often lacks the reliability and rubric alignment necessary for high-stakes assessment preparation. As a result, learners may receive feedback that is either too general, delayed, or misaligned with test-specific writing standards.

Moreover, the effectiveness of feedback is often mediated by learners' ability to interpret and apply it. Learners may not always understand teacher comments or may lack the metacognitive strategies needed to revise their work accordingly. These challenges highlight the need for feedback that is both accessible and, especially in preparation for standardized writing tasks like IELTS Writing Task 2. Recent research has emphasized the value of iterative feedback, particularly when it enables students to revise their writing in cycles, addressing feedback over time rather than as a one-off intervention (Fleckenstein et al., 2023; Mohammadi et al., 2023). In such models, feedback is integrated into a process-oriented pedagogy that encourages continuous learning and deeper engagement with writing tasks. However, implementing such a model in typical classroom settings can be logistically demanding, especially when feedback must be personalized and rubric-specific.

These pedagogical pressures have accelerated interest in AI-powered tools that can deliver feedback at scale—tools that are not only immediate and consistent but also align with the demands of standardized writing rubrics. Advancements in artificial intelligence (AI) and natural language processing (NLP) have enabled automated systems to provide immediate, criterion-based feedback on learner writing. These tools—including commercial platforms like Grammarly and Write & Improve, as well as more recent generative models like ChatGPT—offer corrections, suggestions, and explanations on a range of writing aspects such as grammar, coherence, lexical diversity, and tone (Wei et al., 2023; Shi et al., 2025). AI-generated feedback offers several advantages over traditional approaches. Its primary strengths include speed, consistency, and scalability. Unlike teacher feedback, which may take hours or days to return and may vary across instructors, AI systems can deliver feedback within seconds and apply rubric-based evaluation rules uniformly. Navas Bonilla et al. (2025) suggest that the immediacy of AI feedback creates opportunities for iterative revision and sustained engagement, supporting self-directed learning processes that may contribute to gradual writing development over time.

Emerging research supports the pedagogical utility of AI-generated feedback in EFL and academic writing contexts. For instance, Kohnke et al. (2023) reported that learners perceived ChatGPT's feedback as clearer, more transparent, and more actionable than teacher commentary, especially for surface-level features such as grammatical accuracy and word choice. Moreover, Fleckenstein et al. (2023) found that combining teacher and AI-generated feedback led to higher-quality revisions, suggesting that these approaches can complement rather than replace one another.

In addition to performance gains, studies have documented positive learner perceptions of AI feedback. Students report feeling more confident, less anxious, and more in control of their writing process when interacting with non-judgmental, responsive systems (Aljasser, 2025). This affective benefit may be especially relevant in high-stakes contexts like IELTS, where anxiety can negatively impact performance and motivation. However, concerns remain about the pedagogical limitations of AI-generated feedback. Critics argue that current systems often lack contextual awareness, thereby making them less effective at guiding learners in higher-order skills such as argument development and critical thinking (Zhang, 2025). Furthermore, few studies have explored how these systems function specifically in standardized test preparation, where scoring rubrics and writing formats are tightly prescribed (Uludag & McDonough, 2022).

Despite growing interest in the pedagogical potential of AI-generated feedback, existing research remains limited in several critical respects

of the IELTS. While AI feedback systems have demonstrated effectiveness in general academic and classroom writing settings, relatively few studies have investigated whether such tools can align meaningfully with standardized rubrics such as those used in IELTS Writing Task 2. The feedback generated by platforms like ChatGPT or Grammarly is typically not calibrated to address the four band descriptors that determine IELTS scores: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy (IELTS.org, 2023).

Furthermore, most existing studies have focused on either learner perceptions or single-draft writing outcomes. There is a lack of research examining how learners engage with AI-generated feedback over multiple writing iterations and how this engagement affects their performance across distinct writing criteria. Revision, a core aspect of writing development, is often treated as a one-time event rather than a process. Understanding how learners respond to feedback repeatedly—across tasks, over time—is essential for evaluating the impact of AI feedback systems on writing performance.

While AI-generated feedback shows potential for L2 writing development, notable gaps remain in its application to standardized test preparation, such as the IELTS Writing Task 2. Existing research often focuses on general academic writing without addressing the specific demands of standardized rubrics (Kohnke et al., 2023). Additionally, while studies demonstrate improvements in writing from AI assistance (Han et al., 2024), they typically examine single-draft outcomes rather than the iterative, multi-draft process essential for observing revision behavior and autonomy. Research that combines in high-stakes contexts is still emerging.

Considering previous research, this study aims to investigate the impact of AI-generated feedback across multiple IELTS drafts, measuring improvement against the four band descriptors and exploring learners' perceptions of the feedback's clarity and instructional value. By triangulating this data, the research contributes to understanding how AI tools can be pedagogically integrated into criteria-driven test preparation to support targeted, autonomous learning in EFL settings (Fleckenstein et al., 2023; Mohammadi et al., 2023). The overarching objective of this study is to evaluate the role of AI-generated feedback in enhancing both the performance and perceptions of EFL learners engaging in IELTS Writing Task 2 preparation. Grounded in process-oriented writing pedagogy and informed by prior research on automated feedback systems, the study examines how learners interact with immediate, rubric-aligned feedback over multiple writing iterations. The goal is not only to assess whether AI feedback contributes to measurable writing improvement, but also to

explore how learners perceive its effectiveness, usability, and instructional value. To achieve this objective, the study is guided by the following research questions:

1. To what extent does AI-generated feedback enhance learners' performance in IELTS Writing Task 2 across the four band descriptors: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy?

2. How do learners perceive the usability, clarity, and instructional value of AI-generated feedback in the context of IELTS writing preparation?

