Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.36348

Leveraging Large Language Models for IELTS Learning: A Literature Review

Zhiyang Yang

Xuanhan High School, Sichuan, China
273935296@qq.com

Abstract. Existing research on the application of large language models in IELTS teaching assistance is scattered and fragmented, and there is a lack of systematic reviews on the effectiveness of large models in the complete teaching process. To clarify the role and utility of large language models in the field of IELTS teaching, this paper systematically sorts out relevant studies on large models in the four aspects of listening, speaking, reading and writing, and draws conclusions in combination with the current practical requirements of the IELTS test. The study finds that large language models feature high efficiency and strong customizability in IELTS teaching, and can serve as learning tools to help candidates improve their scores. Meanwhile, they also have limitations and potential risks such as academic integrity crises and insufficient multimodal capabilities. This study enriches the research on the application of large language models in IELTS preparation, and provides references for follow-up research and teaching practice.

Keywords: Large Language Models, IELTS Teaching, Four English Language Skills

1. Introduction

The IELTS test is one of the most widely recognized English language proficiency tests in the world [1]. IELTS scores are often used as a standard to measure English proficiency in scenarios such as overseas study and academic research. Therefore, the number of test takers worldwide grows rapidly every year, and the demand for corresponding educational resources increases accordingly. Traditional offline or online human-led training often suffers from delayed feedback, high costs, limited learning scenarios and insufficient teaching staff [2], which conflicts with IELTS candidates' demand for high-quality, efficient and low-cost education within a short period. In recent years, various large language models represented by ChatGPT have made frequent technological breakthroughs [3]. With their powerful text processing and information generation capabilities, they have become an important auxiliary tool in the field of IELTS teaching, providing new solutions for IELTS preparation.

With the deepening application of large language models in the field of education, a large number of academic studies have confirmed their practical effects and application value in IELTS tutoring [4-9]. Existing studies mostly focus on the performance of large language models in a single dimension such as reading, writing or listening, while research specifically targeting the IELTS examination system and aligning with its four skills of listening, speaking, reading and writing is

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.36348

still relatively scattered. There is a lack of overall analysis of the application scope, advantages and risk limitations of the four modules respectively [5, 7, 9]. Research on large language models in the field of IELTS teaching still needs to be improved.

Against this background, this paper takes relevant literature from recent years as research materials and focuses on three core issues. First, the application of large language models in the four aspects of IELTS listening, speaking, reading and writing; second, the advantages of large language model-assisted learning confirmed by current research; third, the risks and deficiencies of the application of large language models in this field. Based on the above core research questions, this paper starts with relevant concepts of large language models, the IELTS scoring system and intelligent language learning, then sorts out the research status of large language models in recent years, and finally summarizes the risks, challenges and future prospects of their application, as well as generalizes the findings of the full paper.

2. Relevant concepts and theoretical foundations

2.1. Overview of large language models

A large language model is an artificial intelligence model that completes pre-training based on massive multi-domain text corpora and realizes natural language understanding, logical reasoning and fluent natural language generation through deep learning architectures. In teaching practice, the model demonstrates great application potential with its core language capabilities such as semantic recognition, long-text context memory, multi-round interaction, text reconstruction and logical error correction. At present, the mainstream large models at home and abroad are as follows: overseas models include ChatGPT, Anthropic Claude, Google Gemini, etc.; domestic models include Baidu Wenxin Yiyan, Alibaba Tongyi Qianwen, Kimi, iFlytek Spark, etc. Compared with human-led teaching, it has significant advantages in output efficiency and output stability.

2.2. IELTS test structure and scoring system

The IELTS test system was established in 1989 and is jointly owned by the British Council, IDP: IELTS Australia and Cambridge Assessment English. It has now become an international test designed to assess the English proficiency of non-native English speakers who intend to study or work in environments where English is used as the medium of communication, and its popularity continues to rise. The IELTS test fully covers four independent modules: listening, reading, writing and speaking, with a full overall score of 9 bands and independent scoring for each section. Writing and reading consist of objective question types, with corresponding scores converted according to answer keys; speaking and listening focus on comprehension ability and reaction speed.

