

Multimodal Interaction Design and Teaching Reform of AI-integrated Blended English Teaching in Higher Education

Jun Chen. Fuli Li

Faculty of Foreign Languages & Northern Jiangsu Development Research Institute, Huaiyin Institute of Technology, Huai'an 223001, Jiangsu Province, China
Email: 350379069@qq.com

How to cite this paper: Chen, J., & Li, F. (2025). Multimodal Interaction Design and Teaching Reform of AI-integrated Blended English Teaching in Higher Education. *International Journal of Educational Development*, 2(1), 50-56. ISSN Print: 3104-4263; ISSN Online: 3104-4271. <https://doi.org/10.63313/IJED.9034>

Published: 2025-12-29

Copyright © 2025 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

As AI becomes deeply integrated into education, college English teaching faces the challenge of overcoming monotonous interaction modes and insufficient personalization. Supported by multimodal teaching theory and blended learning theory, this paper explores the path of multimodal interaction design for AI-integrated blended English teaching in higher education, constructs an "AI + multimodal" interaction system from three dimensions of pre-class, in-class and post-class sessions, and proposes supporting teaching reform strategies. Teaching practice verifies that the proposed design effectively enhances students' engagement in English learning and their comprehensive application competence, thus providing practical reference for college English teaching reform.

Keywords

AI; College English; Blended teaching; Multimodal interaction; Teaching reform

1. Introduction

At present, blended English teaching has been widely implemented in colleges and universities, yet it still suffers from rigid interaction modes, inadequate utilization of multimodal resources and lack of personalized guidance. The rapid development of artificial intelligence technologies such as ChatGPT, virtual simulation and intelligent speech recognition brings advantages in interactivity, personalization and datatization, which creates possibilities to break through the predicaments of blended teaching and enrich multimodal interaction forms.

Multimodal interaction emphasizes constructing teaching interaction scenarios through diverse semiotic resources including language, images, audio, video and gestures, which aligns with the comprehensive nature of English language learning (Hu, 2007). The integration of artificial intelligence into multimodal interaction

design of college English blended teaching not only exerts the flexibility and resource abundance of online teaching but also highlights the advantages of emotional interaction and real-time feedback in offline teaching. Based on this, this paper focuses on the multimodal interaction design framework and teaching reform practice of AI-integrated blended English teaching in higher education, aiming to provide a new path for improving the quality of college English teaching.

2. Theoretical Foundations

2.1. Blended Learning Theory

Proposed by American scholars such as Garrison, blended learning theory centers on the organic integration of online autonomous learning and offline face-to-face teaching to materialize the student-centered teaching philosophy. It improves learning outcomes through the complementary advantages of the two teaching modes (Garrison, Anderson, & Archer, 2001). The theory emphasizes learners' active participation, collaborative communication and meaning construction in the teaching process, providing a core framework for the multimodal interaction design of AI-integrated blended college English teaching. Specifically, it builds a complete learning cycle by combining online AI-assisted multimodal learning with offline teacher-guided multimodal interaction.

2.2. Multimodal Teaching Theory

Derived from systemic functional linguistics, multimodal teaching theory holds that humans acquire information and construct meanings through multiple perceptual modalities including vision, hearing and touch. Kress & van Leeuwen (2001), the representative scholars in multimodal discourse analysis, argue that teaching processes should fully utilize diverse modal resources to stimulate learners' interest. In English teaching, multimodal instruction can be carried out through texts, audio, video, animation, physical demonstration and interactive games. Artificial intelligence technology enables precise delivery of multimodal resources, real-time feedback of interaction processes and personalized adjustment, thus providing technical support for the efficient implementation of multimodal teaching.

2.3. Theory of Artificial Intelligence Applications in Education

The theory of artificial intelligence applications in education emphasizes the auxiliary and empowering role of technology in education. It advocates optimizing the allocation of teaching resources, realizing precise regulation of teaching processes and facilitating learners' personalized development through functions such as intelligent perception, intelligent analysis and intelligent feedback (Huang, Yang, & Hu, 2020). In college English teaching, artificial intelligence technologies including speech recognition, natural language processing and big data analysis

enable intelligent upgrading of multimodal interaction processes, such as intelligent speech evaluation, personalized exercise delivery and learning behavior data analysis.

3. Multimodal Interaction Design of AI-Integrated Blended English Teaching in Higher Education

Combined with the dual online-offline scenario characteristics of blended teaching, this study takes multimodal interaction as the core and leverages artificial intelligence technology to construct a full-process multimodal interaction system covering pre-class, in-class and post-class stages. This system enables multimodal presentation of teaching resources, intelligent support for interaction processes and precise control over learning outcomes.

