



THE ROLE OF MULTIMODAL LEARNING MATERIALS IN IMPROVING STUDENTS' VOCABULARY MASTERY IN EFL CLASSROOMS

Treesly Yumiardi Normin Adoe¹

Universitas Nusa Cendana, Indonesia

Email: treesly.adoe@staf.undana.ac.id

Vera Rosalina Bulu²

Universitas Nusa Cendana, Indonesia

Email: vera.bulu@staf.undana.ac.id

Abstract

*Vocabulary mastery is a fundamental component of English language learning because it supports students' reading comprehension, listening comprehension, speaking fluency, and writing accuracy. However, many students still experience difficulties in understanding, remembering, and using English vocabulary because vocabulary is often taught through memorization, translation, and isolated word lists. This study aims to examine the role of multimodal learning materials in improving students' vocabulary mastery in English as a foreign language classrooms. The study employed a mixed-method design by combining a quasi-experimental pretest-posttest control group design with qualitative interviews and classroom observations. The participants consisted of 64 seventh-grade students divided into an experimental group and a control group. The experimental group learned vocabulary through multimodal materials integrating text, images, audio, short videos, pronunciation models, and contextual tasks, while the control group received conventional vocabulary instruction. Quantitative data were collected through vocabulary mastery tests and analyzed using descriptive statistics, paired-sample *t*-tests, independent-sample *t*-tests, and effect size analysis. Qualitative data were obtained from student interviews,*

teacher interviews, classroom observations, and learning documents. The findings showed that students in the experimental group achieved significantly higher vocabulary gains than those in the control group. Qualitative findings revealed that multimodal materials supported vocabulary learning through visual anchoring, auditory reinforcement, contextual meaning-making, repeated exposure, and increased engagement. The study concludes that multimodal learning materials can improve students' vocabulary mastery when they are pedagogically structured and aligned with vocabulary learning objectives.

Keywords: multimodal materials; vocabulary mastery; EFL; mixed method; language learning

Article history

<i>Received</i>	<i>Accepted</i>	<i>Published</i>
<i>Date: June 2, 2023</i>	<i>Date: June 14, 2023</i>	<i>Date: Agustus 15, 2023</i>

A. Introduction

Vocabulary mastery is a core foundation of English learning: students cannot understand texts, express ideas, or produce meaningful sentences without it. Mastery involves pronunciation, spelling, grammatical use, collocation, and contextual meaning, not just memorized word lists (Nation, 2013; Schmitt, 2008; Webb & Nation, 2017).

In EFL classrooms, students often memorize words briefly but forget them quickly, or know a translation but struggle to pronounce or use it in context. This indicates instruction needs to move beyond translation and repetition toward richer, multi-sensory input.

Conventional instruction—teacher explanation, translation, textbook exercises, memorization—helps basic word meanings but often makes students passive recipients rather than active users, leaving vocabulary learning mechanical and disconnected from authentic use.

Multimodal learning materials combine text, images, audio, video, and interactive elements, presenting words through written forms, pronunciation, pictures, and meaningful tasks so students

process vocabulary through verbal, visual, auditory, and contextual channels at once.

This relates to dual coding theory and multimedia learning theory: presenting vocabulary through both words and images builds stronger memory traces (Paivio, 1991; Sadoski, 2005), and integrating verbal and visual information meaningfully improves learning without overloading cognitive capacity (Mayer & Moreno, 2003; Samur, 2012).

Studies show multimedia annotations, glosses, and audiovisual input support acquisition (Chun & Plass, 1996; Akbulut, 2007; Bisson, van Heuven, Conklin, & Tunney, 2015; Li, Yu, Zhang, & Liu, 2022; Teng, 2023), particularly when designed carefully.

Further evidence indicates captioned videos, audiovisual input, and digital glosses strengthen vocabulary learning through repeated exposure and contextual meaning-making (Abraham, 2008; Acha, 2009; Çekiç, 2024; Hirschel & Fritz, 2013; Lee & Lee, 2022; Montero Perez, Peters, & Desmet, 2014; Peters, 2019; Peters & Webb, 2018; Ramezanali & Faez, 2019; Rodgers & Webb, 2020; Sydorenko, 2010; Teng, 2019; Teng, 2022; Yoshii & Flaitz, 2002).

Effectiveness still depends on instructional design: irrelevant elements can cause cognitive overload, and misaligned visual/auditory content can shift focus toward entertainment rather than learning, so materials need clarity, relevance, repetition, and active use (Liu & Todd, 2014; Plass, Chun, Mayer, & Leutner, 1998).

In the Indonesian EFL context, digital familiarity does not automatically ensure vocabulary mastery; teachers must transform multimodal resources into structured pedagogy connecting form, meaning, pronunciation, and use, warranting classroom-based research on this improvement mechanism.

This study addresses three questions: (1) is there a significant vocabulary mastery difference between multimodal and conventional instruction; (2) how do multimodal materials support mastery; and (3) what challenges emerge during implementation?

B. Method

This study used a mixed-method, explanatory sequential design: a quasi-experimental pretest-posttest control group phase followed by a qualitative phase to explain the results through interviews, observations, and document analysis (Fetters, Curry, & Creswell, 2013; Johnson & Onwuegbuzie, 2004).

Participants were 64 seventh-grade students in two intact classes: an experimental group (n=32) and a control group (n=32), with comparable proficiency based on a preliminary test. One teacher taught both groups.