By addressing these questions through a convergent mixed-methods approach, the study contributes to the growing field of AI-mediated writing instruction. It provides practical insights for educators, curriculum designers, and assessment stakeholders. The findings are expected to inform best practices for integrating AI feedback systems into standardized writing preparation programs, especially in EFL contexts where access to personalized feedback is often limited.

IELTS Writing Task 2 and Learner Challenges

IELTS Writing Task 2 is widely regarded as one of the most cognitively and linguistically demanding components of the IELTS examination. The task requires test-takers to write a formal essay of at least 250 words in response to a prompt that normally asks them to discuss a point of view, argument, or problem. Within a strict 40-minute timeframe, candidates are expected to formulate a coherent response, present and support their ideas logically, and demonstrate lexical and grammatical competence (IELTS.org, 2023). Writing Task 2 is assessed based on four band descriptors: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. Each criterion plays a distinct role in evaluating the quality and effectiveness of the writing. Task Response focuses on the relevance and development of ideas; Coherence and Cohesion address the logical organization and flow of the text; Lexical Resource assesses vocabulary use and variation; and Grammatical Range and Accuracy evaluate syntactic complexity and correctness.

Despite its structured rubric, many EFL learners continue to struggle to meet the expectations of this task. Studies have found that candidates often struggle with underdeveloped arguments, repetition of ideas, weak paragraphing, and limited lexical variety (Banerjee et al., 2007; Todd et al., 2007). For instance, Nguyen & Le (2022) reported that Iranian EFL learners frequently produced essays that lacked clarity of stance and supporting evidence, as well as Task Response and Coherence criteria.

Time pressure further compounds these challenges. Barkaoui (2010) highlighted that test-related anxiety often impairs the fluency and accuracy of L2 writers, particularly in standardized assessments. In such high-stakes contexts, learners are less likely to monitor their output effectively or revise their ideas in depth. This results in writing that is not only error-prone but also structurally and thematically underdeveloped. These findings indicate that adequate preparation for IELTS Writing Task 2 must address not only the linguistic dimensions of writing but also the strategic and psychological challenges learners face. Central to this preparation is the provision of targeted, criterion-aligned feedback that enables learners to understand their weaknesses and make meaningful improvements in each rubric category.

Feedback on L2 Writing Instruction
Feedback has long been recognized as a cornerstone of effective second language (L2) writing instruction. Defined as information provided to learners to guide revision and improvement, feedback supports the development of both linguistic accuracy and discourse-level competence (Ferris, 2004; Hyland & Hyland, 2006). In formative contexts, feedback can clarify expectations, raise awareness of recurrent errors, and foster metacognitive engagement with writing processes (Biber et al., 2011).

In test-preparation settings like IELTS, feedback that directly aligns with the rubric descriptors is particularly valuable. When learners receive criterion-referenced feedback—e.g., specific to "Task Response" or "Coherence and Cohesion"—they are more likely to focus revisions on the aspects that directly influence their scores. However, such feedback is often constrained by classroom realities. Teacher feedback, while potentially rich and diagnostic, is labor-intensive, inconsistent, and sometimes delayed due to workload and time limitations (Bijami et al., 2013; Man et al, 2018). While peer feedback and peer assessment offer substantial formative benefits, concerns remain regarding reliability, authority, and rubric alignment when peer grading is used in high-stakes summative contexts (Nie, 2022).

Moreover, research has shown that feedback effectiveness depends on how learners interpret and apply it. Learners often misunderstand vague or general comments and may struggle to translate them into concrete revision strategies. This is particularly problematic for lower-proficiency learners or those unfamiliar with rubric-based assessment. Therefore, feedback must be both clear and actionable, and ideally, allow for multiple cycles of revision to support long-term improvement (Ferris, 2004; Man et al, 2018).

Recent studies emphasize the importance of iterative feedback loops in L2 writing pedagogy.

For example, Fleckenstein et al. (2023) found that learners who received and applied feedback across multiple drafts had greater sustained improvement than those who revised only once. Similarly, Mohammadi, Zarrabi, and Kamali (2023) demonstrated that combining human and automated feedback led to statistically significant gains in writing quality among IELTS candidates in Iran, particularly in grammatical range and coherence.

Despite its clear pedagogical value, such iterative, criterion-based feedback remains underutilized in many IELTS preparation programs—primarily due to resource limitations and the pressure to cover test strategies rather than writing development. These constraints have led educators and researchers to explore technology-enhanced solutions, particularly those driven by artificial intelligence, to deliver scalable and rubric-aligned feedback more efficiently.

AI-Generated Feedback in L2 Writing
The integration of artificial intelligence (AI) into writing instruction has introduced new possibilities for delivering immediate, individualized, and consistent feedback. Leveraging natural language processing (NLP) and machine learning technologies, AI-based feedback systems such as Grammarly, Write & Improve, and ChatGPT can evaluate learner writing in real time, identify surface-level and structural issues, and provide suggestions for improvement (Wei et al., 2023; Shi et al., 2025). AI-generated feedback is particularly valuable in contexts where teacher feedback is constrained by time or scalability. One of its core strengths lies in its speed and availability—feedback is delivered within seconds, enabling learners to revise their work multiple times without waiting for instructor input.

Empirical research supports the utility of AI-based feedback in improving L2 writing performance. For instance, Kohnke et al. (2023) reported that undergraduate EFL learners perceived ChatGPT-generated feedback as more readable, specific, and usable than feedback provided by instructors. Similarly, Han et al. (2023) documented substantial vocabulary gains and sentence-level improvements among students who used ChatGPT across multiple essay revisions. Importantly, AI feedback tools also support self-directed learning by allowing students to experiment with revisions in a low-risk environment. Navas Bonilla et al. (2025) note that real-time, personalized feedback allows learners to revise and refine their work autonomously, fostering iterative experimentation without immediate evaluative pressure. This affective benefit is critical in high-stakes settings, such as IELTS preparation, where fear of negative evaluation can inhibit performance and learning.

