

TEACHING ENGLISH THROUGH MULTIMODAL INPUT: HOW COMBINING VIDEO, AUDIO, AND TEXT ENHANCES RETENTION

Salixova Malika Bahtiyor qizi

Toshkent Menejment va Iqtisodiyot Instituti.

eshimbetovaguzel@gmail.com

<https://doi.org/10.5281/zenodo.15634581>

Abstract. *In response to the evolving needs of English as a Foreign Language (EFL) learners and the growing prominence of digital literacy, this study investigates the pedagogical benefits of multimodal input—specifically, the integration of video, audio, and text—in language instruction. A mixed-methods approach was employed involving 60 intermediate-level EFL students, divided into an experimental group exposed to multimodal content and a control group receiving traditional textbook-based instruction over six weeks. Data from pre- and post-tests, surveys, and semi-structured interviews revealed statistically significant improvements in vocabulary retention and listening comprehension among students in the experimental group.*

Furthermore, engagement and motivation levels were notably higher in the multimodal condition. These findings align with theoretical frameworks such as Dual Coding Theory, Cognitive Load Theory, Multimedia Learning Theory, and Krashen's Input Hypothesis, reinforcing the argument for incorporating multimodal strategies in contemporary EFL classrooms.

Keywords: *multimodal learning, EFL, vocabulary retention, listening comprehension, student engagement, digital tools, Dual Coding Theory, multimedia learning, input hypothesis, cognitive load theory*

INTRODUCTION

In today's changing world of English language teaching, traditional methods that only use textbooks and speaking practice are no longer enough to meet the different needs of students.

Classrooms are becoming more culturally and linguistically diverse, and students are more familiar with digital technology than ever before. Because of this, teaching needs to be more flexible, interesting, and adapted to modern learners. Learning a language is no longer just about reading and memorizing vocabulary—it works better when students interact with content through different senses and activities.

Thanks to digital tools and multimedia, teachers now have more ways to present information that can help students learn more effectively. Multimodal learning—where information is shown using visuals (like images or video), sounds (like speech or music), and written text—has become popular because it helps students understand and remember better.

This study looks at how using video, audio, and text together can support English as a Foreign Language (EFL) students, especially by improving their memory and motivation during lessons in different types of classrooms.

MATERIALS

This study used a variety of learning materials designed to support English language development through different types of input. These included video materials with English subtitles, such as educational videos from YouTube and TED-Ed, as well as audio recordings with matching transcripts from sources like BBC Learning English and language-learning

podcasts. In addition, learners used interactive multimodal platforms like FluentU and Edpuzzle, which combine video, text, and quizzes to support active learning. For comparison, students also worked with traditional printed texts. All materials were chosen to match the intermediate level of English as a Foreign Language (EFL) learners and to reflect real-life, natural use of English in everyday situations.

RESEARCH AND METHODS

This study followed a mixed-methods design, combining both quantitative and qualitative research approaches to examine the impact of multimodal input on English language learning. A total of 60 EFL learners between the ages of 18 and 25 participated. They were divided into two groups: the experimental group, which received instruction using multimodal materials (video, audio, and text), and the control group, which followed a traditional textbook-based curriculum.

Both groups studied the same topics over a six-week period to ensure consistency. To measure learning outcomes, researchers used pre- and post-tests focused on vocabulary retention and listening comprehension. In addition, surveys were distributed to evaluate learners' engagement and motivation. For deeper insights into student experiences, semi-structured interviews were conducted with selected participants.

The study was grounded in key theoretical frameworks. Allan Paivio's Dual Coding Theory (1986) argues that people learn better when information is processed through both verbal and visual channels, which supports the design of the experimental group. In contrast, John Sweller's Cognitive Load Theory (1988) warns against overloading learners with too much information at once, which raises the question of whether multimodal input might overwhelm some students. However, Richard Mayer's Multimedia Learning Theory (2001) finds a middle ground, suggesting that well-structured multimedia content—where visual and verbal elements are meaningfully integrated—can enhance understanding and memory if designed properly.

