Journal of Education and Culture Studies
ISSN 2573-0401 (Print) ISSN 2573-041X (Online)
Vol. 9, No. 1, 2025
www.scholink.org/ojs/index.php/jecs

Original Paper

The Correlation Analysis of Multimodal Interaction in English Class on Vocabulary Depth Acquisition

Cuilan Zhao

Chengdu Polytechnic, Guangyuan, Sichuan, China

Received: January 30, 2025 Accepted: February 25, 2025 Online Published: March 8, 2025

doi:10.22158/jecs.v9n1p193 URL: http://dx.doi.org/10.22158/jecs.v9n1p193

Abstract

This study, set against the backdrop of an English classroom, explores the impact of multimodal interaction teaching models on vocabulary depth acquisition and its underlying mechanisms. By integrating Cognitive Load Theory with Sociocultural Theory, the study analyzes the pathways through which multimodal synergies, such as visual, auditory, and kinesthetic, affect vocabulary semantics, grammar, and pragmatic abilities. Combining theoretical explanations with empirical methods, the research reveals how multimodal interaction promotes the deep internalization of vocabulary knowledge through situational construction, cognitive resource optimization, and socialized practice. The results indicate that multimodal teaching can effectively strengthen semantic network connections, enhance the implicit acquisition efficiency of grammatical rules, and increase the contextual adaptability and cultural perception of vocabulary. The study provides theoretical support for innovative English vocabulary teaching, emphasizes the practical value of multimodal task design in achieving multidimensional language ability development, and offers a reference for the transformation of classroom models.

Keywords

English classroom, Multimodal interaction, Vocabulary depth acquisition, Cognitive Load Theory

1. Introduction

Against the backdrop of the deep integration of globalization and digitalization, English teaching is undergoing a paradigm shift from traditional "unimodal" to "multimodal interaction." Traditional vocabulary teaching often relies on text repetition and mechanical memory, resulting in learners' mastery of vocabulary remaining at a shallow cognitive level, making it difficult to achieve coordinated development of semantic understanding, grammatical application, and pragmatic abilities. Vocabulary depth acquisition, as a core dimension of language ability, requires learners not only to master the form

and basic meanings of vocabulary but also to flexibly apply their grammatical rules, cultural connotations, and communicative functions in real contexts. However, in current English classrooms, the inefficiency of vocabulary depth acquisition remains prominent. How to break through this bottleneck through innovative teaching models has become an urgent research issue. Multimodal interaction provides a new theoretical perspective and practical path for vocabulary depth acquisition. With the synergistic stimulation of multiple sensory channels, such as visual (e.g., images, videos), auditory (e.g., dialogues, music), and kinesthetic (e.g., gestures, role-playing), learners can activate cognitive resources through embodied experiences, deeply bind vocabulary information with emotions, contexts, and social interactions, thereby promoting the internalization and transfer of vocabulary knowledge.

This study focuses on the practice of multimodal interaction in English classrooms, aiming to reveal its impact pathways and effects on vocabulary depth acquisition. Through a combination of theoretical analysis and empirical research, this paper attempts to answer two core questions: First, how does multimodal interaction promote the deep processing of vocabulary through the synergistic effects of different modes? Second, how does this interactive mode specifically perform in enhancing vocabulary semantics, grammar, and pragmatic abilities? The research results will provide theoretical and practical bases for optimizing English classroom design, promoting technology-enabled language teaching, and at the same time, open up new directions for cross-modal cognitive research in vocabulary acquisition.

2. Theoretical Basis of Multimodal Interaction

2.1 Definition and Characteristics of Multimodal Interaction

Multimodal interaction refers to a teaching method that integrates visual, auditory, kinesthetic, and other sensory channels, combining linguistic and non-linguistic symbols (e.g., images, sounds, actions, etc.) for information transmission and meaning construction (Lu, 2015). In English classrooms, multimodal interaction encompasses not only traditional text explanations and oral dialogues but also includes diverse forms such as video playback, image display, physical demonstrations, and interactive games. Its core characteristic lies in breaking the limitations of a single mode, enhancing learners' cognitive engagement through multisensory synergistic stimulation. Taking the word "storm" as an example, (visual), (auditory), and hands to mimic a gale (kinesthetic) can help learners deeply integrate the semantics, context, and emotional experience of the vocabulary, thus going beyond surface mechanical memory.