Nevertheless, concerns have been raised about the limitations of current AI systems, especially in providing feedback on higher-order writing skills such as critical thinking, argumentative development, and audience awareness (Zhang, 2025). Because AI models rely on pattern recognition rather than proper discourse comprehension, they may miss nuanced issues or provide misleading suggestions in complex writing tasks. These limitations suggest that while AI-generated feedback offers strong potential in addressing lexical, grammatical, and coherence-related challenges, its use in preparing learners for high-stakes, rubric-governed assessments like IELTS requires careful alignment and pedagogical integration.

Learner Perceptions of AI Feedback Tools

Beyond performance metrics, learner perception plays a pivotal role in determining the effectiveness of AI-generated feedback in writing instruction. How learners interpret, trust, and act upon feedback directly influences their engagement with revision and their long-term writing development (Han et al., 2023; Aljasser, 2025). Consequently, understanding learners' attitudes toward AI feedback is essential for evaluating its pedagogical impact and optimizing its integration into the classroom.

Several studies have highlighted learners' positive attitudes toward AI feedback tools. A common theme is the appreciation for the immediacy and clarity of the feedback. In a mixed-methods study, Kohnke et al. (2023) found that students using ChatGPT were more engaged in the revision process because they could easily understand the suggestions and act on them without delay. Similarly, Aljasser (2025) reported that students viewed AI feedback as more accessible and less intimidating than teacher feedback, especially when revising multiple drafts.

Learners have also expressed appreciation for the non-judgmental tone of AI systems, which creates a psychologically safe environment for trial and error. Fleckenstein et al. (2023) noted that students felt more confident revising when feedback came from an AI system rather than a teacher, particularly in contexts where learners feared embarrassment or negative evaluation.

However, learner perceptions are not uniformly positive. Some students have raised concerns about the overgeneralization or inaccuracy of certain AI-generated suggestions. For example, Sumakul et al. (2021) found that while learners appreciated AI support for grammar and vocabulary development, they also reported concerns about incoherent, unnatural, or unwanted content generated by the tool, indicating reservations about AI recommendations that could affect meaning and organization. Similarly, Han et al. (2023) cautioned that learners may accept AI feedback uncritically,

even when it lacks contextual appropriateness or coherence with the writer's communicative purpose. In light of these findings, learner perception of AI feedback appears to be shaped by both technical performance (i.e., accuracy, clarity) and affective factors (i.e., confidence, autonomy). These perceptions, in turn, influence how feedback is used and whether it leads to meaningful improvement. Therefore, research examining learners' experiences alongside performance outcomes can offer a more comprehensive understanding of how AI feedback contributes to writing development in test-preparation settings, such as the IELTS.

METHOD

Research Design

This study employed a convergent mixed-methods design (Creswell & Plano Clark, 2018) to examine both the quantitative progression of learners' performance in IELTS Writing Task 2 and their qualitative perceptions of AI-generated feedback. This design allowed the researchers to collect and analyze both numerical data (essay scores across four IELTS band descriptors) and textual data (learner reflections via questionnaires) in parallel, followed by a merging of findings for interpretation. On the other hand, quantitative data were derived from students' writing scores over four weekly tasks, evaluated according to the IELTS Writing Task 2 rubric. These scores were analyzed to track changes in Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GRA). Then, qualitative data were obtained from open-ended responses in a post-intervention questionnaire that invited students to reflect on the clarity, usefulness, and motivational impact of the AI-generated feedback.

Participants and Data Collection

The participants were 18 undergraduate students in the English Education Department at a public university in Indonesia. They were selected through purposive sampling based on the criteria that the participants were: (1) in preparation for the IELTS examination as part of their academic coursework, (2) in the intermediate to upper-intermediate English proficiency levels, (3) verified to have institutional CEFR level of B1 to C1, based on prior placement tests, and (4) documented to have Documented IELTS Writing Task 2 practice scores between bands 5.0 and 6.5 within the past three months. These criteria were set to ensure a reasonably homogenous sample in terms of writing proficiency. Also, the sampling approach allowed the study to focus on a well-defined learner group—IELTS-oriented EFL students in a higher education setting—making the findings particularly relevant to similar instructional contexts. During data analysis, participants were kept anonymous by assigning

codes (e.g., P1, P2, through to P18), which were used for both score tracking and qualitative reporting.

Data was collected over a four-week intervention period, during which participants completed a series of four IELTS Writing Task 2 assignments. The types of essays in the intervention were opinion essay, discussion essay, problem-solution essay, and two-part question essay. See Table 1.

Table 1
Writing Sessions

Week	Task Type	Prompt Theme	Writing Phase	Feedback Cycle
Week 1	Opinion Essay	Technology and education	250-word essay (40 min)	AI feedback on TR, CC, LR, GRA → learner revision → final submission
Week 2	Discussion Essay	Government and responsibility	250-word essay (40 min)	AI feedback cycle similar to Week 1
Week 3	Problem-Solution Essay	Environmental sustainability	250-word essay (40 min)	AI feedback cycle similar to Week 1
Week 4	Two-Part Question Essay	Work–life balance	250-word essay (40 min)	AI feedback cycle similar to Week 1

Each task followed a structured cycle comprising three key stages:
1. Initial Essay Submission: Participants wrote a timed, 250-word IELTS Task 2 essay based on a standardized prompt.
2. AI Feedback Delivery: Essays were uploaded to the AI feedback tool, which returned criterion-aligned feedback on TR, CC, LR, and GRA.
3. Essay Revision and Resubmission: Learners reviewed the feedback, revised their drafts, and submitted final versions for evaluation.

