

A Preliminary Exploration of Constructing a Human-in-the-Loop Teaching Model in English Language Testing Courses Empowered by AI

Yajie Shen

Xi'an Fanyi University, Xi'an, Shaanxi, China

Abstract: With the rapid advancement of Artificial Intelligence and Large Language Models in education, English language testing courses face a pressing need to transition from exam-oriented to competency-based approaches. This paper proposes and constructs the “AI-Prep-Activity-AI-Assess” teaching model, elucidating its theoretical foundation, three-stage operational framework, and role division in the classes: Pre-class (AI-Prep) focuses on intelligent diagnosis and personalized guided learning; In-class (Activity) centers on task-driven human-machine collaborative interaction; Post-class (AI-Assess) employs adaptive assessment as its primary means. Through literature review, model design, and case analysis, the study finds that the 3A model effectively enhances students’ practical abilities and learning motivation, supports teachers’ transition from knowledge disseminators to instructional organizers and emotional guides, and demonstrates significant advantages in personalized feedback, formative assessment, and closed-loop teaching. Concurrently, this paper identifies limitations in large-scale empirical validation, tool adaptability, data privacy, and academic integrity, which warrant further research.

Keywords: Human-in-the-Loop; Artificial Intelligence; Teaching Model

1. Introduction

The rapid advancement of Artificial Intelligence (AI) is profoundly transforming societal production and lifestyles, and is gradually permeating the field of education. As education becomes increasingly digitized and intelligent, AI not only provides technological support for the innovation of teaching methods and learning models but also offers new possibilities for

learners’ autonomous development and promotes educational equity. In foreign language education, particularly in English language testing courses, the integration of AI presents both new opportunities and challenges. Therefore, exploring AI-empowered Human-in-the-Loop (HITL) teaching models aligns with the contemporary demands of educational development while addressing practical issues in curriculum reform.

In recent years, the application of AI in education has become increasingly widespread. It has demonstrated significant advantages in areas such as resource delivery, learning analytics, and intelligent assessment, driving education toward greater personalization and intelligence. The application of AI is particularly prominent in the fields of English education and language testing. Some scholars utilize the Coh-Metrix text analysis tool to diagnose students’ writing proficiency through metrics such as cohesion and syntactic complexity (Xu Luyan, 2021)^[9]. This paper demonstrates that the introduction of AI not only shifts course evaluation from summative to formative and developmental approaches but also unlocks new possibilities in teaching organization, learning support, and competency cultivation. However, current English language testing courses still face significant challenges: courses prioritize test-taking skills enhancement, with content heavily focused on question types and strategies, leading to dull classes and low student engagement. And instructors struggle to accommodate individual differences, making personalized guidance and assessment difficult to implement. There exists a disconnect between the transformation of knowledge and skills, with taught content failing to translate into actionable test-taking abilities for students. Against this background, constructing an AI-empowered HITL teaching model has become an essential exploration for advancing the reform of English

language testing courses.

This paper focuses on constructing an AI-empowered HITL teaching model, aiming to provide new insights and pathways for reforming English language testing courses. Academically, this paper contributes to expanding the application scenarios of AI in language teaching and enriching interdisciplinary research outcomes between educational technology and foreign language education. It also offers theoretical references for building a competency-oriented curriculum system. Practically, this paper offers valuable insights for reforming English language testing courses at universities. The HITL model alleviates teachers' workload while delivering personalized learning support and intelligent feedback to students, thereby enhancing their readiness for language tests and professional certification exams. Crucially, this paper facilitates a shift from exam-oriented to competency-based curricula, strengthening the practicality and relevance of English instruction to support talent development.

2. Literature Review

The development of AI has opened up new possibilities for the reform of English language testing courses. In recent years, the emergence of Artificial Intelligence Generated Content (AIGC) and Large Language Models (LLMs) has not only advanced intelligent applications on a technological level but also provided theoretical and practical support for innovations in teaching models and personalized learning. Based on this, this paper first outlines the development of AI and LLMs and then reviews the application of HITL teaching models in foreign language education so as to establish the theoretical and practical foundation.