In the writing section, compared with other test types, the IELTS test has higher requirements for logicality, requiring on-topic viewpoints and a clear, well-organized writing framework. At the language level, it also requires rich lexical resource and extensive, accurate grammatical usage. Speaking is a compulsory component of the IELTS test, and its scoring is mainly based on fluency and coherence, lexical resource, grammatical range and accuracy, and other dimensions.

2.3. Basic concepts related to intelligent-assisted language learning

CALL (Computer-Assisted Language Learning) is the precursor theoretical system of AI-based language education. It refers to the use of computer hardware and software tools to optimize

DOI: 10.54254/2755-2721/2026.36348

language input and output training, breaking through the limitations of time and teaching staff in traditional classrooms. Computer-assisted language learning and teaching provide language teachers and learners with a wealth of activities that are carefully designed in curriculum development and help language learners grasp the language more easily and deeply. The role of computers in the classroom as part of curriculum development has gone through several stages of evolution, and has now become an integrated component of the language learning process [10]. The core logic of AI language education is to use intelligent models to identify learners' weaknesses in vocabulary, grammar and logic, and generate targeted exercises, correction feedback and expanded materials, so as to realize individualized teaching.

3. Research status of the application of large language models in IELTS learning

As a highly internationally recognized oral English proficiency test, the IELTS speaking test attracts a large number of English learners. It is assessed on fluency and coherence (FlC), lexical resource (LR), grammatical range and accuracy (GRA), and pronunciation (Pro) [6]. However, in the learning process, learners often face a series of difficulties that hinder oral English learning and test performance, such as restrictions of time and location. In recent years, the wide application of large language models in various fields has also extended to the field of IELTS speaking learning, and most studies have given positive evaluations of their application in this regard. Xing and Saeed systematically reviewed the application, benefits and obstacles of artificial intelligence in speaking skills in 2025, revealing the role of AI in improving ESL/EFL speaking skills [11]. By analyzing different AI tools, learning environments and research methods, the study concluded that AI tools significantly improved learners' oral fluency and pronunciation. Many participants stated that frequent use of AI-powered tools such as Duolingo and ELSA Speak helped improve their speaking skills. These tools provide real-time feedback, enabling learners to quickly identify and correct pronunciation errors, thus accelerating language learning progress, especially in terms of accent and fluency. Meanwhile, Saraswati et al. pointed out that interacting with AI chatbots like Replika helps learners correct pronunciation errors, expand vocabulary and build speaking confidence [12]. Similarly, the study by Annamalai et al. also confirmed the strong performance of AI tools in correcting pronunciation and listening errors [13], thereby improving learners' overall English proficiency. Nguyen used three types of data—questionnaires, interviews and teaching journals—in 2025 to reveal students' perceptions of ChatGPT's application in IELTS speaking skills, as well as teachers' reflections on the role of ChatGPT in developing IELTS speaking materials [8]. The results show that both teachers and students respond positively to the application of ChatGPT. It not only helps teachers create useful classroom resources and provides scaffolding for students in terms of outlines, vocabulary, ideas and sample responses, but also serves as a valuable self-study tool for students, described as easy to use, time-saving and fast to respond. To sum up, large language models, with their unique efficiency and information processing capabilities, play a positive role in IELTS speaking teaching, but their context awareness and response capabilities still need to be improved.

In the IELTS examination system, listening accounts for a large proportion and is a distinctive feature compared with other test systems. The listening test mainly examines candidates' ability to process details as well as their comprehension and application of the language. The shortage of listening training materials and strict time requirements are common difficulties faced by the majority of test takers. Regarding research on large language models in listening-assisted training, Raigul Zheldibayeva studied the performance of two generative AI tools, ChatGPT and Gemini, in helping non-English major candidates improve their scores in 2025 [14], with Gemini responsible