Writing prompts were selected from publicly available, authentic IELTS preparation materials, ensuring relevance and difficulty comparable to the actual exam. Students were given 40 minutes per task to simulate real test conditions, and all submissions were done digitally through the custom AI interface. Each weekly session was supervised by the researcher and took place in a computer lab to ensure uniform access to the AI tool and minimize distractions. Participants received instructions on interpreting and applying the AI feedback, but no direct assistance with writing or revising was provided. At the end of the four-week cycle, participants completed a feedback perception questionnaire to reflect on their experiences with the AI tool. This two-phase data collection process—writing iterations followed by a perception instrument—enabled the researchers to capture both objective changes in writing performance and subjective learner insights.

The AI feedback used in this study system was developed using the OpenAI GPT-4 API. It was designed to provide consistent, rubric-aligned feedback throughout the study. The system was embedded in a web-based interface designed specifically for IELTS Writing Task 2, allowing learners to submit their essays and receive immediate feedback based on the four IELTS band descriptors: (1) Task Response (TR), (2) Coherence and Cohesion (CC), (3) Lexical Resource (LR) and (4) Grammatical Range and Accuracy (GRA). The system functioned through structured prompt engineering to ensure alignment with assessment criteria. Each essay was processed using a series of GPT-4 prompts pre-configured to analyze the writing according to IELTS standards. For example, a prompt might instruct the model to: *"Evaluate this essay's coherence and cohesion based on IELTS criteria. Provide strengths, weaknesses, and at least two suggestions for improvement."*

Feedback for each criterion was presented in separate blocks. Suggestions were phrased in clear, accessible English and focused on both diagnosis (what the issue is) and actionability (how to fix it). Learners could view color-coded highlights identifying issues in their submitted text, accompanied by model suggestions or rephrased examples for improvement. The tool was also designed to support multiple iterations. After receiving feedback on their initial draft, learners were encouraged to revise and resubmit their essays for a second round of AI-generated evaluation, simulating an authentic revision process guided by formative input. To maintain fairness and consistency, the same version of the GPT-4 model and identical prompt templates were used across all participants and tasks. Logs of both original and revised texts, as well as the feedback generated, were archived for analysis. By controlling both the content and structure of AI-generated feedback, the tool ensured a standardized experience across participants, allowing for valid comparisons of writing performance and revision outcomes over time.

Data Analysis Procedures
The primary quantitative data consisted of learners' writing scores across four IELTS criteria (TR, CC, LR, GRA), collected over four writing tasks (initial and revised drafts). Descriptive statistics—including means, standard deviations, and score ranges—were

calculated to provide an overview of learner performance trends. To assess whether improvements across writing tasks were statistically significant, the study used the Wilcoxon signed-rank test, a nonparametric alternative to the paired-samples t-test, suitable for small sample sizes and ordinal data such as IELTS band scores. Tests were conducted separately for each of the four criteria, comparing scores from Task 1 (Week 1) and Task 4 (Week 4) to determine the magnitude and direction of change. Statistical analysis was conducted using IBM SPSS Statistics (Version 27), with a significance threshold set at $p < .05$.

In the next process, qualitative analysis was conducted using thematic analysis of open-ended questionnaire responses, following the six-step model proposed by Braun and Clarke (2006): data familiarization, initial coding, theme categorisation, theme review, theme definition, and reporting. Data were coded inductively, allowing themes to emerge from the content rather than being imposed a priori. Codes were then grouped into broader themes that captured learner experiences, attitudes, and perceived outcomes. To enhance trustworthiness, coding was reviewed independently by two members of the research team and discussed until consensus was reached. By integrating quantitative progression data with qualitative learner reflections, the analysis aimed to provide a comprehensive understanding of how AI-generated feedback influenced both writing performance and learner engagement over time.

FINDINGS AND DISCUSSION

The results of the study converge into two broad sections. First, the effectiveness of AI-generated feedback was shown across the categories of Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GRA). Then, the quantitative results led to a discussion of learners' perceptions of the usability, clarity, and instructional value of AI-generated feedback in the context of IELTS writing preparation.

Effectiveness of AI-generated feedback

The overall quantitative analyses show the extent to which AI-generated feedback contributed to improvements in learners' IELTS writing performance across multiple tasks (see Table 2). To some extent, this addresses Kohnke et al.'s (2023) concerns about the current focus of overly general rubrics for academic writing assessments. This study shows that effectiveness is operationalized through changes in rater-assigned band scores following iterative cycles of feedback and revision. By tracking score development across four writing tasks, this analysis provides empirical insight into how AI-supported feedback serves as formative support in second-language writing. In the process, each participant completed four writing tasks, each followed by AI-generated feedback and a subsequent revision. The final draft of each weekly task was scored by two certified IELTS raters using the official rubric. Score trends across tasks are reported both descriptively and inferentially. Across 18 participants, writing scores improved consistently from Task 1 to Task 4.

All four criteria exhibited upward trends, with the most pronounced improvements observed in Lexical Resource and Grammatical Range and Accuracy—areas that AI feedback explicitly targeted through vocabulary suggestions, grammar corrections, and sentence-level restructuring. To evaluate whether the observed improvements in writing performance across the four IELTS band descriptors were statistically significant, a Wilcoxon signed-rank test was conducted. This nonparametric test, shown in Table 3, compared the median scores from Task 1 and Task 4 (first and final writing cycles) for each criterion. The test was chosen due to the ordinal nature of IELTS band scores and the modest sample size (N = 18).