2.1 The Development of Artificial Intelligence and Large Language Models

The term AI was first introduced by John McCarthy at a workshop held at Dartmouth College in 1956. Since then, the field of AI has experienced several peaks and valleys. In the past decade, with the rapid advancements in deep learning, big data, and computational power, AI has entered a phase of rapid development. AIGC has gradually become a research and application hotspot, thus attracting widespread attention from both academia and industry. In 2020, Brown et al. proposed GPT-3,

which defined the concept of "Large Language Models" from three dimensions: model scale, capability boundaries, and training paradigms^[1]. By the end of 2022, ChatGPT, developed based on LLMs, was launched and quickly gained widespread attention. Its expanding model scale, multimodal integration capabilities, and global user coverage have made it an intelligent agent with powerful general language understanding and generation capabilities.

2.2 Application Exploration of Human-in-the-Loop Teaching Models

In the field of education, AI is driving the development of personalized learning. Adaptive learning systems can dynamically adjust learning content and paths based on students' learning states and knowledge mastery, providing customized learning plans. Meanwhile, the generative capabilities of LLMs make them a new type of teaching tool, not only serving as intelligent interactive interfaces for real-time feedback but also generating diverse learning materials to enrich the curriculum. AI technologies have already been applied in several mature ways in English education. For instance, Duolingo's Speaking Challenge uses ASR technology to assess pronunciation fluency and help users correct connected speech errors (Jurafsky et al., 2020)^[4]. The Coh-Metrix text analysis tool can conduct readability analysis and provide teachers with recommendations for text leveling (Graesser et al., 2011)^[3].

Teaching models are a crucial part of course design, and AI, along with LLMs, is driving profound changes in foreign language class practices. Yang Zongkai, Director of the National Digital Learning Engineering Technology Research Center of China, pointed out that the education model is shifting "from the traditional 'teacher-student' binary structure to the 'teacher-student-machine' trinary structure" (Zhang Rong, 2023)^[10]. Xu Jiajin (2025) discussed the mediating role of LLMs in personalized foreign language education from two dimensions: at the subject level, the model can generate course outlines and teaching materials tailored to students' cognitive levels, interests, and learning stages; at the interpersonal level, the model can assist teachers in designing cross-cultural teaching plans suited to different interaction scenarios, while it maintains interpersonal and cultural sensitivity^[8].

Chinese universities have also conducted

practical explorations in this field. Shanghai Jiao Tong University launched the “AI Teacher + HI Mentor” classroom reform program, where AI handles foundational teaching, allowing teachers to focus on emotional support and advanced knowledge guidance, thus restructuring role division in classes. Beijing University of Posts and Telecommunications developed the “Code-On” LLM-empowered platform, which, centered on teaching AI agents, provides personalized tutoring for students while offering multidimensional teaching management support to instructors, thereby facilitating the digital transformation of education. Research by several academic teams has also provided empirical evidence for the effectiveness of the HITL teaching model. For example, Chen Wanling’s team (2025) found through a comparison between experimental and control groups that the HITL model not only enhanced students’ English application abilities but also promoted the integration of learning and practice, as well as the realization of holistic education goals^[2]. Wu Jianhao’s team (2024) conducted empirical research with 100 English learners and demonstrated that AIGC could offer more authentic interactions for speaking practice, thus improving language expression skills^[7]. Kong Lei (2024) proposed the “AI-assisted + non-AI-assisted alternating teaching model” and used the LLM-Claude as a practical case, showing that it could not only improve text quality in writing instruction but also provide detailed explanations for revisions^[5].

In summary, existing research shows that the HITL teaching model can enhance learning outcomes, promote personalized teaching, and provide supportive assistance to teachers in foreign language education. Based on these findings, this paper aims to explore the potential application of this model in English language testing courses, offering both theoretical and practical references for curriculum reform.

