Yanmei YU [1]
Nanjing University of Aeronautics and Astronautics, China

The Role of Generative Artificial Intelligence in IELTS Writing

ABSTRACT

This study targeted intermediate-to-low-level learners of IELTS writing, employing generative AI technology as an intervention to systematically examine its influence on syntactic complexity. A mixed-methods approach was adopted, encompassing an 18-week teaching experiment with 50 participants. Theoretically, the research built upon Ortega et al.'s syntactic complexity framework and innovatively established a dual-track data collection system. This system combined automated assessment using natural language processing dependency parsing tools to extract T-unit structures with manual validation by two experienced IELTS instructors to ensure data reliability.

Empirical analysis via paired-samples t-tests revealed statistically significant improvements in key metrics: the mean number of words per T-unit increased notably, and the clause ratio rose substantially (both $p < 0.05$), confirming AI's positive impact on syntactic complexity. Qualitative feedback from interviews and questionnaires highlighted that while AI excelled at providing real-time feedback, human teachers remained indispensable for fostering critical thinking, motivation, and emotional support.

Based on these findings, the study proposed an "AI-Empowered, Teacher-Led" collaborative writing instruction model. This model integrates AI-driven adaptive feedback for syntactic development with teacher-guided activities focusing on argumentation and creativity. The research offers empirical evidence and practical insights for language teaching innovation in the intelligent education era, advocating blended learning that leverages the strengths of both AI and human instructors.

Keywords: *generative artificial intelligence ; syntactic complexity ; IELTS writing*

Introduction

According to the *Big Data Report on IELTS Test Scores in Mainland China for 2023 - 2024* released by the British Council, the application scenarios of the IELTS test are undergoing a structural transformation. Against the backdrop of continuously rising local recognition, the test has transcended the traditional framework of being solely for overseas study purposes and has become an important reference indicator for domestic higher education institutions to assess language proficiency and for employers to select talent. Along with the diversified development of the test's purposes, the language proficiency of IELTS test-takers in Mainland China has shown an overall upward trend. Among them, academic IELTS test-takers have demonstrated particularly remarkable performance in the writing

[1] Corresponding author: Yanmei YU, y.yanmei@gmail.com

component, with the average score increasing by 0.45 points to 5.79 points compared to five years ago, making it the area with the largest improvement among the four language skills.

In the context of the IELTS test evolving from a language proficiency certification system catering to specific groups' needs to a more popularized one, the enhancement of competitiveness in the writing component, which has traditionally been a weak area in language proficiency, holds special significance. Faced with the realistic challenge of the overall rising proficiency level of test-takers, how to specifically improve the academic IELTS writing ability of students with weak foundations in our college's joint-venture educational programs has become a key breakthrough for teaching innovation. Writing, as an output-oriented task, is significantly influenced by the quality of input. Based on the theoretical framework of second language acquisition, Krashen's "i+1 comprehensible input" principle emphasizes that input materials (i+1) that are slightly above the learner's current level (i) are most conducive to language acquisition. However, in large-class teaching scenarios, achieving personalized input adaptation faces practical difficulties: teachers find it challenging to customize differentiated training materials for dozens of students under traditional teaching models. The intervention of generative artificial intelligence (GenAI) technology provides a technological solution to this problem. Its ability to dynamically generate tiered training materials makes it possible to implement precise language input on a large scale.

Since its inception, GenAI has garnered profound attention from both the academic and industrial sectors worldwide due to its exceptional language generation capabilities, exponential technological iteration speed, and natural language processing architecture driven by deep learning. Especially at a critical juncture of educational digital transformation, this technology is reshaping the traditional paradigm of foreign language teaching with revolutionary force and driving a structural transformation in the relationship between "teaching" and "learning." Currently, the research focus of scholars on the application of GenAI in foreign language teaching has shifted significantly: from early theoretical discussions centered on teachers' perspectives to a growing emphasis on students' actual behaviors and cognitive patterns when using AI technology in foreign language learning (such as English writing). Against this backdrop, this study will focus on exploring the impact of GenAI on students' IELTS writing proficiency and employ qualitative methods to analyze its specific effects on students' syntactic complexity in writing.

