Applied Science and Innovative Research
ISSN 2474-4972 (Print) ISSN 2474-4980 (Online)
Vol. 9, No. 4, 2025
www.scholink.org/ojs/index.php/asir

Original Paper

An Exploration of AI-Driven Multimodal Teaching Model Construction for College English Reading Instruction

Yifan Peng

Wuhan College, Wuhan, Hubei, 430212, China

E-mail: 12281254 @ qq.com

Received: September 9, 2025 Accepted: October 1, 2025 Online Published: October 4, 2025

doi:10.22158/asir.v9n4p1 URL: http://doi.org/10.22158/asir.v9n4p1

Abstract

Current college English reading instruction is constrained by traditional single-modal approaches, making it hard to meet students' learning needs in the digital age. The advancement of artificial intelligence (AI) technology has injected new vitality into education, and its integration with multimodal teaching—by incorporating resources like images, audios, and videos—shows great potential to break free from the text-centric paradigm and create a more multidimensional language input environment. Multimodal reading instruction enhances language comprehension, cognitive engagement, and student learning interest. AI integration facilitates individualized learning, shifting teaching from "one-size-fits-all" to "precision-oriented", which is key to college English education reform.

Keywords

Artificial Intelligence, Multimodal Teaching, College English, Reading Instruction, Teaching Model

1. Introduction

For a long time, English reading instruction in Chinese universities and colleges focused mainly on the interpretation of texts, neglecting the essential part of multimodal resources in the language acquisition process. The development of educational informatization is accelerating, while artificial intelligence technology is gradually entering language teaching, providing the technical support for multimodal theory applications. Multimodal reading approaches intentionally combine visual, auditory, and other symbolic resources to help strengthen students' connections between language and context, thus developing cross-cultural cognitive abilities. Artificial intelligence can further extend the potential of multimodal data through resource mobilization and tracking the process of multimodal integration to lead to optimized teaching strategies. Education and reading instruction in English is being powered by

technology and is moving toward a more human-centered and engaging practice.

2. Theoretical Foundations of Artificial Intelligence and Multimodal English Reading Construction

2.1 Multimodal Discourse Analysis Theory

The theory of multimodal discourse analysis is grounded by a significant understanding of meaning-making in the context of social semiotics. At the root of this theory is the acknowledgment that humans make sense of their communication through symbols, and that human communication does not rely only on linguistic symbols. This theory provides a systematic, multi-dimensional approach to describe how different symbolic resources—(images, sounds, layout, color, etc. and even movement)—work to together to create composite meaning. In the context of college English reading intervention, a strong example of this would be a high quality PowerPoint teaching presentation. It is not just a long list of text, but incorporates key phrases, images that illustrate the points it is explaining, background music that sets the context, or animations to add excitement to the presentation. All of these different modal elements are not independent, but rather they are in a complex relationship to one another in which they complement each other, and/or reinforce each other. For example, in an English documentary on Western architecture, the narration conveys factual information and historical context, while the visual elements and background audio communicate the spatial aesthetics and the ambiance of the corresponding era (Chen, Harrison, Stevens, et al., 2024). These modal elements work together to take students beyond the literal text, and experience the cultural aesthetic implicit to the piece. A particularly useful analytical framework within this theory is the visual grammar analytical framework that teachers can use to more systematically analyze visual information. It guides both teachers and students to examine the emotional attitudes and cultural values embedded within aspects like image composition, character perspectives, or color usage. Multimodal discourse analysis theory fundamentally drives the transformation of English reading teaching paradigms, shifting the focus of instructional design from pure textual analysis to the comprehensive interpretation and application of diverse symbolic systems.