3. The Shift in Teachers’ and Students’ Mindsets in the Human-in-the-Loop Teaching Model

As AI technologies, especially LLMs, are increasingly applied in foreign language education, the field is evolving towards a HITL model. Therefore, both university English teachers and students need to reflect on traditional mindsets and proactively adapt to the new AI-empowered teaching paradigm in order

to embrace the profound transformation in educational models and learning methods. As Yu Shengquan, Executive Director of the Future Education High-End Innovation Center at Beijing Normal University, aptly stated, “AI will not replace teachers, but teachers who use AI will replace those who do not” (Su Jin, 2018)^[6]. This highlights the urgency of mindset transformation for both teachers and students.

3.1 Shift in Teachers’ Mindsets

In traditional teaching, the teacher’s role is predominantly centered on knowledge transmission, with students being passive recipients. Classroom interaction and autonomous learning are often not fully realized. Even with abundant resources and technological tools, teachers may struggle to harness the full potential of technology due to outdated mindsets or skills.

In the HITL model, the teacher’s role requires a fundamental shift. Teachers are no longer just knowledge transmitters but also curators of learning content, organizers of teaching activities, facilitators and motivators of classes, as well as trackers and assessors of the learning process. With AI technology, teachers can integrate high-quality materials, design project-based or inquiry-driven tasks, provide personalized feedback, and conduct dynamic assessments, fostering the collaborative development of students’ language and cognitive abilities.

Moreover, teachers should also become innovators in integrating educational technology and guardians of emotional values, creatively applying technology in teaching practice while focusing on students’ emotional interactions, value guidance, and humanistic care. Only through continuous learning and reflection can teachers reposition their roles in the AI era, transforming from “knowledge providers” to “educators”.

3.2 Shift in Students’ Mindsets

In an AI-empowered teaching environment, the role of foreign language students is being reshaped. Compared to teachers, students typically have a stronger digital sensitivity and adaptability to technology, allowing them to quickly master tools such as LLMs. However, the real challenge lies not in the technical operations but in how to stimulate learning motivation, clarify their professional positioning, and set learning goals that align with their

personal development (Xu Jiajin, 2025)^[8].

The ultimate responsibility for personalized learning lies with the students themselves, which requires strong self-regulated learning (SRL) capabilities. Only by actively setting goals, monitoring progress, and reflecting and adjusting in a timely manner can students transform the resources provided by AI into intrinsic motivation and external skill enhancement.

In this process, teachers need to play a guiding and supportive role, helping students clarify their goals, optimize learning strategies, and prevent them from falling into technological dependence or learning deviations. Through the collaborative interaction of “teacher-student-machine”, students can gradually transition from passive receivers to active constructors, thus developing more personalized and practically valuable professional competencies.

4. The Construction and Application of the 3A Teaching Model Empowered by AI

This chapter builds upon the theoretical groundwork discussed in Chapter 3 regarding the shift in teachers’ and students’ mindsets, proposing and defining the core elements and implementation framework of the “AI-Prep-Activity-AI-Assess” (3A) teaching model. On this basis, practical case studies will be discussed to explore the application of this model in English language testing courses, aiming to form an operable teaching loop.

4.1 The Connotation and Framework of the 3A Teaching Model

This section elaborates on the background, application goals, and overall framework of the “3A teaching model”, clarifying the division of roles and operational mechanisms for both teachers and students in the “pre-class + in-class + post-class” loop, providing the foundation for the subsequent implementation of the model.

4.1.1 Course Design and Goal Orientation

English language testing courses focus on competency development while also addressing the assessment requirements of exams like TEM-8 (Test For English Major-8) and CATTI (China Accreditation Test for Translators and Interpreters). These courses cover the five major competencies-listening, speaking, reading, writing, and translation-along with key skills such as cross-cultural awareness. The teaching process follows a progressive structure of

“pre-class diagnosis + training in classes + post-class assessment”, adjusting teaching difficulty and task types based on students’ learning data. This structure aims to promote knowledge input, skill training, and competency transfer, ensuring effective connections between different teaching stages.