While Paivio emphasizes the benefits of using multiple channels for encoding information, Sweller cautions that if the material is not carefully designed, cognitive overload may hinder learning. Mayer bridges these views by showing that the success of multimodal input depends on thoughtful instructional design that reduces unnecessary processing and supports meaningful learning.

By drawing on these differing perspectives, this study aims not only to test the effectiveness of multimodal input but also to examine how its application aligns—or clashes—with the theories of leading scholars in educational psychology.

RESULTS

The findings of this study clearly support the theoretical advantages proposed by Paivio, Sweller, and Mayer. The experimental group, which received multimodal instruction, showed a statistically significant improvement in both vocabulary retention and listening comprehension compared to the control group using traditional methods.

Quantitative data revealed that vocabulary retention scores increased by 32% in the experimental group, while the control group showed a more modest 14% improvement.

Similarly, listening comprehension scores followed the same trend, favoring the multimodal approach. These results align with Paivio's Dual Coding Theory, which suggests that presenting information through both visual and verbal channels improves memory.

Survey data further reinforced the effectiveness of the multimodal method. Eighty-five percent of students in the experimental group reported higher levels of engagement and motivation, while only 43% of students in the control group reported the same. This supports Mayer's Multimedia Learning Theory, which highlights the motivational benefits of combining different media formats when teaching.

Qualitative insights from interviews added depth to the quantitative results. Many learners in the experimental group said that the combination of video, audio, and text helped them better understand context, remember vocabulary more easily, and stay focused during lessons. They also described the content as more relatable, enjoyable, and practical for real-life communication. These responses suggest that multimodal input not only enhances comprehension but also creates a more student-centered and emotionally engaging learning environment.

Taken together, the data supports the idea that, when designed appropriately, multimodal instruction does not overload learners (as Sweller warned), but rather improves both cognitive and emotional engagement—provided that content is clear, structured, and relevant to learners' needs.

DISCUSSION

These results align with and further validate the theoretical foundations of multimodal learning. The findings support the theoretical foundations of multimodal learning. According to Dual Coding Theory (Paivio, 1986), presenting information in both visual and auditory modes enhances memory retention. This was evident in the improved test scores of students exposed to multimodal input. Additionally, Cognitive Load Theory (Sweller, 1988) explains how distributing information across different sensory channels reduces overload and facilitates deeper processing. Mayer's Multimedia Learning Theory (2001) further supports that combining words and images leads to better learning outcomes, as was confirmed in this study.

The high levels of student engagement align with previous research (Mayer, 2009; Sherman, 2003) which shows that multimodal materials enhance motivation. Moreover, the contextual richness provided by videos and real-life audio supports natural language acquisition, a notion echoed by Krashen's Input Hypothesis (Krashen, 1985), which emphasizes the importance of comprehensible input.

CONCLUSION

Teaching English through multimodal input is a pedagogically sound, evidence-based approach that significantly enhances student retention and engagement. By incorporating video, audio, and text, educators can create immersive learning environments that cater to diverse learner needs and capitalize on the brain's natural processing abilities. Future research should explore the long-term effects of multimodal input on productive language skills such as speaking and writing.

REFERENCES

1. Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System*, 36(1), 85–93. <https://doi.org/10.1016/j.system.2007.11.005>
2. Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*. Longman.

3. Mayer, R. E. (2001). *Multimedia Learning*. Cambridge University Press.
4. Mayer, R. E. (2009). *Multimedia Learning: Second Edition*. Cambridge University Press.
5. Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford University Press.
6. Sherman, J. (2003). *Using Authentic Video in the Language Classroom*. Cambridge University Press.
7. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
8. Tschirner, E. (2001). Language acquisition in the classroom: The role of digital video. *Computer Assisted Language Learning*, 14(3–4), 305–319. <https://doi.org/10.1076/call.14.3.305.5796>
9. Vanderplank, R. (2010). Dealing with authentic spoken language: Listening comprehension strategies. In J. Field (Ed.), *Listening in the Language Classroom* (pp. 163–178). Cambridge University Press.