DOI: 10.53469/jerp.2026.08(01).04

Research on Multimodal and AI-Integrated Primary School English Teaching

Wenfeng Zhu

School of Foreign Languages, China Three Gorges University, Yichang, Hubei, China

Abstract: ***Since the implementation of the "double reduction" policy, how to enhance the effectiveness of primary school English teaching has become a key concern for many primary school English teachers. Under the new curriculum standards, the limitations of the traditional teaching model have become increasingly prominent, and teachers urgently need to explore new teaching paths to promote classroom transformation. With the rapid development of science and technology and new media, the education field must keep up with the times and achieve technological innovation. In the context of the new curriculum standards issued in 2022, the multimodal teaching model has emerged. The multimodal teaching provides an effective carrier for the application of AI technology in primary school English teaching. This article analyzes the specific application paths of AI technology in multimodal teaching, aiming to provide primary school English teachers with an innovative and efficient teaching method reference. By deeply integrating AI technology with multimodal teaching, it can break the single-modal limitation of traditional teaching, fully mobilize students' multiple senses, stimulate students' interest and enthusiasm in learning, thereby effectively improving the quality and efficiency of primary school English teaching, better adapting to the requirements of the "double reduction" policy and the new curriculum standards, and promoting the all-round development of students.***

Keywords: Multimodal teaching, Artificial Intelligence, Primary School English**.**

1. Introduction

Primary school English is an English course designed for primary school students, which focuses on enabling students to master basic English knowledge and skills through systematic English learning activities. The teaching at this stage covers the recognition and application of letters, words, and sentence patterns, as well as basic grammar rules. It aims to comprehensively cultivate students' language skills in listening, speaking, reading, and writing. The teaching characteristics of primary school English mainly lie in its fundamental, entertainment, practicality, and comprehensiveness. Fundamentality emphasizes imparting solid language knowledge to students, laying a solid foundation for their English learning journey. Entertainment involves using lively and interesting teaching content and forms, such as songs, games, and stories, to stimulate students' learning interest and make them more actively involved in the learning process. Practicality focuses on the practical application of language through a large amount of listening, speaking, reading, and writing exercises, allowing students to constantly exercise and improve their language skills in practice. Comprehensiveness emphasizes that English teaching not only focuses on the transmission of language knowledge but also aims to cultivate students' cross-cultural awareness, emotional attitudes, and values, in order to achieve the all-round development of students.

The traditional primary school English teaching model has the following drawbacks: Firstly, the teaching methods are rather monotonous, mostly relying on teachers' lectures, and students are often in a passive state of accepting knowledge, lacking opportunities for active participation and interaction. This leads to low learning enthusiasm. Secondly, the teaching resources are limited, mainly relying on textbooks and simple teaching aids, which are unable to provide rich and diverse learning materials and a real language environment, making it difficult to cultivate students' comprehensive language application abilities. Lastly, the teaching evaluation methods are not scientific and comprehensive enough. They overly focus on students' test scores while neglecting the evaluation of their learning process and overall qualities. As a result, they cannot accurately reflect students' learning situations and progress.

With the rapid development of science and technology and new media, the education field must keep up with the times and achieve technological innovation. Against the backdrop of the new curriculum standards issued in 2022, the multimodal teaching model has emerged. This model utilizes multiple senses, such as hearing, vision, and touch, to convey information, emphasizing students' perception and experience, and enhancing their participation, enabling them to receive and understand knowledge through different media. As an emerging teaching model, multimodal teaching combines static resources (such as text, pictures, and charts) with dynamic resources (such as language, sound, actions, body language, and gestures) to stimulate students' multi-dimensional thinking, further deepen their understanding, and consolidate their memory.

The application of multimodal teaching in primary school English classrooms can provide students with more diverse and realistic learning scenarios. For instance, by playing original English animations, presenting vivid pictures or objects, and using body language and gestures to assist teaching, students can more intuitively understand the meanings of words and sentence patterns, thereby enhancing their language perception ability. Additionally, multimodal teaching also encourages students to participate in the learning process through imitation, performance, and interaction, thereby improving their language application skills and communication abilities.

Meanwhile, the rapid development of artificial intelligence technology has brought unprecedented changes to the education field. AI technology can analyze students' learning data and behavioral patterns intelligently, providing teachers

with more accurate teaching feedback and suggestions, helping teachers better understand students' learning needs and difficulties, and thus adjusting teaching strategies to achieve personalized teaching (Xiong & Zheng, 2024). In addition, AI technology can also provide intelligent tutoring and answer services for students, solving the problems they encounter in the learning process and improving learning efficiency.