3.1. Pre-class: AI-driven Multimodal Resource Delivery and Preview Interaction

Pre-class learning centers on online autonomous study. Relying on intelligent learning platforms such as Chaoxing Learning and XuetangX, AI technology realizes personalized delivery of multimodal preview resources and real-time monitoring of preview effectiveness. Based on teaching objectives and students' English proficiency data, AI screens and pushes preview resources covering visual, auditory and other modalities (e.g., original English videos, audio-books, interactive exercises) and accurately matches learning tasks to address weaknesses in listening and vocabulary acquisition (Min, Fang, Zhou, & Pan, 2025). Meanwhile, the AI intelligent Q&A function allows students to ask questions via voice or text to obtain instant answers and extended resources. The system synchronously records preview data and generates personalized preview reports, laying a precise foundation for offline teaching.

3.2. In-class: AI-assisted Multimodal Interaction and Targeted Guidance

In-class teaching prioritizes offline face-to-face instruction, with AI enriching multimodal interaction forms. On one hand, intelligent speech recognition systems facilitate interactive activities such as follow-up reading and role-play, providing real-time pronunciation evaluation and visualized feedback. Virtual simulation platforms are employed to create intercultural communication scenarios, enhancing students' language application competence (Li, 2021)—a typical example being negotiation simulation in business English teaching. On the other hand, AI classroom analysis systems collect real-time data on students' facial expressions and interaction frequency, adopting multimodal fusion analysis to assess learning status. This helps teachers dynamically adjust teaching pace, strengthen key-point explanation and push personalized exercises to implement student-oriented

teaching.

3.3. Post-class: AI-enabled Multimodal Review and Personalized Feedback

Post-class learning focuses on consolidation and improvement. Integrating pre-class preview and in-class learning data, AI generates personalized review plans and pushes multimodal review resources including knowledge summary videos and English film clip appreciation (Zhang, & Li, 2024). For post-class assignments, AI writing assistance systems conduct multi-dimensional grading covering grammar correction and logical analysis, offering revision suggestions through multimodal forms such as text and voice. Speech evaluation systems comprehensively assess pronunciation and fluency of oral assignments and formulate targeted improvement plans. Additionally, an AI-assisted collaborative review platform is built to support students in establishing groups for multimodal collaborative activities like English speech recording and intercultural theme discussion. AI monitors the collaborative process in real time and records relevant data to support teachers' assessment.

4. Supporting Teaching Reform Strategies

To ensure the effective implementation of multimodal interaction design for AI-integrated blended English teaching in higher education, supporting teaching reforms must be advanced from four dimensions: teaching objectives, teaching content, teaching evaluation and teacher development.

4.1. Reconstructing Multimodal Teaching Objectives

Guided by core competencies, college English teaching objectives are reconstructed. On the basis of cultivating students' linguistic knowledge and skills, emphasis is laid on developing their multimodal interaction competence and innovative thinking. Objectives such as "mastering the application of multimodal English learning resources", "conducting intercultural communication through multimodal approaches" and "possessing the ability to conduct autonomous learning via AI tools" are incorporated into the teaching objective system, realizing the organic integration of knowledge, ability and literacy objectives.

4.2. Optimizing Multimodal Teaching Content

Combined with the characteristics of artificial intelligence technology and students' learning needs, the teaching content structure is optimized to establish a multimodal teaching content system. Firstly, textbook content is integrated with multimodal resources, incorporating original English films, documentaries and online interactive courses to enrich the presentation forms of teaching content. Secondly, practical thematic content such as intercultural communication, business English and technical English is introduced; multimodal interaction scenarios are designed to

enhance students' language application competence. Meanwhile, AI big data analysis functions are utilized to real-time monitor students' acceptance of and interest in teaching content, enabling dynamic adjustment of teaching content.

4.3. Establishing a Multimodal Teaching Evaluation System

Breaking away from the traditional single-score evaluation model, an "AI + manual" multimodal teaching evaluation system is established to integrate formative assessment and final assessment. Formative assessment mainly relies on AI systems to collect multimodal data including students' online learning records, in-class interaction data and post-class assignment completion, conducting objective quantitative evaluation. Final assessment adopts diverse forms such as written tests, oral tests and multimodal project presentations (e.g., English speech videos and intercultural communication simulation reports). In addition, student self-assessment and peer assessment are introduced. Through multimodal evaluation platforms, students can reflect on their own learning processes and outcomes and evaluate peers' collaborative performance, improving the comprehensiveness and objectivity of evaluation.