DOI: 10.54254/2755-2721/2026.36348

for the listening training component. The study selected 93 undergraduate students with CEFR B1 level in Kazakhstan as subjects, 48 of whom were in the experimental group (with large models integrated into the training process) and 45 in the control group, who only used traditional paper-based learning materials for training. The experiment set three assessment time points: pre-intervention, post-intervention and a summer delayed retest. Senior foreign teachers scored the candidates' responses and collected experimental data. Quantitative data showed that the listening scores of the experimental group were significantly higher than those of the control group after the intervention, but at the summer retest, the gap between the two groups was significantly smaller than at the post-intervention point. The experimental results indicate that large models have a significant effect in helping candidates improve their scores in the short term, but the effect is short-lived, and long-term intervention is required to better enhance candidates' listening ability. The experiment also found that the listening content generated by large models has problems such as unbalanced difficulty and repetitive question types. Effatpanah investigated the learning performance of 310 Iranian IELTS candidates in his study, aiming to explore which cognitive diagnostic assessment (CDA) model can reasonably reflect the underlying interaction of second language listening comprehension and identify Iranian candidates' strengths and weaknesses in the IELTS listening section [6]. The results show that cognitive diagnostic assessment shows potential in providing rich diagnostic information about students' learning status. Therefore, more attention should be paid to the design and development of educational assessment based on the CDM framework in second language/foreign language contexts. This urgently requires collaboration among experts from different fields such as subject teaching, measurement science and pedagogy. Panahi discussed learners' strengths and weaknesses in each subsection of the IELTS listening test based on three cognitive diagnostic models: DINA, DINO and the hybrid DINA-DINO model [15]. The results show that all three models set lower requirements for fill-in-the-blank questions than for other subsections. However, they did not show consistent results for multiple-choice sub-skills. Therefore, it can be concluded that CDM-based diagnostic information for IELTS listening subsections can be used to explore the underlying structure of the test in the education system, and to detect learners' potential strengths and weaknesses (including mastered and unmastered skills) on target items. To sum up, large language models have great value in providing listening materials by virtue of their advantages in capability, customization, multilingual processing, accuracy and real-time updating.

In terms of writing and reading, large language models have also shown considerable potential. In May 2025, Areeg Fahad Rasheed, Shimam Amer Chasib et al. explored the scoring reliability of four types of large models for IELTS writing: ChatGPT (including versions 4, 4O, 4O-Mini and 3.5), Claude 3.5 Sonnet, Gemini Flash, and Copilot [9]. The study used 34 Task 1 essays and 34 Task 2 essays as materials, which were scored by large models and Cambridge University experts respectively, with data collected accordingly. For the large model group, the researchers used two types of prompts to generate results respectively. The two groups of data were compared using R^2 and degree of bias to examine the differences between the human scoring group and the large model group. The results show that ChatGPT-4 can accurately score all 68 essays, performing far better than other models, followed by ChatGPT-4o. In a 2026 study by Michael James Day, 76 students were recruited as subjects to explore their views on using generative artificial intelligence (GenAI) to improve reading and writing skills and prepare for language learning and examinations [5]. The results show that about 61% of students held an optimistic attitude towards AI-assisted learning tools supporting EAL in VLE discussions. Another 30% held a cautiously optimistic attitude, acknowledging the potential benefits while noting their limitations. About 10% of responses emphasized critical perspectives. In educational practice, generative artificial intelligence tools can

DOI: 10.54254/2755-2721/2026.36348

be integrated as scaffolding learning resources (rather than replacements for independent practice) and play an effective role in IELTS preparation. Therefore, educators can combine AI-assisted activities (such as guided writing review, vocabulary exploration and exercise generation) while encouraging students to critically evaluate AI-generated answers. The above studies all show that large models have certain functional capabilities in corresponding fields, and have delivered outstanding performance in writing scoring and reading question analysis, with broad development prospects in the future.

Studies on all four skills generally give positive evaluations of the role of large models in IELTS teaching. Taken together, research on the four skills shows that large language models can provide IELTS candidates with a low-cost, high-efficiency learning support system. For writing, large models can enable real-time correction, logical optimization and sample essay generation for daily writing practice. For speaking, the simulation of oral interview scenarios by large models can alleviate test takers' speaking anxiety [16] and improve their oral English ability. For reading, large models excel at breaking down long and difficult sentences for grammatical analysis and sorting out the structure of articles [17]. For listening, large models support script creation and detailed explanation of difficult points. On the whole, large models can serve as efficiency-enhancing tools for IELTS learning, but they still cannot replace human teaching. Reasonable integration with teacher-led instruction can maximize the effectiveness of test preparation.