2.2 Compatibility of Cognitive Load Theory and Multimodal Interaction

Cognitive Load Theory provides a scientific basis for the effectiveness of multimodal interaction in vocabulary teaching. The theory posits that the capacity of human working memory is limited, and multimodal information can be processed in parallel through different sensory channels, thereby optimizing the allocation of cognitive resources (Zhang & Wang, 2025). In English classrooms, when vocabulary information is presented simultaneously through visual, auditory, and kinesthetic channels,

learners can reduce cognitive load on a single channel while enhancing the depth and persistence of information processing. For example, when learning the word "symmetry," combining visual symmetry displays of geometric shapes with origami operations can help learners reduce cognitive pressure when understanding abstract concepts, thereby improving the efficiency of semantic integration of vocabulary. Studies have shown that the hierarchical design of multimodal interaction can balance intrinsic cognitive load, extraneous cognitive load, and germane cognitive load, ultimately achieving "cognitive relief" in vocabulary depth acquisition.

2.3 Multimodal Interaction from the Perspective of Sociocultural Theory

Sociocultural Theory emphasizes that language learning is essentially a process of social interaction and meaning negotiation. Multimodal interaction creates authentic sociocultural contexts, shifting vocabulary acquisition from individual cognition to group collaboration. In English classrooms, teachers guide learners to engage in language practice with the support of multimodal resources through interactive activities such as role-playing and group discussions. This process not only helps to strengthen the pragmatic functions of vocabulary but also prompts learners to actively adjust language strategies in social interactions, deepening their understanding of the cultural connotations of vocabulary. Sociocultural Theory further points out that multimodal interaction helps learners cross the "zone of proximal development" through "scaffolding" mechanisms (e.g., teacher demonstrations, peer feedback), gradually internalizing external support into autonomous vocabulary application abilities. This dynamic process from social interaction to internalization is the key path for vocabulary depth acquisition to move from "formal mastery" to "meaning generation."

3. The Practice of Multimodal Interaction in English Classrooms

3.1 Visual Modality: From Images to Dynamic Contexts

The visual modality is the most intuitive presentation method in multimodal interaction, with its core lying in activating learners' spatial cognitive abilities through concrete symbols. In English vocabulary teaching, static images (such as illustrations, charts) can transform abstract vocabulary into perceptible visual information (Zhang, 2023). Taking the word "ecosystem" as an example, teachers can display ecological scenes that include forests, rivers, and animals, helping students establish a connection between vocabulary and the real world, thereby deepening their understanding of the concept of "biotic community interaction." Dynamic visual resources (such as videos, animations) further expand the possibilities for context construction. Taking "hurricane" as an example, playing documentary clips of the hurricane formation process can not only display the physical characteristics of the vocabulary but also convey its destructive power and emergency situations through dynamic images, prompting learners to strengthen the emotional color and pragmatic context of the vocabulary in addition to semantic memory. Studies have shown that the superimposed use of visual modalities (static + dynamic) can increase the long-term memory rate of vocabulary by about 23%, especially when it comes to complex concepts, visual aids significantly lower cognitive thresholds.

3.2 Auditory Modality: Immersive Penetration of Sound

The auditory modality mainly constructs a three-dimensional network of language input through speech, music, and environmental sound effects, directly affecting the phonetic encoding and pragmatic perception of vocabulary. In the classroom, teachers can reinforce the phonetic memory of vocabulary through "listening to meaning" activities. For example, playing dialogue recordings containing verbs like "whisper," "shout," "murmur," learners need to match the meaning based on differences in tone, volume, and tone, a process that not only consolidates the pronunciation rules of vocabulary but also deepens the identification of the emotional function of vocabulary. The creative integration of background music can regulate the learning atmosphere and enhance the emotional stickiness of vocabulary. For instance, when teaching "melancholy," teachers can play Chopin's nocturnes, and learners, through the emotional rendering of the music, more easily understand the implied meaning and cultural associations of the vocabulary. Moreover, the simulation of real-world sound effects can help learners master the communicative function of vocabulary in virtual scenarios. For example, in the context of airport announcements and restaurant ordering, the high-frequency occurrence of "boarding pass" and "menu" in corresponding sound effects can reinforce the pragmatic conditioned reflexes of vocabulary.

3.3 Kinesthetic Modality: Situational Experience through Physical Participation

The kinesthetic modality emphasizes achieving "learning by doing" through body movements and hands-on activities, its value lies in transforming vocabulary memory from passive reception to active construction. In lower-grade classrooms, teachers can use "TPR (Total Physical Response)" to design vocabulary command games, such as requiring students to complete corresponding actions based on verbs like "jump," "crouch," "rotate," allowing learners to internalize the semantic and grammatical structures of vocabulary (such as the transitivity of verbs) through physical movement. Upper-grade classrooms can introduce role-playing and project-based tasks. For example, when simulating an "interview" scenario, students need to interact using gestures, expressions, and props (such as resumes, business cards) to complete the interaction, students in this process need to accurately use vocabulary such as "qualification," "strength," and also adjust language strategies based on the other party's response, thereby enhancing the communicative adaptability of vocabulary. The intervention of kinesthetic modality can increase the active output rate of vocabulary, especially when it comes to abstract concepts, physical participation can compensate for the limitations of language expression and promote the multi-dimensional internalization of vocabulary (Li, 2015).