Literature Review

In the past two years, the application of GenAI in college English teaching reform has become increasingly widespread, emerging as a crucial driving force for the digital and intelligent

transformation of education. Scholars have conducted systematic explorations in areas such as technology empowerment pathways, teaching effectiveness evaluation, and practical bottlenecks, forming a multi-dimensional research landscape.

At the theoretical construction level, Wang Haixiao (2024), drawing on the teaching reform practice of "General Academic English Writing," innovatively constructed an eight-dimensional framework for generative AI-empowered teaching. This framework systematically integrates core elements such as curriculum design reconstruction, intelligent resource development, and hybrid model innovation, providing an actionable guide for technology-driven teaching transformation that combines theoretical depth with practical effectiveness.

In terms of implementation pathway exploration, scholars have achieved complementary findings by approaching from different teaching scenarios. Jiao Jianli et al. (2023), using ChatGPT as the technological carrier, systematically analyzed its empowerment pathways from four major scenarios: brainstorming, personalized feedback, higher-order thinking cultivation, and automated assessment, and proposed specific strategies for teacher practice. Chen Mo and Lv Mingchen (2024) divided the writing process into four stages -- pre-writing, writing, reviewing, and reflection -- and analyzed the advantages and disadvantages of AI involvement in writing teaching activities at each stage from the perspective of activity theory.

Regarding effectiveness evaluation, Xu Linlin et al. (2024) investigated learners' perceptions and behavioral patterns of AI-assisted academic English writing, finding that AI enhances learners' academic English writing in four dimensions, including language optimization. Both Wang Yabing et al. (2025) and Qin Lili et al. (2025) demonstrated that AI intervention can positively influence students' writing feedback engagement and emotional experiences, thereby enhancing learning motivation and strategy application.

Meanwhile, existing studies have also revealed limitations in AI applications. Yang Linxiu and Ping Jiapeng (2024), based on a metadiscourse perspective, conducted a comparative analysis of the discourse production effects of generative AI and human writing in English academic paper abstracts. They found that while generative AI exhibits unique "human-like" advantages in terms of in-text interactivity and engagement markers, it still shows significant deficiencies in semantic cohesion and coherence, the richness of frame markers, evidentiality expressions, and the construction of academic authority.

Currently, scholars have achieved rich research outcomes on the integration of AI and college English writing teaching from multiple perspectives and dimensions. However, the research still

predominantly focuses on the construction of macro-theoretical frameworks, with insufficient exploration of micro-level aspects such as words, phrases, and sentences in writing assessment criteria.

Methodology

Research Design

The subjects of this study were freshmen majoring in Accounting and International Trade in the SQA Joint-Venture Educational Program at our college. These students generally had below-average English proficiency and moderate learning motivation. A pre-test was conducted at the beginning of the semester, followed by a post-test after a semester of GenAI-assisted writing instruction. For both the pre-test and post-test, students were required to write an IELTS Task 2 essay, which was completed on the Pigai.net platform and graded according to the IELTS essay scoring formula, with the overall evaluation set to an intermediate level. Due to attendance issues, some students only participated in the pre-test, while others only took part in the post-test. Therefore, only essays from students who participated in both tests were selected, and a few extremely low-scoring essays (2 points or below) were excluded, ultimately retaining essays from 50 students as the sample. During the course between the pre-test and post-test, students were required to submit one to two Task 2 essays per week. When writing, students were required to complete the initial draft independently. After finishing the writing, they used GenAI tools for collaborative revision to generate personalized model essays. Finally, students had to submit records of their revision process based on GenAI feedback. Assignments that only included the final draft without the revision process were considered unqualified and returned for rewriting.