4.1.2 Background and Application Goals of the 3A Model

English language testing courses currently face the dual demands of competency development and exam pass rates. The traditional “lecture-practice-test” model falls short in providing personalized support and formative feedback. With the development of AI tools such as LLMs, personalized guided learning, learning diagnostics, and adaptive assessments have become feasible. In this context, the 3A model (AI-Prep-Activity-AI-Assess) aims to enhance students’ engagement and participation without compromising exam objectives. It seeks to promote the transfer of knowledge and skills in real-world contexts, offering an innovative teaching model for professional education in English language testing courses.

4.1.3 Core Elements and Implementation Framework of the 3A Model

The 3A model consists of three stages: AI-Prep (pre-class), Activity (in-class), and AI-Assess (post-class), emphasizing the collaboration among “teacher-student-machine”. Figure 1 systematically illustrates the structure of the 3A model.

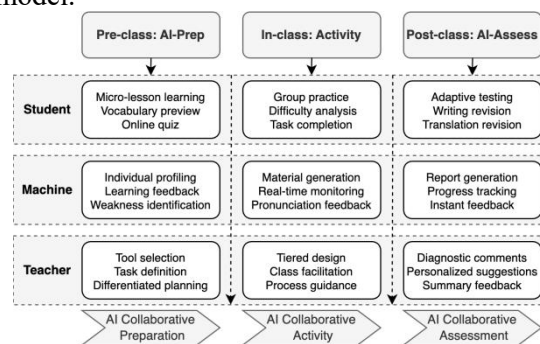


Figure 1. Structure of the 3A Model

In the AI-Prep stage, the focus is on intelligent diagnostics and guided learning. In this phase, teachers should select and evaluate AI tools that are suitable for preparing for exams like TEM-8, integrating them into the learning platform and determining task types to ensure alignment between technological tools and teaching objectives. Students complete foundational preparation through micro-lessons, vocabulary previewing and online quizzes. The learning

platform generates profiles of individual learners and provides feedback on students' weaknesses. Teachers then adjust class activities and prepare differentiated lessons accordingly.

In the Activity stage, the emphasis is on a task-based class with human-machine collaboration, highlighting interaction and instant feedback. Based on pre-class diagnostics, teachers design tiered tasks such as group oral simulations and reading comprehension exercises, utilizing AI-provided corpus, examples and real-time prompts to support students in completing these tasks. The machine's role in this phase is to monitor learning progress in real-time and provide prompts, such as ASR-based pronunciation corrections or LLM-generated example outputs. The teacher, meanwhile, organizes and guides the class, complementing the machine's role and facilitating the transfer of language knowledge to practical application.

Finally, in the AI-Assess stage, the focus shifts to precise assessment and extended practice. Students receive immediate feedback through adaptive tests, iterative writing improvements, and oral practice. The online learning platform generates periodic reports and highlights changes in competency. Teachers then provide diagnostic feedback and personalized learning recommendations based on these reports.

These three stages are interconnected through students' learning data, forming a closed-loop teaching system: AI is responsible for generating materials, issuing tasks, monitoring progress and analyzing learning data, while teachers handle goal setting, activity design and assessment. Students transition from passive recipients to active learners through continuous feedback.

4.2 Initial Exploration of ChatGPT's Application in the 3A Teaching Model

In the 3A teaching model, ChatGPT, as a representative tool of LLMs, holds significant

potential across all three stages: AI-Prep (pre-class), Activity (in-class) and AI-Assess (post-class). Its core advantage lies in its ability to provide immediate feedback and diverse resources through natural language interactions, thus enabling personalized support and formative assessment.

In the pre-class stage, ChatGPT can assist students in self-guided learning. For instance, when students encounter complex sentences or specialized vocabulary while reviewing study materials, they can input these into the model to receive definitions, example sentences and usage explanations, thereby reducing the difficulty of reading comprehension. Additionally, teachers can analyze students' interactions with ChatGPT through logs to identify areas of weakness, providing a foundation for differentiated teaching in the classes.