3.4 Multimodal Integration: The Generation Mechanism of Synergistic Effects

The isolated use of a single modality is difficult to achieve the optimal effect of deep vocabulary acquisition, while the organic integration of multimodalities can activate higher-order cognition through cross-channel complementarity. For example, when explaining "photosynthesis," teachers can design a "three-step integration" task: first, display a plant cell structure diagram (visual), guiding students to label key parts such as "chloroplast," "glucose"; then play a scientific commentary audio

(auditory) on the photosynthesis process, requiring students to recount the steps of energy conversion; finally, group (kinesthetic), using different materials to simulate the absorption of light energy and the release of oxygen. Throughout the process, the visual modality lays the semantic foundation, the auditory modality strengthens the logical chain, and the kinesthetic modality promotes knowledge transfer. The synergistic effect of these three modes allows learners to complete the deep processing of vocabulary in the cycle of "observation-listening-manipulation." Empirical data shows that multimodal integration strategies can significantly improve the complexity of the semantic network of vocabulary and effectively help learners reduce grammatical error rates and pragmatic errors. This "1+1>2" synergistic effect is the underlying logic of multimodal interaction empowering deep vocabulary acquisition.

4. The Impact Mechanism of Multimodal Interaction on Deep Vocabulary Acquisition

4.1 Definition and Dimensions of Deep Vocabulary Acquisition

Deep vocabulary acquisition refers to the systematic internalization of vocabulary knowledge by learners, covering three core dimensions: semantics, grammar, and pragmatics. The semantic dimension requires learners not only to master the literal meaning of vocabulary but also to understand its polysemy, metaphorical extensions, and the associative network with other vocabulary. The grammatical dimension involves the syntactic rules and collocation restrictions of vocabulary, such as the transitivity of verbs, the positional attributes of adjectives, etc. The pragmatic dimension focuses on the functional adaptability of vocabulary in real communication, including cultural context, emotional color, and social conventions. These three dimensions together constitute the "three-dimensional" goal of deep vocabulary acquisition.

4.2 The Impact of Multimodal Interaction on Semantic Memory

The strengthening effect of multimodal interaction on semantic memory originates from the distributed activation of cognitive resources. When visual, auditory, and kinesthetic modalities act synchronously on the same word, the information input from different sensory channels forms multiple encoding paths in the learner's brain (Yuan, 2016). For example, the visual modality activates the right brain's visual-spatial processing area through images or videos, the auditory modality mobilizes the left brain's language processing area through speech input, and the kinesthetic modality associates the motor cortex through body movements. This cross-brain area coordination can transform the abstract semantics of words into embodied experiences, thereby enhancing the depth and persistence of memory encoding. Neuroscience research indicates that multimodal stimulation can significantly improve the information integration efficiency of the hippocampus, causing the semantic network of words to form denser connections in long-term memory. Multimodal interaction strengthens the context-dependency of words through contextualized input, reduces the fragmentation of semantic memory, and prompts learners to shift from isolated word memory to the construction of systemic semantic frameworks.

4.3 The Impact of Multimodal Interaction on Grammar and Pragmatic Competence

The implicit acquisition of grammatical rules and the dynamic adaptation of pragmatic competence depend on the "cognitive-social" dual context created by multimodal interaction. At the grammatical level, multimodal interaction uses visual aids (such as syntactic structure diagrams) and kinesthetic operations (such as sentence reassembly games) to visualize and operationalize abstract rules. For example, when learners construct sentences by dragging word cards on the screen (kinesthetic), visual feedback (such as color marking subjects, verbs, and objects) can instantly correct grammatical errors. At the pragmatic level, multimodal interaction provides "social input" by simulating real communication scenarios (such as video dialogues, role-playing), helping learners capture the contextual restrictions and cultural connotations of words in interaction. For example, teachers can introduce dialogue videos with different cultural backgrounds (visual + auditory), and learners can gradually internalize the politeness levels and emotional tendencies of words by observing the speaker's expressions, tones, and body language, thereby avoiding pragmatic errors.