Integrating multimodal teaching with AI technology can fully leverage the advantages of both, bringing more revolutionary changes to primary school English teaching. Through multimodal teaching, students can understand and master English knowledge more deeply under the stimulation of multiple senses; while AI technology can provide teachers with more scientific and comprehensive teaching support, helping students better achieve personalized learning.

2. Application of Multimodal Teaching in Primary School English

The multimodal teaching method is a teaching approach that integrates various symbolic modalities such as language, images, sounds, and gestures in the classroom, stimulating the participation of learners' multiple senses, and promoting the construction of meaning and the understanding of knowledge (Zhao & Wu, 2025). Specifically, multimodal teaching is manifested in that teachers utilize various teaching resources such as pictures, videos, audio, and physical objects, as well as non-verbal means like body language and facial expressions, to create a vivid, realistic, and interesting language learning environment (Zhang, 2010). This teaching method not only stimulates students' interest in learning, enhances their classroom participation, but also helps them understand English knowledge more intuitively and improve their language perception and application abilities. For example, when teaching animal-related words, teachers can use methods such as showing animal pictures, playing audio of animal sounds, and imitating animal actions to allow students to perceive and understand the meaning of the words from multiple perspectives, thereby deepening their memory.

2.1 Application of Multimodal Approach in Primary School English Vocabulary Teaching

Students in the primary school stage have cognitive characteristics such as well-developed image thinking, short attention span, and strong sensory experience demands. This provides a theoretical basis and practical need for the application of multimodal resources in vocabulary teaching. Multimodal resources work together through multiple sensory channels, such as vision, hearing, and touch, to create a richer and three-dimensional vocabulary learning environment, effectively stimulating students' interest and participation. Let's take the "How do you feel" unit 6 of the 6th grade of the People's Education Edition as an example.

2.1.1 Design and implementation of multimodal vocabulary introduction

Visual and auditory modal resources are the main carriers of vocabulary teaching. By integrating image, text, and sound symbol systems, they construct multi-dimensional vocabulary cognitive schemas for students (Li, 2022) Using multimedia courseware to play animated short films without text, the short films are set against familiar campus and family scenes, such as "jumping and cheering after achieving a full score in the exam", "lowering one's head and crying after losing a beloved school supply", "frightenedly running when chased by a big dog", etc. Students can intuitively perceive different emotional states through the facial expressions and body movements of the characters. During the playback, pause key frames to guide students to observe and describe "what the characters are doing", "how you think he is feeling now", and naturally lead to the English expressions of the core vocabulary of this unit. Then, present the corresponding static graphic cards, with the front side being the high-definition close-up of the characters corresponding to the animation scenes, and the back side marking the English word, phonetic symbol, and simple Chinese explanation. Through flashcards, help students accurately connect the emotional experience in the dynamic situation with the static vocabulary symbols, strengthening the visual memory of the vocabulary.

2.1.2 Consolidation of students' vocabulary connections

In addition to visual and auditory modal resources, which can be well applied in primary school English vocabulary teaching. Kinesthetic modal resources use physical activity forms such as body movements, gesture expressions, and spatial movement to add an experiential and interactive dimension to vocabulary learning. The participation of body movements can activate students' kinesthetic intelligence, transforming vocabulary learning from a simple cognitive activity into a full-body experience. After learning the previous emotion words, the word flashcards can be used in the form of words. When a word is given, students need to make the corresponding emotions. This can fully mobilize students' body senses, enabling them to remember the meaning of the word more deeply. For example, when showing the "happy" word flashcard, students immediately show bright smiles, some even jump excitedly, using cheerful movements to interpret the joyful mood represented by "happy"; when seeing the "sad" flashcard, students lower their heads, frown, make crying or sighing movements, truly expressing the emotion of sadness; when the "angry" flashcard appears, students clench their fists, frown, some even stomp their feet, showing their understanding of the word with an angry posture. Through the application of this kinesthetic modal resource, students no longer memorize words through rote learning, but connect the vocabulary with specific emotional experiences through body movements. This not only deepens their understanding of the vocabulary but also significantly improves the firmness of memory. Moreover, this interesting learning method greatly stimulates students' learning enthusiasm, making them more actively and actively engage in vocabulary learning, further enhancing the learning effect.