4. Challenges, limitations and risks

4.1. Subjectivity and reliability of scoring

As shown in the study by Chasib et al., even the best-performing large model still shows discrepancies with human scoring. The official IELTS scoring system demands extremely high professionalism, standardization and humanistic subjectivity, which are difficult for large models to fully satisfy. Meanwhile, when facing essays with flexible sentence structures, original ideas, or only minor logical and grammatical errors, large language models cannot guarantee accuracy in all cases. This means that the scores and feedback given by LLMs may not be consistent with official IELTS scoring, and deviations may exist.

4.2. Lack of multimodal capabilities

In a real speaking interview, variables such as the test taker's pronunciation accuracy, accent and tone need to be processed simultaneously, which poses a significant challenge to large models and places extremely high demands on their audio processing capabilities. Factors such as contextual changes, logical rationality, intonation and speech rate control in listening texts are difficult for text-only large models to handle. Therefore, their application in formal IELTS examinations still needs to be improved.

4.3. Over-reliance and academic integrity

The recent development of large models has enabled them to assist with tasks such as article conception and language polishing. Long-term reliance on large models for such assistance can easily cause users to lose their independent thinking ability [18]. At the same time, the application of large models poses new challenges to the detection of academic misconduct. Using large models to

DOI: 10.54254/2755-2721/2026.36348

generate academic papers undermines the rigor of academic research [19], while generating IELTS essay templates will affect the fairness of the examination.

5. Future prospects

The rapid development of artificial intelligence in recent years has demonstrated remarkable progress in assisting IELTS teaching, indicating its great potential and broad development space in the field of education. In view of the advantages and disadvantages of large language models in IELTS teaching summarized above, future iterations of large models can be advanced in two directions: function expansion and standardized application mechanisms. At the technical level, with the growth of computing power and sample size, the gap between large models and human scoring will gradually narrow. Their capabilities in audio processing, text analysis, and handling complex contexts and grammar will continue to improve with increasing computing power and data volume. When their accuracy and stability reach a certain level, they can serve as an alternative scoring reference parallel to human scoring. If the accuracy meets the standard, they can be used to improve the precision of final scoring, or act as an independent IELTS tutoring unit to realize comprehensive and integrated IELTS preparation support. Meanwhile, this requires the improvement of relevant mechanisms. At the application management level, application guidelines can be formulated and gradually improved in the future according to the development status of large models, so as to avoid problems such as over-reliance, academic fraud and scoring errors, thereby enhancing the credibility of large model applications. On the whole, refinement, multimodality and standardization will become the core development trends of large models. Their role in the field of IELTS teaching will also gradually expand, with a focus on further embodying the features of being "all-round" and "high-efficiency".

6. Conclusion

This paper systematically sorts out the application scenarios and effectiveness of large language models in the four aspects of IELTS listening, speaking, reading and writing. The study finds that large language models have significant advantages in simulating oral dialogues and generating customized listening texts. Compared with traditional human-led teaching, large language models have the advantages of high efficiency, instant feedback, customizability and low cost, and can be used as supplementary resources to address the problems of delayed feedback and time/location constraints in traditional teaching. While boasting many advantages, large language models also have several shortcomings, including deficiencies in multimodal perception, a writing scoring mechanism and accuracy that have not yet reached a practical level, as well as academic integrity crises and over-reliance caused by imperfect relevant mechanisms. Although AI tools like ChatGPT can significantly improve resource availability and provide basic knowledge, their integration into IELTS preparation should be combined with human interaction. At the same time, despite their limitations and inability to fully replace human tutors, AI tools like ChatGPT offer valuable benefits and should be integrated into teaching practice [20]. The limitation of this study is that it only reviews published literature without empirical verification. So far, large models are positioned as auxiliary efficiency-enhancing tools in IELTS teaching, helping candidates improve their English proficiency with their unique high efficiency. Future research can focus on model iteration, multimodal integration and the improvement of application standardization to further optimize the role of large language models in the teaching field. This study can provide references for IELTS

DOI: 10.54254/2755-2721/2026.36348

teaching systems that integrate artificial intelligence, and point out directions for further in-depth exploration.