This study employed a multi-modal approach for data collection, utilizing the Pigai.net platform for pre-tests and post-tests, and complementing it with semi-structured interviews, questionnaires, and classroom observations. The specific implementation paths included: leveraging Pigai.net's word count and sentence-by-sentence analysis functions, along with manual proofreading by two full-time English teachers, to ensure data authenticity and accuracy; using SPSS 26.0 statistical software to conduct quantitative analysis of students' mean number of words per T-unit and clause ratio in the essays from both the pre-test and post-test. In the middle of the research period, 30 students who volunteered for interviews were selected to participate in approximately half-hour semi-structured interviews focusing on their experiences and perceptions of using AI. During the interviews, teachers recorded key points from the students' conversations and extracted keywords. Meanwhile, an anonymous survey space was created on the Chaoxing platform, and the 50 research subjects were enrolled in a dedicated virtual

class. Structured questionnaires were distributed for data collection, and the system automatically generated word frequency statistics to support qualitative analysis.

Research Findings

To rigorously examine the impact of the one-semester GenAI intervention on students' syntactic complexity--operationalized as the mean word count per T-unit and clause ratio, quantitative analysis was conducted using a paired-samples t-test via SPSS 26.0.

As presented in the paired samples statistics table (Table 1-1), the descriptive data revealed notable differences between the pre-test and post-test results. The post-test mean word count per T-unit stood at 19.1795, with a standard deviation (SD) of 3.14645 and a standard error of the mean (SEM) of 0.70357. In contrast, the pre-test yielded a lower mean of 17.0365, accompanied by a slightly higher SD of 3.74960 and an SEM of 0.83844. The reduced variability in the post-test data suggests that the intervention may have also contributed to greater consistency in students' T-unit word count across the cohort.

The paired-samples t-test results (Table 1-2) further validated the intervention's effectiveness. The mean difference in word count per T-unit was 2.14300, indicating an average increase of approximately 2.14 words per T-unit in the post-test. The SD of this difference was 3.08732, with an SEM of 0.69035, and the 95% confidence interval of the difference ranged from 0.69809 to 3.58791--a range that does not include zero, reinforcing the reliability of the observed increase.

Statistically, the t-value was calculated as 3.104 with 49 degrees of freedom, and the two-tailed significance level was 0.006. This p-value is substantially below the conventional alpha level of 0.05, demonstrating a highly statistically significant difference in the mean word count per T-unit between the pre-test and post-test. Such results confirm that the observed increase in syntactic complexity, as reflected by longer T-units, is not attributable to random variation but is likely a direct outcome of the GenAI intervention.

The paired samples statistics table (Table 2-1) presents clear disparities between the pre-test and post-test clause ratio results. The post-test mean clause ratio reached 0.6205, with a SD of 0.27194 and a SEM of 0.06081. In contrast, the pre-test yielded a substantially lower mean of 0.3785, accompanied by a smaller SD of 0.18559 and an SEM of 0.04150. The higher SD in the post-test data suggests that while the intervention generally improved clause skills, there was greater variability in post-test performance among students—potentially reflecting differences in individual learning trajectories, prior language proficiency, or engagement with instructional activities. This variation, however, does

not diminish the overall significance of the intervention's impact, as the mean difference remains statistically meaningful.

The paired-samples t-test results (Table 2-2) further confirmed the intervention's effectiveness in enhancing students' clause ability. The mean difference in clause ratio was 0.24200, indicating an average increase of 0.242 in the proportion of clauses used by students in the post-test. The SD of this difference was 0.29790, with an SEM of 0.06661, and the 95% confidence interval of the difference ranged from 0.10258 to 0.38142. Notably, this interval excludes zero, which reinforces the reliability of the observed increase and suggests that the intervention consistently promoted clause skill development across the cohort.

Statistically, the t-value was calculated as 3.633 with 49 degrees of freedom, and the two-tailed significance level was 0.002. This p-value is considerably below the conventional alpha level of 0.05, demonstrating a highly statistically significant difference in clause ratios between the pre-test and post-test. Such findings indicate that the observed improvement in students' ability to use clauses is not a product of random variation but is likely the direct result of the AI assistance in the English writing course.

The semi-structured interviews, conducted with 30 participating students, centered on two core themes: students' specific strategies for using GenAI writing tools and their assessments of the generated texts. The findings revealed distinct variations in students' usage patterns, reflecting differences in digital literacy, learning autonomy, and familiarity with AI tool functionalities.