4.4. Enhancing Teachers' Multimodal Teaching Competence

Efforts are made to strengthen college English teachers' ability to apply artificial intelligence technology and their multimodal teaching competence. AI educational technology training, multimodal teaching workshops and teaching observation activities are organized to help teachers master the operation of AI tools including intelligent learning platforms, virtual simulation systems and speech recognition systems. Teachers are encouraged to conduct multimodal teaching research and reform practices, establish interdisciplinary teaching teams and cooperate with information technology teachers and educational technology experts to enhance their capacity for multimodal teaching design and implementation.

5. Analysis of Teaching Practice Effects

To verify the effectiveness of the proposed multimodal interaction design and teaching reform strategies, two sophomore classes of non-English majors from a university were selected as research subjects. The experimental class (45 students) adopted the AI-integrated multimodal interactive blended teaching mode, while the control class (45 students) used the traditional blended teaching mode, with the experiment lasting one semester. Data were collected through questionnaires, academic performance analysis and interviews to conduct a comprehensive evaluation of teaching effects.

5.1. Improved Learning Engagement

Online learning data collected by the AI system indicated that the average weekly

online learning duration of students in the experimental class (4.2 hours) was significantly longer than that in the control class (2.8 hours), with better performance in resource access frequency and interactive question accuracy. Questionnaire results showed that 86.7% of students in the experimental class regarded multimodal interactive teaching as interesting and vivid, which effectively stimulated their learning interest; 77.8% of them were willing to actively participate in in-class multimodal interactive activities, compared with only 53.3% in the control class.

5.2. Enhanced Language Competence

Semester-end performance analysis revealed that the average comprehensive English score of the experimental class (82.5 points) was significantly higher than that of the control class (75.3 points). The most notable improvements were seen in listening and speaking skills: the experimental class scored an average of 81.2 points and 79.8 points respectively, while the control class obtained 73.5 points and 72.1 points. Interviews showed that students generally agreed that AI-assisted multimodal interactive scenarios helped them better comprehend linguistic knowledge and improve listening comprehension and oral expression abilities.

5.3. Promoted Autonomous Learning Ability

Its findings demonstrated that 82.2% of students in the experimental class believed AI tools effectively facilitated autonomous learning and helped them master the application of multimodal learning resources; 75.6% reported adjusting their study plans proactively based on AI-generated learning reports. In contrast, only 57.8% of students in the control class perceived themselves as having strong autonomous learning ability, indicating that AI-integrated multimodal interactive teaching can effectively enhance students' autonomous learning capacity.

6. Conclusions and Prospects

This study explores the multimodal interaction design path of AI-integrated blended English teaching in higher education, constructs a full-process multimodal interaction system covering pre-class, in-class and post-class stages, and proposes supporting teaching reform strategies. Teaching practice verifies that this design effectively improves students' learning engagement, comprehensive English application competence and autonomous learning ability, providing a feasible practical scheme for college English teaching reform. However, the study has certain limitations, such as a short experimental period and limited sample size. Future research can expand the sample scale and extend the experiment duration to further verify the long-term effectiveness of the teaching mode. Meanwhile, exploration may be conducted on the application of different types of AI technologies (e.g., generative AI) in multimodal interactive teaching to enrich interaction forms and teaching content.

Acknowledgements

This work was supported by the General Project of Philosophy and Social Sciences Research in Colleges and Universities of Jiangsu Province: A Study on Blended C-E Translation Teaching in College English Based on the Pigai.org Platform (Project No.: 2022SJYB1936).

References

- [1] Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, 15(1), 7-23.
- [2] Huang, R. H., Yang, J. F., & Hu, Y. B. (2020). The three realms of smart education: From environmental intelligence to ecological intelligence. *Educational Research*, 41(7), 54-63.
- [3] Hu, Z. L. (2007). Multimodalization in social semiotics. *Language Teaching and Linguistic Studies*, (1), 1-10.
- [4] Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.
- [5] Li, L. L. (2021). The application of virtual simulation technology in college English intercultural communication teaching. *Computer-Assisted Foreign Language Education*, (3), 89-95.
- [6] Min, S. C., Fang, F. M., Zhou, S. B., & Pan, X. Y. (2025). A longitudinal study on the effects of GenAI-empowered personalized adaptive English learning. *Foreign Language World*, 5, 42-49.
- [7] Zhang, Y., & Li, M. (2024). The application of generative AI in college English multimodal teaching. *Computer-Assisted Foreign Language Education*, (4), 45-52.