2.2 Theory of Artificial Intelligence Applications in Education

The theoretical framework for artificial intelligence applications in education is founded at the intersection of several fields, such as computer science, cognitive psychology, and pedagogy, and it provides a lens on technology as a functional analogue to particular roles carried out by human teachers in the process of studying or learning. The theory is based on understanding how technology can detect learners' unique characteristics and respond with customized instructional content. Presently, many educational AI systems depend on the analysis of big data to interpret behavioral data that educators would ordinarily collect as students read. For example, some systems can automatically identify how long students hover over specific types of text and how often they reread portions of text. Advances in natural language processing of text allow machines to have at least a preliminary understanding of

open-ended questions that students present to the machine and automatically retrieve relevant answers from inherited knowledge bases. Additionally, some machine learning models enable systems to leverage AI's pattern recognition ability to filter teaching materials—from large educational libraries—based on how well they align with individual feedback and specific learning goals.

2.3 Second Language Acquisition Theory

The theory of second language acquisition investigates the cognitive processes and social and environmental factors that shape human learning to communicate in a second language. Its key concepts emphasize that comprehensible language input is essential to the developing of language proficiency. Most researchers agree that learners need to be exposed to language a fraction above their current level of competence, with the understanding that adjustments during social interaction, assists in the absorption of new information. The filter hypothesis suggests that psychological factors like anxiety, or lack of confidence, may negatively interfere with the learners' effective use of input. The Interaction Hypothesis emphasizes that social discussion of language allows for negotiation of linguistic forms, and that negotiating meaning promotes deeper language processing mechanisms. The Output Hypothesis underlines the importance of language output for language acquisition and development, suggesting learners must put their hypotheses to test through form and thus receive feedback. These theories provide a rationale for the design of multimodal reading tasks. Teachers might combine visual and auditory materials to create environments rich in language, which helps students offload cognitive load and build understanding through effective interaction between modalities (María, Martínez, & Lirola, 2016).

3. Current Status of Multimodal Reading Construction in College English and Key Challenges of AI Application

3.1 Insufficient Integration of Multimodal Teaching Resources

At present, multimodal reading instruction in college English teaching is challenged by the absence of coherent ties among the various teaching materials. When preparing for lessons, instructors often need to gather materials separately from whatever external online platforms they find helpful, from self-made courseware, or from textbooks they are accustomed to using. These resources are often lacking in coherent ties among different multimodal resources, adopt different formatting standards, and follow widely different design principles. For example, video clips may lack textual explanations for viewers, and audio recordings may not have images or visual representations. Many instructors rely on their past personal experience to piece together resources in a rudimentary format, thus engaging students in ways that may remain superficially layered. Resource quality will always vary too, as instructors may use online video recordings not edited to remove incorrect language expressions or culturally inappropriate content. Instructors' limited out-of-class time limits their ability to analyze and assess the appropriateness of available resources in relation to the content and students' reading processes. This fragmented resource use is without synergy. Students would receive fragmented and

random multimodal inputs that do not support their reading process holistically.

3.2 Limited Depth in AI Application

The existing use of artificial intelligence technology in multimodal English reading instruction has not yet moved beyond superficial assistance. Several teaching platforms have implemented intelligent tools that focus primarily on the basic functions of automatically grading multiple-choice questions or providing demonstrations of word pronunciation. These tools had limitations in the evaluation of multimodal texts. For example, automated systems may have difficulty determining how images relate to text in the understanding of deep meaning, and they do not adequately identify how cultural background elements in videos affect detaining and understanding language. Intelligent, recommendation algorithms often rely on keyword matching, hence never actually understand students' cognitive challenges and students' interest preferences. Technology vendors who offer solution tools often lack sufficient understanding of principles of teaching an English subject, which exacerbate this lack of knowledge and understanding. The teachers who use these systems report that generated reading reports sometimes provide outcomes, but reading reports present very low level of information, at best providing basic statistics that do not inform precise instructional decision-making. Artificial intelligence technology has not yet developed a deep integration with multimodal teaching theory and principles, and existing applications of technology in reading instruction do not deliver the value potential of technology. This means that the technology itself enables personalized cognitive and meaning construction—but this value has not been realized in current practices.