2.2 Application of Multimodal Approach in Primary School English Reading Teaching

In primary school English teaching, reading mainly consists of English stories. The English story class in primary school is the core domain for language acquisition and thinking development, and its teaching quality directly affects the

effectiveness of students' English core literacy. The multimodal teaching method, by integrating various resources such as text, audio, images, and body language, can create a three-dimensional perception scene for English story learning, allowing language elements such as vocabulary and sentence patterns to naturally permeate through rich modalities. We take the "Story Time" of Unit 5 in the 6th grade of the People's Education Edition as an example.

2.2.1 Warm-up and Introduction: Multimodal activation of prior knowledge, creating an atmosphere

1) Audio + Action Modality: Play a cheerful English career song (such as "Jobs Song"), and simultaneously display the song lyrics and corresponding action diagrams on the presentation. The teacher leads the students to sing along and imitate the actions (such as the doctor listening, the chef cooking, the writer writing, the farmer planting rice) while following the music. Through auditory input and kinesthetic input, students activate their existing vocabulary related to careers and quickly immerse themselves in the English learning atmosphere.

2) Image + Questioning Modality: On the interactive whiteboard, display high-definition pictures of different careers (without text annotations), and the teacher asks: "What can you see?" Guess what job it is?" Guide students to express job titles in English. For uncertain words, play the standard pronunciation recording of the words (audio modality) and have students follow along to correct their pronunciation. Through the visual impact of images and the accurate input of audio, consolidate the old knowledge and lay the groundwork for the new lesson.

2.2.2 Pre-reading prediction: Image + Context Modality, trigger thinking

1) Animation segment + Prediction Modality: Play the animation segment of Story time (without dialogue sound, only the background sound), showing the discussion between Zoom and Zip about jobs. Guide students to make predictions based on the scene in the picture (visual modality) and life experience, stimulating reading interest.

2) Real objects + Vocabulary Modality: Display the prepared real objects (clown mask, magic wand), introduce the core job vocabulary of this lesson together with the props, and let students touch and observe the props. Through the multi-modal combination of touch and vision, deepen the understanding and memory of the vocabulary. For example, pick up the clown mask and ask the students, "What other unusual jobs can you think of?"

2.2.3 Reading exploration: Multi-modal assistance for reading, breaking through key points

1) Initial reading: Animation + Text Modality, overall perception. Play the complete Story time animation video (with dialogue sound and English subtitles), and students follow the video and read softly. Require students to pay attention to the expressions, actions, and scene changes in the picture, and combine the audio dialogue and text subtitles to initially understand the main idea of the text, and circle the unfamiliar words.

2) Re-reading: Image + Task Modality, sort out details. Divide the Story time text into 4 key scenes, and make task cards with graphics and corresponding dialogue segments and questions (each card contains a scene picture, the corresponding dialogue segment, and a question). Distribute these cards to each group. Group cooperation to complete the tasks: ① Match the scene pictures with the dialogue segments; ② Read with questions (such as "What does Zoom want to do?"). What does Zip want to do?"). Teachers use interactive whiteboards to display pictures of various scenarios and invite groups to share their answers. Through visual images, they assist students in organizing the logical structure of the text and focusing on the core information.

3) Intensive Reading: Audio + Imitation Mode, Strengthening Language. Select the core dialogue segments from the text, play the audio (adjusting the speed, starting slow and then speeding up), and have students follow and imitate. The teacher uses the interactive whiteboard to display the dialogue text, mark the pronunciation and intonation, and combine the characters' expressions and tones in the pictures to guide students to imitate the tones of different characters, improving the accuracy of oral expression. For difficult sentences, break them down, follow them sentence by sentence, and have teachers and students interactively answer questions to overcome them.

2.2.4 Post-reading Output: Multimodal Contextual Expression, Enhancing Skills

1) Role-playing: Props + Context Mode, Internalizing Language. Divide the students into groups, and each group is given a set of physical props and the simplified version of the Story Time scene script. The groups cooperate to perform role-playing. Students are required to combine the props, imitate the actions and tones of the characters in the animation, and fully interpret the content of Story time. The teacher walks around to guide and encourage students to add their own understanding, such as adding simple greetings. Through the combination of multimodal elements such as body language, physical props, and language expression, students internalize the learned language in a real context.

2) Extension Expression: Images + Creation Mode, Transfer, and Application. The interactive whiteboard displays pictures of different job scenarios, guiding students to use the sentence patterns learned in this lesson for extension expression. Then, students draw pictures of other special occupations they know and describe the job content of that occupation in 3-5 sentences. After completion, they share within the group. Through the combination of visual modality, such as drawing and language expression, knowledge is transferred and applied.