References

[1] Pilcher, N., & Richards, K. (2017). Challenging the power invested in the International English Language Testing System (IELTS): Why determining 'English' preparedness needs to be undertaken within the subject context. *Power and Education, 9*(1), 3–17.

[2] Sharif, Y., Fakhar, E., & Lodhi, M. A. (2023). Challenges faced by IELTS candidates during attaining reading proficiency and designing 'reading practice plan'. *Jahan-e-Tahqeeq*.

[3] Faja, S., & Rajkumar, P. (2025). Exploring the use of ChatGPT as a learning tool in programming education. *Journal of Instructional Pedagogies, 31*.

[4] Dashti, L., & Razmjoo, S. A. (2020). An examination of IELTS candidates' performances at different band scores of the speaking test: A quantitative and qualitative analysis. *Cogent Education, 7*(1), 1770936.

[5] Day, M. J., & Zhang, T. (2026). Leveraging generative AI for IELTS preparation: Student perspectives on language learning. *Education Sciences, 16*(5), 673.

[6] Effatpanah, F. (2019). Application of cognitive diagnostic models to the listening section of the International English Language Testing System (IELTS). *International Journal of Language Testing, 9*(1), 1–28.

[7] Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System, 121*, 103254.

[8] Nguyen, P., & Nguyen, T. (2025). Using ChatGPT to develop students' IELTS speaking skill in an IELTS course. In *VietTESOL International Convention Proceedings (Volume 5)*, 92–119.

[9] Rasheed, A. F., Chasib, S. A., Alabbas, H. N., Zarkoosh, M., & Abbas, S. F. (2024). IELTS writing tasks evaluation using LLMs: A comparative study of ChatGPT, Claude, Copilot, and Google Gemini. In *2024 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)*, 32–37.

[10] Dina, A. T., & Ciornei, S.-I. (2013). The advantages and disadvantages of computer assisted language learning and teaching for foreign languages. *Procedia-Social and Behavioral Sciences, 76*, 248–252.

[11] Xing, C., & Saeed, M. A. (2025). A systematic review on artificial intelligence (AI) technologies in ESL/EFL speaking skills. *International Journal of TESOL Studies, 8*(2), 240–270.

[12] Saraswati, G. P. D., Farida, A. N., & Yuliati, Y. (2023). Implementing AI Replika in higher education speaking classes: Benefits and challenges. *ELT Forum: Journal of English Language Teaching, 12*(3), 207–215.

[13] Annamalai, N., Uthayakumaran, A., Bervell, B., & Kumar, R. (2025). Examining the use of ChatGPT in argumentative writing: A mixed methods study. *Journal of Advanced Academics, 36*(3), 521–550.

[14] Zheldibayeva, R. (2025). GenAI as a learning buddy for non-English majors: Effects on listening and writing performance. *Educational Process: International Journal (EDUPIJ), 14*(1), 1–16.

[15] Panahi, A., & Mohebbi, H. (2022). Cognitive diagnostic assessment of IELTS listening: Providing feedback from its internal structure. *Language Teaching Research Quarterly, 29*, 147–160.

[16] Hopârtean, A. M. (2019). Self-evaluation: How to ease speaking anxiety in exams and improve performance. *Journal of Linguistic and Intercultural Education*, 65.

[17] Mandour, W. S. (2025). AI-powered tools for linguistic and sentiment analysis in language learning: Applications using Google Sheets and large language models. In *Using AI tools in text analysis, simplification, classification, and synthesis* (pp. 67–112). IGI Global Scientific Publishing.

[18] Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*(1), 28.

[19] Pudasaini, S., Miralles-Pechuán, L., Lillis, D., & Llorens Salvador, M. (2025). Survey on AI-generated plagiarism detection: The impact of large language models on academic integrity. *Journal of Academic Ethics, 23*(3), 1137–1170.

[20] Zhou, S. (2024). Facilitating English teaching with the use of ChatGPT: Can AI be more effective than human tutors? In *ICERI2024 Proceedings* (pp. 9139–9147). IATED.