A small subset of students (approximately 15%) adopted a passive usage mode, treating AI merely as a basic search tool. Their interactions with AI were limited to inputting simple prompts (e.g., "Write an essay on environmental protection") without leveraging advanced features such as follow-up questioning, multi-perspective inquiry, or iterative revision requests. These students reported that the AI-generated model essays were often "overly complex" and "far beyond their current English proficiency level"—characterized by dense syntactic structures, specialized vocabulary, and sophisticated argumentation that they struggled to comprehend or emulate. For this group, AI's utility was constrained by a mismatch between the generated content and their linguistic competence, leading to limited learning gains from the tool.

In contrast, the majority of students (around 65%) demonstrated proactive and strategic usage of AI tools. They employed targeted techniques to refine AI outputs, such as asking follow-up questions to clarify ambiguous points, like "Can you explain how this argument is supported?", requesting multi-perspective analyses, or demanding iterative revisions. These students engaged in a collaborative "dialogue" with AI, using the tool as a co-constructor of writing rather than a passive content provider.

They reported feeling satisfied with the final AI-generated texts, as the iterative refinement process ensured the content aligned with their proficiency level, writing goals, and personal expression needs.

Notably, across both usage groups, students exhibited a high level of trust in AI-generated results. Only a small minority (about 20%) expressed mild skepticism—primarily regarding the originality of AI's arguments or the appropriateness of its vocabulary in specific contexts. Most students perceived AI outputs as "reliable" and "authoritative," citing the tool's ability to generate coherent, grammatically accurate content as a key reason for their trust. This high trust level suggests that students generally view GenAI as a credible learning aid, though it also raises potential concerns about over-reliance or uncritical acceptance of AI-generated content. When asked about their primary focus when using GenAI to revise their own essays, the interview responses revealed clear priorities. The largest proportion (33.2%) of students identified advanced vocabulary as their main concern, expressing a strong interest in acquiring and applying words they could not independently produce. For these students, AI served as a "vocabulary expansion tool," helping them replace simple or repetitive words with more precise, academic, or contextually appropriate alternatives. Additionally, 27.8% of students highlighted AI's ability to generate unique perspectives as a key draw -- particularly for unfamiliar writing topics, where AI's multi-dimensional insights helped them break through creative blocks. Smaller proportions of students focused on grammar (16.7%) and essay structure (11.1%), using AI to identify grammatical errors, improve sentence coherence, or optimize the logical flow of their writing. These priorities reflect students' self-awareness of their weaknesses in English writing and their strategic use of GenAI to address specific gaps.

The keyword frequency analysis from Chaoxing platform (Figure1) revealed four core areas of perceived benefit: viewpoint generation, advanced vocabulary, grammar support, and structural guidance -- consistent with the priorities identified in the interviews. Among these, "viewpoint" emerged as the most frequently mentioned keyword, indicating that students perceive AI's greatest value in helping them generate ideas and develop arguments. For many students, especially when facing abstract or unfamiliar topics, AI's ability to propose diverse, well-structured viewpoints addressed a key pain point: the lack of concrete ideas or logical frameworks for writing.

Following viewpoint generation, "advanced vocabulary" and "grammar" were the next most commonly cited benefits. Students reported that AI not only provided them with new vocabulary but also offered context-specific usage examples, helping them understand how to integrate advanced words into their own writing effectively. Similarly, AI's grammar-checking and revision functions were valued for their ability to identify errors that students might overlook, such as tense inconsistencies, preposition misuse, or subject-verb agreement issues. Finally, "structure" was

identified as a key benefit, with students noting that AI-generated model essays provided clear templates for organizing introductions, body paragraphs, and conclusions -- helping them improve the coherence and logical flow of their own writing.

Collectively, the questionnaire results reinforce the interview findings, confirming that AI tools address multiple layers of students' writing needs—from idea generation to linguistic refinement and structural organization. The consistency between qualitative and quantitative data enhances the reliability of the study's conclusions, highlighting AI's multifaceted utility in supporting L2 writing development.