Table 2

Overview of the Mean Band Scores by Criterion Across the Four Writing Tasks

Criterion	Task 1	Task 2	Task 3	Task 4
Task Response (TR)	5.39	5.61	5.78	6.06
Coherence & Cohesion (CC)	5.33	5.50	5.72	6.00
Lexical Resource (LR)	5.22	5.44	5.72	6.06
Grammar (GRA)	5.17	5.39	5.67	6.00

Table 3

Overview of Statistical Significance of Score Gains

Criterion	Z-score	p-value	Interpretation
Task Response (TR)	-3.32	0.001**	Significant improvement
Coherence & Cohesion (CC)	-3.05	0.002**	Significant improvement
Lexical Resource (LR)	-3.41	0.001**	Significant improvement
Grammar (GRA)	-3.12	0.002**	Significant improvement

Note: $p < 0.05$ = significant; $p < 0.01$ = highly significant

All four writing criteria showed statistically significant improvements over the course of the intervention. The strongest gains were observed in Lexical Resource ($Z = -3.41$, $p = 0.001$), suggesting that the vocabulary enhancement features of the AI feedback system were particularly effective. Similarly, improvements in Grammatical Range and Accuracy were both statistically and pedagogically meaningful. These results support the conclusion that iterative, criterion-based AI feedback, when used across multiple writing cycles, can contribute to measurable gains in learners' IELTS writing performance. In the following section, each category is elaborated further.

Task Response: Performance Patterns and Learner Reflections

While overall score gains indicate the effectiveness of AI-generated feedback across writing criteria, closer examination is needed to understand how these improvements manifest in specific performance dimensions. While Hyland (2000) acknowledged the benefits of peer feedback in promoting reflective learning, this study found that, with AI assistance, individual students can develop reflective learning. For example, in terms of Task Response, score patterns were combined with learners' reflective accounts to illustrate how AI feedback shaped task interpretation, idea development, and alignment with prompt requirements. In the Task Response (TR) criterion, learners demonstrated progressive improvement in their ability to fully address prompts, develop ideas logically, and maintain relevance throughout the essay. The mean band score increased from 5.39 (Task 1) to 6.06 (Task 4), representing a notable shift from "limited" to "generally effective" argumentation, according to IELTS rubric benchmarks. Initial submissions often included incomplete thesis statements, superficial topic sentence development, or underdeveloped conclusions. By the third and fourth tasks, however, most learners were constructing clearer positions, elaborating with appropriate support, and staying more consistently on topic.

Open-ended questionnaire responses revealed that AI-generated feedback was instrumental in helping learners improve task relevance and idea development. Participants cited feedback such as "You need to directly answer the question in your introduction" or "Include specific examples to support this point" as especially helpful. In addition, a participant noted, "I used to write off-topic introductions. The AI told me exactly what part was confusing and gave me a better example. That helped me fix it in my next draft" (P07). Then, another participant explained that "The system always reminded me to link my body paragraph to the main argument. It made me more aware of staying focused" (P12).

These comments highlight the awareness-building function of AI feedback. Many participants reported being unaware that they had misinterpreted or incompletely addressed the task prompt until the system. Learners also appreciated that feedback for TR was not only corrective but also directive, offering suggestions for improving future responses. Overall, the improvement in TR can be attributed to the system's ability to pinpoint misalignment with prompt expectations, suggest clearer thesis formulations, and prompt elaboration of underdeveloped ideas.

Coherence and Cohesion: Score Trends and Feedback Impact

While improvements in Task Response reflect learners' increasing ability to address prompts and develop relevant ideas, effective communication also depends on how those ideas are organized and connected across the text. As noted by Biber et al. (2011), feedback is important in providing learners with information on how to improve content development, coherence, and rhetorical organization in written work. This study is relevant because it shows increasing trends in coherence and cohesion, specifically in how AI-generated feedback influenced learners' use of logical sequencing, paragraphing, and cohesive devices. The criterion of Coherence and Cohesion (CC) saw a consistent upward trajectory across the four tasks. Mean scores increased from 5.33 (Task 1) to 6.00 (Task 4). Early drafts were often characterized by unclear paragraph structure, overuse of basic linkers (e.g., *and*, *but*, *because*), and weak logical progression between ideas. By Task 4, most learners had begun incorporating appropriate paragraphing, clearer sequencing, and a wider range of cohesive devices.

Student reflections suggest that AI feedback played a formative role in helping them structure their essays more logically. Several learners noted that before the intervention, they had little awareness of paragraph-level coherence. For example, a participant noted that "Before, I just wrote everything in one flow. The AI showed me where to separate my ideas and how to connect them better" (P03). Hence, this brought about a reflection for another participant who said, "It told me not to use 'firstly–secondly–finally' in every essay. It gave alternatives like 'one contributing factor' or 'a contrasting view,' which made my writing feel more natural" (P15).

In addition to cohesion markers, students were prompted to add transitional phrases that clarified relationships between ideas. The AI also provided suggestions on topic-sentence formulation and on how to end paragraphs more purposefully, reinforcing the flow of ideas. So, the improvement in CC appears to result from both explicit instruction (e.g., "Use a clearer transition to

introduce this paragraph") and subtle modeling (e.g., AI-generated rephrased examples showing improved cohesion). Learners actively engaged with these features, leading to structurally more coherent essays by the end of the final writing cycle.

Lexical Resource: Progress and AI Vocabulary Support

Among the four assessment criteria, Lexical Resource (LR) showed the most substantial improvement. The average band score rose from 5.22 (Task 1) to 6.06 (Task 4). Early writing samples typically featured limited vocabulary, frequent repetition of common words, and inappropriate word choices. As the intervention progressed, learners demonstrated increasing control over collocation, register, and word variation, meeting the IELTS rubric expectations for lexical range and flexibility.

Student reflections clearly linked vocabulary development to the AI feedback system's rephrasing suggestions and synonym prompts. Many participants reported that the tool encouraged them to move beyond their habitual word choices and experiment with more precise or academic alternatives. A participant reported that "The AI suggested I change 'big problem' to 'pressing issue.' It sounded better and more formal" (P02) and "It told me I used 'important' five times. Then it gave me options like 'significant,' 'crucial,' or 'essential.' I didn't know about those before" (P10).

Additionally, learners appreciated that the vocabulary suggestions were contextual rather than generic. The AI explained why a certain word was inappropriate or repetitive and offered suitable substitutions. This form of just-in-time lexical support appeared not only to enhance individual essays but also to increase students' lexical awareness over time.