In the in-class stage, ChatGPT can serve as a tool for collaborative learning, supporting task-based and inquiry-based learning approaches. During group work, students can use it to generate example answers or reference translations, compare them with their own answers, and engage in peer review and critical discussions. Teachers can guide students to analyze the strengths and weaknesses of ChatGPT's outputs, enhancing their language judgment and critical thinking, while avoiding over-reliance on the technology.

In the post-class stage, ChatGPT's role becomes especially prominent, offering immediate feedback and personalized revision suggestions beyond class time. For example, in TEM-8 translation practice, students first complete their translation independently, then input their answers into ChatGPT to request feedback on issues related to grammar, vocabulary choice, or discourse coherence. The model can also generate specific revision suggestions tailored to TEM-8 exam requirements. Table 1 presents a complete translation revision case.

Table 1. Translation Revision Case Provided by ChatGPT

Translation Questions in TEM-8 Exam	The first sentence of the 2021 TEM-8 translation task conveys the meaning of "commemorating youth".
Student's Answer	Your youth is a journey, one that takes you further and further away from your childhood and adolescence.
Prompts	As a TEM-8 candidate, please analyze my Chinese-English translation from three dimensions: 1. Accuracy Critique - Identify semantic deviations, terminology misuse, and coherence gaps 2. Stylistic Assessment - Evaluate register appropriateness (academic/literary), syntactic complexity, and rhetorical devices

	3. TEM-8 Alignment - Check compliance with band descriptors for “distinction” level translations Provide: - 3 prioritized revision suggestions with concrete alternatives - Highlight where native speaker phrasing differs from my version - Preserve the original text structure unless compromising idiomaticity
ChatGPT Feedback	1. Accuracy Meaning is largely correct, but “journey” is too plain for “distant journey”. “Childhood and adolescence” is acceptable but not the closest rendering of “boyhood” or “early youth”. 2. Style Wording is a bit colloquial; lacks the literary tone expected at TEM-8 level. Expressions like further and further away could be elevated. 3. TEM-8 Alignment Meets accuracy, but refinement in diction and register needed for distinction level. Key Revision Suggestions: Replace journey → distant journey / long voyage. Refine childhood and adolescence → childhood and boyhood / early youth. Improve further and further away → ever farther away / receding into the distance.
Revised Version by ChatGPT	Your youth is a distant journey, one that carries you ever farther away from your childhood and boyhood.

The above case demonstrates the specific application of ChatGPT in the post-class phase. By setting prompts tailored to the TEM-8 exam, students can receive multidimensional feedback that balances accuracy, style, and exam standards. This allows them to not only identify shortcomings in their translations but also learn more authentic expressions through the suggested revisions. However, the design and use of prompts require students to have a certain level of technical literacy, such as being able to clearly articulate and accurately express their needs. Without this, the feedback may be superficial and fail to reach its full potential. Therefore, when guiding students to use ChatGPT, teachers should help them gradually master the basic methods of prompt construction to ensure the technology’s effectiveness in the learning process. Compared to the traditional exam preparation process, where students rely on one-way feedback from teachers, ChatGPT provides instant, personalized, and reusable learning support, which helps enhance students’ self-directed learning abilities. This model not only reduces the burden on teachers in large-class teaching environments but also creates a more open, interactive, and continuously optimized learning space for students.

5. Conclusion

Based on the preceding research, this chapter summarizes and reflects on the value, limitations, and future development directions of the

“AI-Prep-Activity-AI-Assess” (3A) teaching model. By synthesizing the conclusions, analyzing its limitations, and outlining suggestions for future studies, this chapter aims to present the academic significance and practical implications of this paper, offering references for subsequent research and application.