Discussion

Syntactic complexity stands as a cornerstone concept in second language (L2) writing teaching and research, as learners' syntactic development is not merely a peripheral skill but a core component of target language acquisition that reflects the depth of their linguistic mastery (Ortega, 2003). This concept encapsulates two interrelated dimensions: the diversity of syntactic structures a learner can deploy and the degree of sophistication with which these structures are integrated to convey meaning. Unlike basic grammatical accuracy, which focuses on avoiding errors, syntactic complexity emphasizes the ability to manipulate language to express complex ideas—an essential skill for academic and communicative success in L2 contexts. As defined in existing literature, syntactic complexity is generally operationalized through two key measurable indicators: T-unit length and clause ratio (Bao Gui, 2009). Notably, Ortega (2003) offered a nuanced perspective, arguing that syntactic complexity should not be simplistically equated with overall language proficiency. She highlighted that high-proficiency L2 writers often shift from clausal complexity to phrasal rhetorical complexity, such as nominalization or prepositional phrases. This shift reflects a move toward the concise, information-dense syntactic style valued in academic writing. However, this distinction is particularly relevant for advanced learners; for lower-proficiency learners, clausal complexity remains a critical marker of development. Jayanti Banerjee's research found that the mean clause length per T-unit is a more sensitive indicator of proficiency differences among lower-level L2 learners than phrasal complexity metrics. This is because lower-proficiency learners typically lack the linguistic repertoire to deploy phrasal sophistication, making clausal complexity, such as integrating relative clauses or adverbial clauses, the primary way to demonstrate syntactic growth. Consistent with this framework, the pre-test and post-test IELTS Writing scores of the study's participants primarily ranged from 3.0 to 5.5, with only one outlier score of 7.0—confirming that the cohort was predominantly composed of lower-proficiency learners. For this group, the statistically significant increase in the average number

of words per T-unit and clause ratio observed post-intervention is not merely a quantitative change but a meaningful reflection of improved writing proficiency. It indicates that learners were able to elaborate ideas more fully, integrate dependent clauses more effectively, and deploy syntactic structures with greater control -- all key indicators of progress in L2 acquisition for this learner demographic..

As a productive language assessment, IELTS Writing imposes higher requirements on candidates' language proficiency. Effective input is the guarantee of efficient output. The changes in the average number of words per T-unit and clause ratio between the pre-test and post-test indicate that the revised versions provided by GenAI based on students' drafts are appropriate input materials for students. When students see their revised essays become more readable, they gain a stronger sense of self-identity and enhanced self-efficacy, which improves their output quality and motivates them to invest more in exam preparation. Meanwhile, the emphasis on process over results in regular assignment evaluation corrects students' perception of tasks -- completing a task is not merely submitting a perfunctory outcome. This individual understanding of task requirements directly influences their behavior, turning each assignment into a learning process. Such immersive learning is more likely to sustain students' learning enthusiasm compared to cramming, thereby facilitating the improvement of their English proficiency. Long-term use of AI-generated model essays as learning templates and AI scoring may lead to measurable progress in students' overall writing scores. This progress, especially for students with lower English proficiency, provides tangible learning motivation and further improves their overall learning attitude. However, it is crucial to remain vigilant: to perform effectively, lower-level language learners may focus more on imitating or even directly copying so-called "universal sentence patterns" from previously generated model essays in subsequent tasks. This may make it seem that their English proficiency has improved in a short period, but such improvement comes at the cost of sacrificing self-expression -- they choose to make fewer mistakes over expressing their own ideas better. Therefore, a key focus repeatedly emphasized in cultivating students' information literacy is the development of critical thinking. This is particularly important in IELTS Writing. If candidates blindly rely on model essays and dare not attempt to express themselves, they will struggle to achieve qualitative breakthroughs in English learning. The suggestion proposed by Wen Qiufang et al. (2006) to avoid the tendency of emphasizing language skill training over thinking ability development in writing teaching is equally applicable to IELTS Writing instruction.