A few participants also mentioned keeping personal glossaries of new words encountered through the system, indicating the emergence of independent vocabulary-learning behaviors. This aligns with literature on AI feedback fostering autonomy and strategic vocabulary use (Aljasser, 2025). Therefore, the increase in LR scores appears to reflect both expanded lexical range and improved accuracy of word choice, outcomes directly supported by the AI system's tailored feedback.

Grammatical Range and Accuracy: Development Across Cycles

The criterion of Grammatical Range and Accuracy (GRA) demonstrated clear and steady improvement throughout the four writing tasks. The mean band score increased from 5.17 (Task 1) to 6.00 (Task 4). Initial drafts frequently included tense inconsistencies, run-on sentences, article errors, and limited sentence structure variety. By the end of the intervention, participants were producing writing that featured greater syntactic diversity and fewer grammatical errors, aligning more closely with the descriptors for Band 6 in IELTS writing.

AI-generated feedback was particularly effective in identifying common grammatical issues and offering corrective suggestions with explanations. Students consistently cited grammar-related feedback as the most immediate and actionable component of their learning experience. About this issue, a participant commented that "The system always caught when I used the present tense in the past. It didn't just say 'wrong,' it told me why" (P05). Another participant noted that "I was afraid to try complex sentences before. After the AI showed me how to combine ideas, I started using more clauses" (P14).

Learners reported that the color-coded highlights and rewritten model sentences provided by the AI helped them understand the nature of their errors and visualize correct structures. In several cases, students remarked that they began consciously reviewing sentence variety and checking for fragments before submission—behaviors indicating growing grammatical self-monitoring. Additionally, many participants found the feedback motivating because it did not feel punitive. Unlike traditional red-pen correction, the AI's tone was described as "neutral," "helpful," and "less stressful," reducing the emotional resistance some learners associate with grammar correction. The combined impact of corrective clarity, error pattern identification, and supportive tone appears to have facilitated tangible improvement in both the range and accuracy of sentence construction.

Overall, this study's findings demonstrate that iterative, AI-generated feedback significantly contributed to learner improvement across all four IELTS Writing Task 2 band descriptors. These improvements were not only statistically significant but also pedagogically meaningful, aligning closely with the instructional objectives of IELTS preparation programs. The increase in Task Response scores suggests that learners developed a better understanding of how to interpret prompts and structure argumentative responses. This aligns with Ferris (2004) and Hyland and Hyland (2006), who emphasize that specific, targeted feedback enhances content relevance and idea development. In this study, AI-generated prompts that highlighted vague thesis statements or underdeveloped body paragraphs directly helped learners address the prompt more effectively in subsequent drafts.

The improvement in Coherence and Cohesion reflects the AI tool's ability to scaffold logical sequencing and paragraph unity. Learners were not only exposed to lexical alternatives for transitions but were also encouraged to improve macro-organizational skills, a skill that can be difficult to teach in classroom settings due to time constraints.

This supports the findings of Biber et al. (2011), who advocate feedback that fosters paragraph-level cohesion through both instruction and modeling.

Lexical gains were particularly strong, suggesting that the system's vocabulary enhancement functions (e.g., synonym suggestions and rephrased sentences) effectively addressed one of the most common weaknesses among EFL IELTS candidates: repetitive and simplistic word choice (Todd et al., 2007). Unlike teacher feedback, which may highlight lexical errors, the AI provided context-aware alternatives—a strength echoed in a study by Fleckenstein et al. (2023).

The improvement in grammatical control further confirms that AI feedback systems can serve as powerful tools for syntax awareness and self-correction. Learners responded positively to direct, non-punitive grammar feedback, aligning with Navas Bonilla et al. (2025) and Shi et al. (2025), who found that AI feedback promotes confidence in experimenting with more complex structures. The tool's consistent identification of tense errors, fragment issues, and overreliance on simple sentence forms led to greater grammatical diversity and accuracy over time.

Learner Perception of AI-Generated Feedback
Thematic analysis of responses to three open-ended questions in the post-intervention questionnaire revealed four dominant themes in learners' perceptions of AI-generated feedback: clarity and specificity of feedback, motivation and confidence, revision awareness and strategy, and perceived limitations.

In terms of clarity and feedback, students overwhelmingly praised the directness and structured presentation of the AI-generated feedback. For example, a participant stated, "It gave me feedback in four parts—just like the IELTS. That helped me focus" (P09). In addition, a participant wrote, "The system told me *exactly* what was wrong and how to fix it. It felt more organized than teacher comments" (P06). This means the feedback aligned with each IELTS criterion, helping them internalize what each band descriptor meant in practice.

Participants in the study reported a boost in motivation and confidence due to the AI's tone and immediacy. Comments are found to have positive tones such as "It didn't judge me. Just helped me improve. That made me more willing to revise" (P08) and "I was more confident to try again because I didn't have to wait for a teacher" (P01). These responses suggest that AI feedback can create a low-anxiety environment, which is particularly beneficial in high-stakes exam preparation contexts.

Learners noted that they became more strategic and deliberate in their revisions. A participant admitted, "I used to just correct grammar. Now I look at structure, logic, and vocabulary too" (P13).

Meanwhile, several other students mentioned they began keeping checklists or vocab journals based on repeated feedback patterns, signaling increased self-regulated learning behavior. On the other hand, the data analysis also shows that a minority of students noted limitations in the feedback system, especially in higher-order skills. For example, they commented, "Sometimes it didn't understand my meaning. It gave suggestions that were too simple" (P11) or "Good for grammar and vocab, but not always helpful with argument strength" (P17). These remarks indicate that while the AI tool was well received, it was not perceived as a full substitute for teacher guidance—particularly in areas such as idea logic and nuance. Overall, learner perceptions highlight the system's accessibility, emotional safety, and strategic usefulness, while also identifying its limitations, namely its limited ability to interpret semantic depth or to engage in critical reasoning in argumentative writing.