3. Application Practice of AI Technology in Multimodal Teaching of Primary School English

3.1 The Supporting Role of AI Technology in Multimodal Teaching

In our multimodal teaching, it is necessary to combine multiple modalities to fully engage students' concentration and enthusiasm.

The integration of AI technology can add more intelligent support to this process.

In multimodal teaching, a lot of modal data is needed for support, and collecting this modal data requires a lot of time. AI technology can, with its powerful data processing capabilities, quickly and accurately screen and integrate suitable modal data for different teaching contents and students' characteristics. For example, when teaching an English course on career themes, AI can select high-definition and realistic career pictures and vivid and interesting career animation videos from massive online resources. These materials not only enrich the teaching content but also stimulate students' senses in multiple dimensions, greatly enhancing their learning interest and participation. At the same time, when teachers are learning about relevant career information, they can use AI to provide accurate information. In addition, apart from using AI to help search for pictures and videos, teachers can combine the characteristics of the class to generate more personalized videos and pictures suitable for their teaching mode with the help of AI.

AI technology can also adjust the presentation method and difficulty of modal data in real time based on students' learning progress and feedback. For students who are learning faster and have a stronger understanding ability, AI can provide more challenging and expansive materials, such as animal science videos with more professional vocabulary and complex sentence structures; for students who are slightly struggling with learning, it can provide more basic and intuitive materials, such as simple animal cartoon pictures accompanied by simple English descriptions, helping students gradually build confidence and better master knowledge.

In terms of classroom management, AI technology can also play an important role. Through intelligent monitoring of students' classroom performance, such as their concentration level and the frequency of interaction, teachers can promptly understand the students' learning status and adjust teaching rhythms and strategies to ensure that every student can actively participate in classroom learning.

Moreover, AI technology can also perform intelligent analysis and optimization of multimodal materials. Through the analysis of a large amount of teaching data, AI can understand which type of multimodal materials is most effective in improving students' learning outcomes, thereby providing scientific suggestions for teachers to help them choose and use multimodal materials more reasonably, further improving the quality and efficiency of multimodal teaching.

3.2 The Supporting Role of AI in Students' Post-Class Multimodal Learning

Learning is not limited to the classroom; post-class learning is also an important part of improving students' English proficiency. For primary school students, it is difficult to master vocabulary just by the time in class. For teachers, it is impossible to implement personalized teaching for every student. Therefore, AI technology can create a personalized post-class multimodal learning environment for students. Students can provide AI software, such as Douba, to check their pronunciation accuracy and have Douba conduct word listening and writing for them.

At the same time, AI can provide students with intelligent learning partners. This learning partner can interact with students in the form of voice communication, answer students' questions encountered in post-class learning, and guide students to expand their learning. For example, after students complete their post-class homework, the AI learning partner can have a simple English conversation with them, focusing on the content learned that day, to exercise students' oral expression ability. At the same time, AI technology can also provide timely assessment and feedback on students' post-class learning outcomes. By analyzing students' homework completion, interaction performance with the learning partner, etc., AI can generate detailed learning reports, point out students' strengths and weaknesses, and provide targeted suggestions for students' subsequent learning, helping students learn more efficiently through post-class multimodal learning.

References

[1] LI Chuanxin. (2022). Application of Multimodal Teaching Mode in English Vocabulary Teaching. *The Guide of Science & Education, 24*, 29–31. https://doi.org/10.16400/j.cnki.kjdk.2022.24.009

[2] Xiong ying & Zheng jipeng. (2024). AI Technology Enables Innovative Practices in Large-Scale Personalised English Teaching. *Chinese University Science & Technology, 9*, 6–10. https://doi.org/10.16209/j.cnki.cust.2024.09.004

[3] Zhang delu. (2010). Preliminary Investigation into the Concept of Design and the Selection of Modalities in Multimodal Foreign Language Teaching. *Foreign Languages in China, 7*(3), 48–53, 75. https://doi.org/10.13564/j.cnki.issn.1672-9382.2010.03.011

[4] Zhao xiufeng & Wu yuxin. (2025). Integration of Multimodal Pedagogy and Embodied-Cognitive Linguistics. *Foreign Languages in China, 22*(4), 4–10, 22. https://doi.org/10.13564/j.cnki.issn.1672-9382.20250721.001