5. Empirical Research and Data Analysis

5.1 Research Design and Implementation

To verify the correlation between multimodal interaction and deep vocabulary acquisition, this study adopted a mixed research method, selecting two parallel classes from a middle school (experimental group 60 people, control group 60 people) for a 12-week controlled experiment. The experimental group fully integrated multimodal interaction strategies into vocabulary teaching, including visual aids (dynamic courseware, image matching), auditory tasks (contextual audio, voice imitation), and kinesthetic activities (role-playing, manual modeling); the control group continued to use traditional text explanations and repetitive practice patterns. The research tools include vocabulary depth test papers (including semantic analysis, grammatical error correction, and pragmatic context questions), classroom observation scales, and learning logs. The reliability and validity of the test papers were verified through expert review and pre-experiments, and data collection was divided into three stages: pre-test, mid-test, and post-test.

5.2 Data Results and Significance Analysis

The vocabulary depth scores of the experimental group and the control group showed significant differences in the post-test. In the semantic dimension, the experimental group's accuracy rate for distinguishing polysemous words (89.2%) was 38.3% higher than that of the control group (64.5%); in the grammatical dimension, the experimental group's error rate in correcting verb collocations and sentence structures (76.8%) was higher than that of the control group (51.2%), with a 33.3% reduction in errors; in the pragmatic dimension, the experimental group's accuracy rate in situational multiple-choice questions (82.1%) far exceeded that of the control group (57.4%). Independent sample t-tests showed that the significance levels of all three sets of data reached $p<0.01$, with effect sizes ranging from 0.83-1.21, indicating that multimodal interaction has a medium to strong effect on

vocabulary depth acquisition. Classroom observation data further revealed that the participation rate of the experimental group students in multimodal tasks (91.7%) was significantly higher than that of the control group's mechanical practice participation rate (62.3%).

5.3 Mechanism Discussion and Teaching Implications

The data results verified the multi-level facilitative effect of multimodal interaction on vocabulary depth acquisition. The improvement in semantic scores comes from the complementary encoding of cross-modal information. The visual modality anchors the core word meaning through images, the auditory modality strengthens polysemous associations through contextual audio, and the kinesthetic modality deepens the accessibility of the semantic network through embodied operations. The progress in grammar and pragmatic abilities is related to the "implicit input-explicit output" cycle of multimodal interaction. For example, role-playing (kinesthetic) requires students to autonomously invoke grammatical rules in communication, while video analysis (visual + auditory) internalizes pragmatic norms through real-world language input. Analysis of teaching logs indicates that the experimental group students' emotional identification with vocabulary has significantly increased, confirming the additional value of multimodal interaction on the emotional dimension. Based on this, English classrooms need to systematically design multimodal task chains, such as breaking down vocabulary teaching into three stages: "visual input - auditory imitation - kinesthetic output," and using technological tools (such as interactive whiteboards, VR scenarios) to optimize modality synergy efficiency.

6. Conclusion

This study confirms that multimodal interaction significantly enhances the effectiveness of vocabulary depth acquisition in English classrooms through the integration of information across sensory channels and cognitive synergy. Empirical data show that the experimental group's overall performance in vocabulary depth tests is stronger than that of traditional teaching groups, especially in pragmatic appropriateness and polysemy analysis. Future research can further explore the differential impact of multimodal interaction on special learners (such as young children, second language learners with disabilities), and extend the experimental period to observe long-term memory effects. From a cross-disciplinary perspective, the integration of neurolinguistics and educational technology will provide finer neural evidence for multimodal teaching mechanisms, promoting vocabulary depth acquisition research to delve deeper into brain science.

References

Li Qionghua. (2015). Research on Multimodal Interactive Teaching in College English Listening and Speaking Teaching from the Perspective of Output-driven. *Journal of Hubei Normal University: Philosophy and Social Sciences Edition*, *35*(6), 3.

Lu Chenchen. (2015). A Review and Prospect of Domestic Multimodal Teaching Research. *Journal of Hubei University of Economics: Humanities and Social Sciences Edition*, *12*(8), 3.

Yuan Wenjuan. (2016). Teacher-student Interaction in Multimodal College English Classroom Teaching. *New Campus: Reading Edition*, *2016*(4), 1.

Zhang Yue. (2023). *The Impact of Multimodal Vocabulary Teaching Model on the Breadth and Depth of English Vocabulary Knowledge of Junior High School Students*. Qufu Normal University.

Zhang Yuxin, Wang Honggang. (2025). Application Analysis of Multimodal Theory in Senior High School English Vocabulary Teaching. *Advances in Education*, *2025*, 15.