One of the most compelling findings of this study concerns the learner-reported behavioral shifts in how participants approached the revision process. Rather than treating feedback as a one-time correction mechanism, students began to engage in strategic, multi-layered revisions—attending not only to grammar and vocabulary but also to structure, argumentation, and clarity. This aligns with contemporary views on process-oriented writing pedagogy, in which revision is conceptualized as an iterative, reflective act (Hyland & Hyland, 2019; Ferris & Hedgcock, 2014).

AI-generated feedback functioned as a catalyst for learner autonomy and metacognitive growth. Participants frequently reported that the clarity and accessibility of the feedback empowered them to revise independently and with greater purpose, reflecting the transition from reactive revision to deliberate rewriting described by Aljasser (2025). Many also began to internalize the feedback language, mentally archiving and reusing recurrent prompts in future drafts. Such metacognitive awareness is indicative of self-regulated writing competence (Wenden, 1991). In line with Navas Bonilla et al. (2025), the immediacy and personalized nature of AI feedback may help reduce affective barriers by creating a more supportive, less evaluative learning environment. Learners reported feeling more comfortable engaging with AI-mediated feedback, which in turn encouraged experimentation and autonomous revision.

Taken together, these observations suggest that the AI feedback system served not only as a corrective tool, but also as a pedagogical agent that enhanced learner agency, ownership of the writing process, and reflective practice—outcomes that are essential for sustainable writing development. While the AI feedback system yielded measurable improvements across all IELTS scoring criteria, participant responses also highlighted some

limitations, particularly in areas requiring higher-order cognitive skills, such as critical reasoning, logical argumentation, and nuance in ideas. These concerns echo those of Zhang (2025), who caution that AI systems, although effective at surface-level revisions, often struggle with global discourse functions.

A few participants in the study noted that the AI was less helpful when evaluating the strength of their arguments, identifying bias, or offering suggestions for deeper idea development. For instance, a participant remarked that "the feedback didn't know what I was really trying to say" (P11), indicating a lack of semantic sensitivity in the AI model's interpretation of the essay's rhetorical intent. Similarly, another participant observed that feedback was "too focused on grammar" and "not very helpful when I needed to improve the logic of my ideas" (P17).

These limitations may affect pattern recognition and syntactic analysis, but they lack a true understanding of pragmatic intent and argumentative depth. Although the tool was prompted to assess alignment with IELTS descriptors, it did not consistently offer feedback on idea originality, logical fallacies, or persuasive strength—skills that are essential for high-band IELTS responses (IELTS.org, 2023). It further implies that while AI tools can serve as effective revision scaffolds, they may need to be supplemented with teacher-led instruction or guided peer review—particularly for tasks that involve critical thinking and extended reasoning. Teachers may, for example, use AI feedback as a first layer of revision support, followed by targeted conferencing to address argument structure and content quality.

CONCLUSION

This study examined the impact of AI-generated feedback on the writing performance and revision behavior of EFL learners preparing for the IELTS Writing Task 2. Using a convergent mixed-methods design, we analyzed quantitative improvements in scores across four official IELTS band descriptors—Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy—alongside qualitative reflections on learners' perceptions of the feedback system. The results offer compelling evidence that AI-driven feedback, when carefully aligned with test rubrics and used iteratively, can be a powerful tool in promoting writing development in high-stakes assessment contexts. Quantitative analysis of the study showed that learners demonstrated statistically significant improvements across all four writing criteria, with particularly notable gains in lexical resources and grammatical accuracy—areas often underserved in traditional classroom feedback due to time constraints. Qualitatively, learners described the AI

feedback as clear, immediate, and confidence-building. It encouraged more frequent revision, reduced anxiety, and fostered strategic writing behaviors such as self-monitoring and vocabulary variation. Crucially, learners began to engage more deeply with the feedback process, shifting from superficial corrections to more purposeful revisions that addressed structure, logic, and coherence. This engagement reflects not only writing improvement, but also the development of feedback literacy and self-regulated learning skills essential for academic and professional communication beyond the test itself. The study extends current understandings of automated writing evaluation and validates the claim that AI feedback can be more than a surface-level correction tool. Instead, it can function as a scaffold for writing autonomy, learner agency, and iterative skill-building when appropriately framed. In a more practical sense, the value of integrating AI tools into IELTS curricula lies in their potential as complementary instructional support. Teachers can use such systems to offload repetitive corrective work, focus more on higher-order concerns like argument strength, and ensure that learners receive individualized feedback—even in large or resource-limited classrooms.

The findings of this study offer several practical insights for educators, curriculum designers, and assessment specialists interested in integrating AI feedback into IELTS preparation. AI-generated feedback can be used not only as a corrective mechanism but also as a formative instructional tool that promotes revision, reflection, and skill development across the IELTS writing rubric. In many EFL contexts, especially in large classrooms or self-directed learning environments, it is difficult to provide consistent and timely feedback to each learner. The AI system used in this study addressed this challenge by offering immediate, criterion-aligned responses for each submission. This allowed learners to engage in frequent and low-pressure revision cycles, which are often not feasible in conventional classroom settings due to time constraints. The system's structured presentation of feedback—mirroring IELTS band descriptors—helped students take control of their own learning. As students became familiar with the criteria and the types of feedback provided, they began to internalize the test and revise more strategically. This supports the view that feedback literacy can be fostered through structured, AI-supported environments (Carless & Boud, 2018). While AI feedback was effective in improving grammar, vocabulary, and organization, its limitations in addressing argumentative strength and critical analysis suggest that it is best positioned as a complementary tool rather than a replacement for human instruction. Teachers may integrate AI systems into a blended approach, using automated feedback for initial drafts and reserving deeper

rhetorical guidance for follow-up conferences or peer-review sessions.