3.3 Teachers' Inadequate Technical Application Skills

A number of English language teachers exhibit a level of deficiency in technological application when faced with intelligent teaching systems. The training system established in teacher education programs has not satisfactorily addressed operational training in intelligent education technologies; therefore, novice teachers do not have the technical preparation they require. Professional development for teachers is often limited to a basic introduction to the functions of classroom instruction tools, with little or no focus on the organic integration of technology and multimodal reading instruction. When teachers attempt to undertake multimodal reading tasks using intelligent platforms, they run into practical issues, such as operational complexity and complex, nested tool functionalities. As a result of their experience and seniority in their careers, older teachers expressed a strong aversion to new technologies and firmly favored familiar traditional approaches to teaching. Technical support services in the college context relate primarily to equipment maintenance and not instructional application aspects of intelligent teaching components. The demands of work-related time and energy limits the opportunity for teachers to explore technologies, and, as a result, many of the teachers note, as a matter of course, that they are unable to use a lot of the intelligent aspects of teaching (Paudel, 2025).

3.4 Monolithic Teaching Evaluation Methods

Currently, the assessment system of multimodal reading instruction in English courses for colleges is still dominantly uniform. Most schools still use final written exam grades to measure students' reading

ability; this traditional method does not fully represent students' performance in multimodal reading tasks. When teachers create assessment activities, they often limit themselves to formats that are easy to grade, such as multiple-choice questions or short answer questions. They do not adequately assess students' understandings of modal features, including images, audios, and videos. Existing assessment tools largely do not assess students' critical thinking and cultural interpretation that students displayed while engaging in multimodal reading. Learning Analytics, generated by intelligent teaching systems, typically only measure higher-order actions (i.e., completion rates and accuracy rates) rather than capturing the learning process and pathway to deepen understanding. Assessment results are often just scores or grades for students without providing specific analysis to guide students to enhance ways to improve learning.

4. Innovative Approaches to AI-Driven Multimodal Reading Teaching Model

4.1 Construction of Intelligent Multimodal Teaching Resource Repositories

Building an intelligent multimodal teaching resource repository requires the systematic integration of dispersed high-quality teaching materials. The repository development team should establish a clear resource classification system and metadata standards based on college English curriculum standards and practical teaching needs. Technical personnel can utilize intelligent crawling tools to obtain certified text, image, audio, and video resources from authoritative educational platforms, while implementing rigorous academic review mechanisms to ensure content quality. Each resource should be accompanied by comprehensive metadata descriptions, including difficulty ratings, subject tags, linguistic characteristics, and intermodal resource associations. Intelligent annotation systems can automatically generate keyframe markers for video resources and add text transcripts to audio materials, significantly enhancing resource discoverability. The repository must feature a user-friendly teacher upload interface to encourage frontline educators to share their self-created multimedia lesson plans and courseware. Intelligent recommendation algorithms can automatically push tailored resource combinations based on teachers' instructional objectives and student proficiency levels. For example, they can match specific lesson topics with relevant documentary clips, news recordings, and cultural background image galleries. The repository should establish a dynamic update mechanism that continuously optimizes resource quality and relevance through usage data analysis, forming a continuously evolving multimodal teaching resource ecosystem (Xu & Li, 2022).

4.2 Design of Personalized Reading Learning Pathways

The personalized reading learning paths' design should be based on students' actual cognitive characteristics and reading behaviors to develop a flexible and responsive learning process. Teachers can use intelligent analysis tools to help make sense of each student's sensitivity to different variations of learning information (for example, some students are able to pick out key points rapidly from videos while other students are strong at learning from text analysis). The system then uses those characteristics to develop customized differentiated reading task sequences for each student that offers

combinations of disparate materials including texts, images, as well as audio narration. In addition, the learning path adjusts the tasks' difficulty liquidity ratios in real-time while the students are learning. If the system detects that the student is struggling with a cultural concept, it will append graphics and videos relevant to the concept. The material and path design focus on student interests first, and offer other thematic materials in addition to reading tasks. This strategy ensures that each student receives genre- or text-specific reading instruction aligned with their cognitive profile, maintains momentum, while respecting the individual student.