The standard deviation of students’ clause ratios in the post-test (.27194) was notably higher than that in the pre-test (.18559), a statistically observable shift that signals a widened range of variation in syntactic complexity across students’ IELTS Task 2 essays following the intervention. Beyond a mere

quantitative discrepancy, this increase in variability reflects a deepening performance gap among learners—one that can be directly linked to the integration of AI tools in the writing learning process and its uneven impact on different student groups. In traditional in-class teaching contexts, the learning environment is characterized by relative uniformity in instructional input. Teachers typically deliver the same quality and type of content, guidance, and feedback to all students, regardless of individual differences in proficiency, learning styles, or motivation. While inherent variations in students' cognitive abilities, prior knowledge, and engagement levels do lead to differences in learning outcomes, these gaps are often constrained by the structured nature of classroom instruction. Teachers can adjust their pacing, offer targeted scaffolding, or provide supplementary support to mitigate extreme disparities, ensuring a baseline level of equity in learning opportunities. In this model, the "playing field" remains relatively level, as the primary driver of performance differences is individual effort rather than differential access to high-quality learning resources.

However, the advent of the AI era has disrupted this balance, introducing new dynamics that amplify existing inequalities. A critical observation from the study is that even when students interact with the same AI writing tool, the outcomes of their engagement can vary drastically—often independent of the tool itself. Unlike traditional teaching resources that deliver fixed content, AI tools operate as "responsive systems" whose outputs are shaped by users' input strategies, digital literacy, and proactive engagement. Students who pose vague, generic questions, like "Improve my essay", may receive superficial, one-size-fits-all feedback, while those who use precise, iterative prompts, like "Can you rephrase this paragraph using more complex subordinate clauses?", can elicit targeted, high-quality support tailored to their specific needs. This means that the same AI tool can function as a "learning accelerator" for some and a "passive resource" for others, depending on how effectively students leverage its capabilities.

The root of this widening gap lies in two interrelated factors: students' AI literacy and their intrinsic learning motivation. For students who approach AI with a passive, perfunctory attitude -- treating it as a "shortcut" to complete assignments rather than a learning partner -- engagement is limited to minimal interactions. They are unwilling to invest time in mastering advanced usage techniques, such as follow-up questioning, multi-perspective inquiry, or iterative revision requests. As a result, they gain little from the tool beyond surface-level adjustments and fail to develop the syntactic complexity, critical thinking, or rhetorical skills that the intervention aimed to foster. For these students, AI does not supplement their learning but merely replaces their own effort, leading to stagnant or minimal progress. In stark contrast, students with strong learning motivation and proactive learning habits view AI as a collaborative tool to enhance their capabilities. They actively experiment

with different prompt strategies, reflect on the feedback received, and use AI to address specific gaps in their writing . Over time, this iterative engagement not only improves their immediate writing outcomes but also builds their "AI literacy". As their AI literacy grows, they receive increasingly relevant and high-quality feedback, creating a positive feedback loop: better use of AI leads to improved English proficiency, which in turn enables more sophisticated interactions with AI. This group thus pulls further ahead, widening the performance gap with their passive peers.

If this trend remains unaddressed, the implications for educational equity are profound. Educational equity, at its core, emphasizes fair access to learning opportunities and the elimination of barriers that prevent students from reaching their full potential. However, AI-driven learning environments risk creating a "digital divide 2.0"--not one based on access to technology, as most students in this study had access to the same AI tools, but on the ability to use technology effectively. This divide marginalizes students who lack AI literacy, learning autonomy, or supportive learning contexts, trapping them in a cycle of limited progress while their more engaged peers advance. Over the long term, this could exacerbate existing social inequities, as academic success in an AI-integrated education system becomes increasingly tied to skills that are not evenly distributed across student populations.

These concerns align with the warnings outlined in UNESCO's Guidelines on the Use of Generative Artificial Intelligence in Education and Research, which highlights that while generative AI offers significant potential to enhance teaching and learning, it also poses inherent risks to educational equity. The document notes that AI's technical limitations—such as its reliance on user input quality and its tendency to reinforce existing biases—can exacerbate issues like information cocoons and digital divides. These risks, if unmitigated, can intensify social inequities, marginalize vulnerable groups, and perpetuate unbalanced development in education. For IELTS writing instruction specifically, this means that the benefits of AI tools may be concentrated among already advantaged students, while others are left behind—undermining the goal of providing equitable support for all learners.