Several limitations were noted in the present study. The study involved a relatively small sample of 18 undergraduate students from a single university, limiting the generalizability of the findings to broader EFL populations. Although the participants represented typical IELTS learners in terms of proficiency and preparation stage, larger and more diverse samples across multiple institutions would increase the external validity of the results. Future research may consider a quasi-experimental design, including control groups receiving traditional teacher feedback or no feedback at all, to more clearly attribute outcomes to the intervention. The intervention spanned four writing cycles over four weeks, which, while sufficient to observe short-term improvement, may not fully capture long-term retention or transfer of writing strategies. Although the feedback system was aligned with IELTS descriptors, the AI model—like all large language models—could not contextualize arguments deeply or assess critical reasoning.

ACKNOWLEDGEMENTS

REFERENCES

Aljasser, A. (2025). Investigating EFL students' perceptions of feedback: a comparative study of instructor and ChatGPT-generated responses in academic writing. *Education and Information Technologies*, 1-22. https://doi.org/10.1007/s10639-025-13649-8

Banerjee, J., Franceschina, F., & Smith, K. (2007). Documenting features of written language production typical at different IELTS band score levels. *IELTS Research Reports, 7*, 1–84. https://ielts.org/cdn/Research/documenting-features-of-written-language-production-typical-at-different-ielts-score-levels-banerjee-et-al-2007.pdf

Barkaoui, K. (2010). Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly, 7*(1), 54-74. https://doi.org/10.1080/15434300903464418

Biber, D., Nekrasova, T., & Horn, B. (2011). *The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis (TOEFL iBT Research Report No. RR-11-05)*. Educational Testing Service.

Bijami, M., Kashef, S. H., & Kashef, S. M. (2013). Peer feedback in learning English writing: Advantages and disadvantages. *Journal of Studies in Education, 3*(4), 91–97. https://www.macrothink.org/journal/index.php/jse/article/viewFile/4314/3623

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Ferris, D. R. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? *Journal of Second Language Writing, 13*(1), 49–62. https://doi.org/10.1016/j.jslw.2004.04.005

Ferris, D. R., & Hedgcock, J. S. (2014). *Teaching L2 composition: Purpose, process, and practice* (3rd ed.). Routledge.

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence, 6*, 1162454. https://doi.org/10.3389/frai.2023.1162454

Han, J., Yoo, H., Myung, J., Kim, M., Lee, T. Y., Ahn, S. Y., & Oh, A. (2023). ChEDDAR: Student-ChatGPT dialogue in EFL writing education. *arXiv*. https://arxiv.org/abs/2309.13243

Han, J., Yoo, H., Myung, J., Kim, M., Lee, T. Y., Ahn, S. Y., & Oh, A. (2024). RECIPE4U: Student-ChatGPT interaction dataset in EFL writing education. *arXiv*. https://arxiv.org/abs/2403.08272

Hyland, K., & Hyland, F. (2019). Contexts and issues in feedback on L2 writing. *Feedback in second language writing: Contexts and issues*, 1-22. https://doi.org/10.1017/9781108635547.003

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

IELTS.org. (2023). *IELTS Writing Task 2 band descriptors (public version)*. https://www.ielts.org

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Man, D., Xu, Y., & O'Toole, J. M. (2018). Understanding autonomous peer feedback practices among postgraduate students: A case study in a Chinese university. *Assessment & Evaluation in Higher Education*, *43*(4), 527-536. https://doi.org/10.1080/02602938.2017.1376310

Mohammadi, M., Zarrabi, M., & Kamali, J. (2023). Formative Assessment Feedback to Enhance the Writing Performance of Iranian IELTS Candidates: Blending Teacher and Automated Writing Evaluation. *International Journal of Language Testing*, *13*(1), 206-224. https://doi.org/10.22034/ijlt.2022.364072.1201

Navas Bonilla,C. d .R., Viñan Carrasco, L. M., Gaibor Pupiales, J.C., Murillo Noriega, D. E., (2025). The future of education: A systematic literature review of self-directed learning with AI. *Future Internet , 17*, 366. https://doi.org/10.3390/fi17080366

Nguyen, L. Q., & Le, H. V. (2022). Improving L2 learners' IELTS task 2 writing: the role of model essays and noticing hypothesis. *Language Testing in Asia*, *12*, 58. https://doi.org/10.1186/s40468-022-00206-0

Nie, D. (2022). Advancing student learning experience: Peer assessment in advanced legal research classes. *Law Library Journal, 114*(4), 369-394.

Shi, H., Chai, C. S., Zhou, S., & Aubrey, S. (2025). Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer Assisted Language Learning*, 1-28. https://doi.org/10.1080/09588221.2025.2454541

Sumakul, D. T. Y., Hamied, F.A., & Sukyadi, D. (2022). Students' perceptions of the use of AI in a writing class. In the *67th TEFLIN International Virtual Conference & the 9th ICOELT 2021* (TEFLIN ICOELT 2021), 52-27. Atlantis Press. https://doi.org/10.2991/assehr.k.220201.009

Todd, R. W., Khongput, S., & Darasawang, P. (2007). Coherence, cohesion and comments on students' academic essays. *Assessing writing*, *12*(1), 10-25. https://doi.org/10.1016/j.asw.2007.02.002

Uludag, P., & McDonough, K. (2022). Validating a rubric for assessing integrated writing in an EAP context. *Assessing Writing*, *52*, 100609. https://doi.org/10.1016/j.asw.2022.100609

Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. *Frontiers in Psychology*, *14*, 1249991. https://doi.org/10.3389/fpsyg.2023.1249991

Wenden, A. (1991). *Learner strategies for learner autonomy: planning and implementing learner training for language learners.* Prentice Hall

Zhang, K. (2025). Enhancing critical writing through AI feedback: A randomized control study. *Behav. Sci.,15*(5), 1-13. https://doi.org/10.3390/bs15050600