4.3 Development of Intelligent Reading Assistance Tools

The advancement of intelligent reading support tools should be based on the real needs of multimodal reading processes, with a focus on developing practical functions that support students in comprehending challenging content. Development teams may consider building intelligent texts that parse tools that can automatically identify the complex sentences in the text and visually represent the syntactic relationships. To promote comprehension of multimodal resources, support tools would include cross-modal queries, whereby a student could pull relevant image or video explanations by inputting text-based keywords. The speech interaction modules would contribute by offering real-time assessments of accuracy in pronunciation and could immediately correct or highlight areas for students to consider while reading aloud the content integrated into the lesson. Enrichment around cultural contexts, which may apply to various settings, could be automatically added through illustrated annotations to concepts or words related to the reading materials, providing cultural contextual explanations to the student. All of these technology tools need to be integrated for students to receive coherent intelligent support, without switching applications whilst reading. The designs should also consider the ability of students to operate the technology, ensuring it is as intuitive as possible while also allowing for tiered assistance to meet various levels of technology literacy. It is also important for English language teachers to be involved in testing during the development and revision stages, providing feedback on the functionality during use, and providing feedback on the overall user experience. The final integrated intelligent reading support suite of tools should effectively lower the cognitive load for students and increase processing efficiency for multimodal information (Yuan & Zhu, 2020).

4.4 Establishment of Diversified Teaching Evaluation Systems

Establishing a diversified teaching evaluation system requires breaking through the limitations of traditional single-score assessments and constructing a multidimensional evaluation framework that covers the entire reading process. Intelligent systems can automatically collect students' process-oriented data during multimodal reading, including dwell time on different modal materials, interaction frequency, and depth of comprehension. Teachers should design diverse assessment tasks, such as text-image association analysis, video content interpretation, and cross-modal conversion exercises, to comprehensively evaluate students' multimodal reading abilities. Intelligent analysis tools can perform semantic analysis on students' reading notes to evaluate their information integration and

critical thinking abilities. The system should incorporate peer review mechanisms, enabling students to evaluate and provide feedback on each other's multimodal reading outcomes. The evaluation system should include a self-assessment module, guiding learners to reflect on their reading strategies and progress. Teachers generate comprehensive evaluation reports using multidimensional data provided by intelligent platforms, incorporating both quantitative metrics and qualitative descriptions. This evaluation approach holistically reflects students' reading competency development in multimodal environments, offering richer reference points for instructional refinement.

5. Implementation Recommendations for AI-Driven Multimodal Teaching Model Construction

5.1 Enhancement of Teachers' AI Instructional Competencies

A tiered training program is to be developed with the aim of helping teachers progressively master the skills for using smart teaching tools. Entry-level training can be organized through workshops demonstrating the basic functions of smart platforms, enabling teachers to become familiar with methods for accessing and integrating multimodal resources. Intermediate courses should guide teachers in using learning analytics data to identify students' reading difficulties and adjust teaching strategies accordingly. Colleagues can be provided with one-on-one coaching by tech-savvy teachers through a digital mentor system established by colleges, to solve practical problems in smart teaching. A smart teaching resource repository containing typical case studies and ready-to-use multimodal course design solutions should be developed by education departments. Regular smart teaching demonstration events, featuring teacher-led showcases of successful practices and organized professional discussions, should be hosted. Training content should emphasize authentic teaching scenario replication to help teachers learn—through hands-on experience—how technology enhances multimodal teaching effectiveness. Institutionalization of these measures is necessary to ensure teachers continuously update their smart education skills and adapt to the rapidly evolving educational technology landscape.

5.2 Development and Improvement of Intelligent Teaching Platforms

The development of an intelligent teaching platform should be guided by practical instructional needs, adopting a modular design philosophy to progressively refine its functional framework. The technical development team must collaborate with English instructors to design the platform architecture, ensuring each functional module aligns with the characteristics of multimodal reading instruction. The platform's core modules should include an intelligent resource repository management component, enabling teachers to rapidly retrieve and integrate diverse multimedia teaching materials. The learning analytics module must aggregate student reading behavior data to generate visual learning progress reports and competency maps. Interactive functional modules should provide real-time collaboration tools, enabling students to annotate, share, and engage in discussions during the reading process. Platform interfaces must maintain openness and scalability to facilitate future integration of new intelligent tools and external resources. System operations teams should establish a regular update

mechanism, continuously optimizing platform performance and user experience based on user feedback. Platform design must fully consider the network environments and hardware conditions of different institutions, delivering adaptable and stable solutions (Yumin, 2022).