Interview and questionnaire results show that students have a high degree of recognition for AI-generated model essays and feedback, and they will earnestly learn the parts that are helpful to them. Although students did not specifically focus on sentence construction skills, there was a substantial improvement in their sentence writing. This provides important insights for IELTS teaching. In IELTS Writing, especially when facing topics far removed from students' life experiences, the "ice-breaking" role of ideas is particularly crucial. Unfamiliar topics often increase students' cognitive load and risk a lack of ideas. When dealing with unfamiliar topics, students need to invest more energy in

understanding the topic, collecting materials, and organizing ideas, which may occupy resources that could otherwise be used for syntactic processing, leading to a decline in syntactic complexity. At the same time, unfamiliar topics are more likely to result in a lack of ideas, making sentences empty and logically confusing, which further affects the display of syntactic complexity. AI-generated model essays can provide targeted ideas and writing materials, helping students familiarize themselves with writing frameworks for different topics, enhance their adaptability to unfamiliar topics, and encourage them to use more complex syntactic structures to express their views, thereby improving their IELTS Writing scores. Therefore, IELTS Writing teaching should not be confined to mere language training; expanding students' knowledge and helping them enhance their perception of social life should also be part of the teaching syllabus.

Additionally, during the semi-structured interviews, a consistent theme emerged from students' reflections: the initial novelty of interacting with generative AI writing tools tends to fade over time, leading many to gradually experience fatigue, monotony, or even a loss of enthusiasm for sustained interaction. This shift in engagement is rooted in the inherently functional, task-driven nature of AI tools—while they excel at delivering technical feedback and generating content, they lack the dynamic, empathetic, and adaptive interaction that characterizes human communication. In the early stages of use, students are often curious about AI's ability to generate essays, suggest vocabulary, or revise sentences, which motivates active exploration. However, as this novelty wears off, the repetitive nature of AI interactions can lead to a sense of disconnection. Students reported feeling "unmotivated to continue asking questions" or "indifferent to AI's responses" after several weeks, as the tool fails to recognize their evolving learning needs, celebrate small wins, or adapt its tone to match their emotional state. This decline in engagement underscores a critical limitation of AI: its inability to sustain long-term learner motivation through relational connection, making teachers irreplaceable in maintaining students' investment in the writing process—especially in writing evaluation, where personalized guidance and humanized feedback are key to keeping learners engaged.

Another fundamental reason AI cannot substitute teacher feedback lies in the multi-dimensional nature of human guidance—teachers' feedback encompasses far more than just language knowledge; it includes emotional affirmation, tailored encouragement, and genuine spiritual support that directly fosters students' self-efficacy. Self-efficacy, defined as an individual's belief in their ability to succeed in specific tasks, is a key predictor of learning outcomes in L2 writing. Teachers nurture this belief through intentional, personalized interactions: they might highlight a student's progress in using complex sentences, validate their unique ideas, or reassure them during setbacks. These moments of emotional support are not peripheral to teaching; they are integral to helping students build the

confidence to take linguistic risks, persist through challenges, and view themselves as capable writers. In contrast, AI feedback is narrowly focused on linguistic form and content, lacking the warmth, empathy, and personalization that make emotional affirmation meaningful. Even the most advanced AI tools cannot replicate the sincerity of a teacher's encouragement or the reassurance of someone who understands a student's individual journey, strengths, and struggles.

Conclusion

This study targeted intermediate and low-level IELTS candidates, a group facing persistent writing barriers: over-reliance on simple sentences, poor clause integration, limited advanced vocabulary, and logical incoherence with complex topics. It explored GenAI as a targeted intervention to address these needs and foster syntactic growth. After a 18-week AI-assisted program, paired-samples t-tests revealed highly significant progress in mean word count per T-unit and clause ratio. These findings validate AI's efficacy for lower-level learners, and highlight its potential to democratize high-quality, personalized writing support—hard to deliver in traditional classrooms due to time constraints and diverse needs.

The study also signals a paradigm shift from the traditional teacher-student binary to a tripartite ecosystem: teachers, students and AI. In this ecosystem, AI handles routine tasks like grammar feedback, idea generation, and vocabulary suggestions, freeing teachers to focus on personalized pathways, critical thinking facilitation, emotional support, and AI literacy guidance. For students, this model enables anytime access to scaffolded feedback, boosting learning ownership and confidence. In IELTS Writing instruction, teachers should leverage AI's adaptability and data-analytic capabilities to develop "one student, one plan" interventions. To mitigate widened performance gaps, teachers must instruct students on effective AI use to ensure equitable access to benefits.