5.1 Major Findings

This paper focuses on the “AI-Prep-Activity-AI-Assess” (3A) teaching model, exploring its construction and application path in English language testing courses. This study finds that the 3A teaching model has positive significance at multiple levels. First, in terms of teaching effectiveness, the 3A model strengthens students’ practical abilities and exam skills, overcoming the limitations of traditional classrooms, which often prioritize theory over practice. Second, in terms of learning motivation, the use of intelligent diagnostics, personalized learning materials, and adaptive assessments significantly enhances students’ autonomy and interest in learning, which in turn improves learning efficiency. Third, from a career-oriented perspective, the 3A model aligns with the skill requirements of language testing and professional certification, emphasizing practical skills and problem-solving abilities, thus enhancing students’ career competitiveness. Furthermore, the model not only facilitates the transformation of the teachers’ role but also promotes teachers’ mastery of AI tools and data

analysis, thus enhancing their professional development. Overall, the 3A teaching model improves teaching effectiveness and learning motivation at the course level, while also showing strong practical value in educational reform, teachers' development, and talent cultivation.

5.2 Limitations

Although this study provides a preliminary exploration of the 3A teaching model's construction and application, there are still certain limitations. First, the study is primarily based on theoretical analysis and framework construction, lacking large-scale empirical data, which may affect the general applicability of the conclusions. Second, the study focuses on English language testing courses and has not been validated in a broader range of subjects or diverse learning contexts, and its cross-disciplinary applicability requires further testing. Third, there are still adaptation issues in selecting and using AI teaching tools. Different tools vary significantly in terms of functionality, stability, and user experience, requiring teachers to perform secondary screening and adjustment in practice. Lastly, the implementation of AI-powered teaching faces challenges such as data privacy, academic integrity, and the technological literacy of both teachers and students, all of which may impact the effectiveness of the model's deployment.

5.3 Suggestions for Future Study

Future studies can be developed in three main directions. First, empirical studies should be conducted on a larger scale and across diverse teaching contexts to further validate the applicability of the 3A teaching model in different types of courses and among students at different levels. Second, with the continuous advancement of artificial intelligence technologies, the operational process of the 3A model should continually be optimized in practice, and existing AI tools should be selected and adapted to enhance the model's effectiveness and operability. Lastly, research should focus on improving teachers' AI literacy and fostering students' self-regulation abilities, exploring the establishment of comprehensive support mechanisms to ensure the sustainable implementation and long-term effectiveness of

the HITL model in English language teaching.

References

- [1] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. *Advances in neural information processing systems*, 2020, 33: 1877-1901.
- [2] Chen Wanling, Wang Yangyang. Construction and Effectiveness of a Human-in-the-Loop Teaching Model in Vocational College English Courses: An AI-Empowered Perspective [J]. *Journal of Guangdong Industry Polytechnic University*, 2025, 24(01): 34-41.
- [3] Graesser A C, McNamara D S, Kulikowich J M. Coh-Metrix: Providing multilevel analyses of text characteristics[J]. *Educational researcher*, 2011, 40(5): 223-234.
- [4] Jurafsky D, Chai J, Schluter N, et al. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics[C]/Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020.
- [5] Kong Lei. Application of Generative Artificial Intelligence in Foreign Language Teaching: A Case Study of *College Critical English Course: Intensive Reading* [J]. *Frontiers of Foreign Language Education Research*, 2024, 7(01): 11-18+90.
- [6] Su Jin. What kind of teachers are needed in the era of artificial intelligence? [N]. *China Education Daily*. 17-05-2018.
- [7] Wu Jianhao, Zhou Wanting, Cao Chao. An Empirical Study of Generative Artificial Intelligence Technology Empowering Oral English Teaching [J]. *China Educational Technology*, 2024, (04): 105-111.
- [8] Xu Jiajin. Implementation Strategies for Personalized Foreign Language Education in the Context of Large Language Models [J]. *Foreign Language Teaching and Research*, 2025, 57(01): 81-91.
- [9] Xu Luyan. A Coh-Metrix-Based Study on Cohesion and Coherence in English Major Students' Expository Writing [D]. *Zhejiang Normal University*, 2021.
- [10] Zhang Rong. The Impact of Generative Artificial Intelligence Technology on Education: An Interview on ChatGPT [J]. *E-education Research*, 2023, 44(02): 5-14.