5.3 Exploration and Practice of Novel Teaching Models

Exploring new teaching models requires teaching teams to conduct teaching experiments and innovative practices based on intelligent technologies. Educators can experiment with blended learning approaches, seamlessly integrating personalized learning from online intelligent platforms with interactive discussions in physical classrooms. Teaching teams should design project-based learning activities, guiding students to utilize multimodal resources for various inquiry tasks. The flipped classroom model enables students to independently master foundational knowledge through intelligent tools before class, allowing classroom time to focus on in-depth discussions and applications of multimodal reading. Teachers can organize collaborative reading activities, leveraging the grouping features of intelligent platforms to facilitate joint interpretation of multimodal texts. These teaching models need ongoing refinement based on actual instructional outcomes to develop student-body-specific teaching plans. Regular teaching research mechanisms should be set up by teaching teams for the periodic exchange of experiences and improvement strategies related to new teaching model implementation.

5.4 Reform and Innovation of Teaching Evaluation Mechanisms

Reforming teaching evaluation mechanisms requires establishing a multidimensional assessment system supported by intelligent technology. Educational institutions should develop digital platforms integrating multiple evaluation tools to automatically collect behavioral data and performance outcomes from students during multimodal reading processes. Teachers can design comprehensive assessment tasks encompassing skills such as text analysis, image interpretation, and cross-modal conversion to holistically measure students' reading literacy. Intelligent systems must incorporate self-assessment and peer-review functionalities to cultivate learners' metacognitive abilities and critical thinking. Evaluation outcomes should generate visual developmental trajectory charts, dynamically illustrating trends in students' evolving reading competencies. Based on diagnostic reports generated by intelligent systems, educators can promptly adjust instructional strategies and deliver personalized guidance. Process-oriented and developmental assessment is emphasized by this evaluation mechanism, effectively overcoming the limitations of traditional single-score evaluations.

6. Conclusion

The integration of artificial intelligence and multimodal theory has opened new pathways for college English reading construction. Its value lies not only in technological innovation but also in the reconstruction of pedagogical concepts. The deep integration of multimodal resources and intelligent technologies can transcend the temporal and spatial constraints of traditional classrooms, fostering a dynamic and open learning ecosystem. Future teaching must prioritize the appropriateness and

humanistic dimensions of technology application, avoiding the pitfall of technocentrism. The transformation of the teacher's role, the optimization of assessment systems, and the intelligent development of learning environments will become key drivers for advancing English reading instruction.

References

Chen, Y. H., Harrison, S., Stevens, M. P., et al. (2024). Developing a multimodal corpus of L2 academic English from an English medium of instruction university in China. *Corpora*, *19*(1), 5. https://doi.org/10.3366/cor.2024.0295

María, Martínez, & Lirola. (2016). Multimodal teaching proposal in an English as a foreign language subject. *ASIAN TEF: Journal of Language Teaching and Applied Linguistics*, *1*(1), 15. https://doi.org/10.21462/asiantefl.v1i1.5

Paudel, P. (2025). Use of multimodal glosses in teaching English vocabulary for non-English specialised undergraduates in public university in Nepal. *Frontiers in Education*, 2025. https://doi.org/10.3389/feduc.2025.1443803

Xu, J., & Li, T. (2022). Application of multimodal NLP instruction combined with speech recognition in oral English practice. *Mobile Information Systems*, *2022*, 6. https://doi.org/10.1155/2022/2262696

Yuan, T., & Zhu, X. (2020). *Multiliteracies in the English reading classroom of senior high school in China*. London: Francis Academic Press.

Yumin, C. (2022). *Interpersonal Meaning in Textbooks for Teaching English as a Foreign Language in China: A Multimodal Approach*. London: Bloomsbury Publishing.