Ultimately, integrating AI aims to help lower-level candidates overcome linguistic barriers, build confidence, and improve IELTS Writing scores—key to their academic aspirations. This study provides a roadmap for responsible AI integration, centering learner needs, equity, and sustainable writing development.

Mean word count per T-unit

Table 1-1

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Post-test	19.1795	50	3.14645	.70357
	Pre-test	17.0365	50	3.74960	.83844

Table 1-2

Paired Samples Test									
		Mean Difference					t	df	Sig.（2-tailed）
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Post-test - Pre-test	2.14300	3.08732	.69035	.69809	3.58791	3.104	49	.006

Subordination ratio

Table 2-1

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair1	Post-test	.6205	50	.27194	.06081
	Pre-test	.3785	50	.18559	.04150

Table 2-2

Paired Samples Test									
		Mean Difference					t	df	Sig.（2-tailed）
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Post-test - Pre-test	.24200	.29790	.06661	.10258	.38142	3.633	49	.002

Figure 1 *Word cloud of questionnaire-based keyword statistics*

References

Banerjee J. Franceschina F.& Smith A, M. (2007). Documenting features of written language production typical at different IELTS band score levels. *IELTS Research Reports Volume, 7*, 1-56

Bao, G. (2009). A study on the changes of syntactic complexity in English learners' compositions. *Foreign Language Teaching and Research*, *4*, 291-297.

British Council. (2024). IELTS test taker performance data report: Mainland China [EB/OL]. https://www.chinaielts.org/press-office/whitepaper-2024

Chen, M., & Lü, M. C. (2024). College English writing teaching in the ChatGPT environment. *Contemporary Foreign Language Studies*, *1*, 161-168.

Hunt KW. (1965). *Grammatical structures written at three grade levels*. Champaign: The National Council of Teachers of English.

Jiao, J. L., & Chen, T. (2023). Large language models empowering English teaching: Four scenarios. *Computer-Assisted Foreign Language Education*, *2*, 12-17.

Krashen S. D. (1985). *The input hypothesis: Issues and implications*. London: Longman Group Limited.

Ortega L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing. *Applied Linguistics*, *24*, 492-518.

Qin, L. L., Dong, J. J., & Zhang, A. N. (2025). A study on English learners' emotional experiences in narrative writing revision mediated by generative AI: A Q-method research on continuation writing teaching. *Foreign Languages in China*, *2*, 61-70.

Qiu, Y., Liu, J. X., Jiang, M., & Ding, Q. F. (2023). *AI questioning made easy*. Posts & Telecom Press.

UNESCO. (2023). Guidance for generative AI in education and research [EB/OL]. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000386693

Wang, H. X. (2024). An exploration of the application of generative AI in college English teaching reform: Taking the teaching reform practice of "General Academic English Writing" as an example. *Foreign Language Education Research in China*, *4*, 41-50, 95.

Wang, Y. B., Xuan, Q. Y., & Zhang, J. L. (2025). The impact of generative AI intervention on college students' engagement with feedback in foreign language writing. *Foreign Languages in China*, *2*, 71-78.

Wen, Q. F., & Liu, R. Q. (2006). An analysis of college students' abstract thinking characteristics through English argumentative essays. *Journal of Foreign Languages*, *2*, 49-58.

Xu, L. L., Hu, J. H., & Su, Y. (2024). A study on learners' cognition and behaviors in AI-assisted academic English writing. *Foreign Language World*, *3*, 51-58.

Yang, L. X., & Ping, J. P. (2024). Generative AI and academic discourse writing from the perspective of metadiscourse. *Contemporary Rhetoric*, *6*, 11-23.

Zhang Z ＆ Hyland K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing, 36*, 90-102.

Zhao, Q., & Wang, S. Y. (2020). A study on the syntactic complexity of English compositions by students of different proficiency levels. *Foreign Language Testing and Teaching*, *3*, 21-28.