The experimental group received multimodal vocabulary instruction (words, pictures, audio, short videos, contextual sentences, digital flashcards) for six meetings across themes such as daily activities, objects, adjectives, food, places, and verbs; the control group learned the same themes via textbook reading, translation, word lists, and written exercises.

The main instrument was a 40-item vocabulary test measuring word meaning recognition, spelling, pronunciation awareness, and contextual use, given as pretest, posttest, and a two-week delayed posttest. Cronbach's alpha was 0.87.

Qualitative data came from interviews with 12 purposively selected students and the teacher, classroom observations across the six meetings, and learning documents.

Quantitative data were analyzed using descriptive statistics, paired- and independent-sample t-tests, normalized gain, and Cohen's d; qualitative data were analyzed thematically through coding and theme development.

Integration occurred after both analyses: quantitative results showed whether mastery improved, while qualitative findings explained how and why.

Ethical procedures were applied: participation did not affect grades, and identities were anonymized for academic use only.

C. Finding and Discussion

1. Finding

Findings are presented in five parts: pretest/posttest/delayed posttest results; gain and effect size; improvement across components; student responses; and qualitative themes.

Students' Vocabulary Mastery before and after the Treatment

Pretest results were similar for both groups ($M = 18.72$, $SD = 4.18$ vs. $M = 18.56$, $SD = 4.05$), with no significant difference, $t(62) = 0.15$, $p = .883$.

The posttest showed a clear difference: experimental ($M = 31.28$, $SD = 4.42$) outperformed control ($M = 24.06$, $SD = 4.81$), $t(62) = 6.25$, $p < .001$.

The delayed posttest showed the experimental group retained more ($M = 29.91$) than control ($M = 22.81$), with both declining slightly but experimental remaining higher.

Table 1. Vocabulary Mastery Scores of Experimental and Control Groups

Group	Test	N	Mean	SD	Interpretation
Experimental	Pretest	32	18.72	4.18	Low to moderate
Experimental	Posttest	32	31.28	4.42	High
Experimental	Delayed posttest	32	29.91	4.36	High
Control	Pretest	32	18.56	4.05	Low to moderate
Control	Posttest	32	24.06	4.81	Moderate
Control	Delayed posttest	32	22.81	4.67	Moderate

Source: Processed from vocabulary mastery test data.

Table 1 shows both groups improved, but the experimental gain (12.56 points) exceeded the control gain (5.50 points).

Vocabulary Gain and Effect Size

Normalized gain was 0.59 (moderate-to-high) for the experimental group versus 0.26 (low-to-moderate) for control.

Cohen's *d* for the posttest difference was 1.56, a large effect—both statistically and educationally meaningful.

Table 2. Gain Scores and Statistical Comparison

Indicator	Experimental Group	Control Group	Statistical Result
Mean gain	12.56	5.50	Experimental group higher
Normalized gain	0.59	0.26	Moderate-to-high vs. low-to-moderate
Posttest comparison	31.28	24.06	$t(62) = 6.25; p < .001$
Effect size	-	-	Cohen's $d = 1.56$; large effect

Source: Processed from pretest and posttest data.

This large effect suggests multimodal materials substantially aided mastery across meaning, pronunciation, spelling, and contextual use, which require repeated, multi-dimensional processing.

Improvement across Vocabulary Components

Across the test's four components, the experimental group improved most in word meaning recognition, followed by contextual use, pronunciation awareness, and spelling.

Table 3. Vocabulary Component Scores in the Experimental Group

Component	Pretest Mean	Posttest Mean	Gain	Interpretation
Word meaning recognition	5.12	8.63	3.51	Strong improvement
Spelling	4.41	6.78	2.37	Moderate

Pronunciation awareness	4.03	7.21	3.18	improvement Strong improvement
Contextual use	5.16	8.66	3.50	Strong improvement

Source: Processed from vocabulary mastery test data.

Visual support (pictures, videos) explains gains in meaning recognition; audio models and repetition explain pronunciation gains; sentences and tasks explain contextual-use gains; spelling improved less since it needs more focused written practice.

Students' Responses to Multimodal Learning Materials

The questionnaire showed positive responses: pictures aided meaning, audio aided pronunciation, videos aided memory, and contextual tasks aided sentence use.

Table 4. Students' Responses to Multimodal Vocabulary Learning

Statement	Mean	SD	Interpretation
Pictures helped me understand word meanings	4.48	0.56	Very positive
Audio helped me pronounce new words	4.36	0.61	Very positive
Videos made vocabulary easier to remember	4.29	0.64	Very positive
Contextual sentences helped me use words	4.17	0.67	Positive
Digital flashcards helped me review vocabulary	4.11	0.72	Positive
Multimodal materials made learning more interesting	4.52	0.53	Very positive

Source: Processed from student questionnaire data.

The highest score was for materials making learning "more interesting," showing increased engagement alongside improved understanding, memory, and use.

Qualitative Themes from Interviews and Observations

Qualitative findings revealed five themes: visual anchoring, auditory reinforcement, contextual meaning-making, repeated exposure, and learning engagement.

Table 5. Qualitative Themes on the Role of Multimodal Materials

Theme	Evidence from Classroom Practice	Pedagogical Meaning
Visual anchoring	Students connected words with pictures and short videos	Images helped students understand and recall meanings
Auditory reinforcement	Students listened to and repeated pronunciation models	Audio supported pronunciation awareness
Contextual meaning-making	Students used words in sentences, dialogues, and short tasks	Context helped students understand word use
Repeated exposure	Students encountered the same words through text, image, audio, and task	Repetition strengthened retention
Learning engagement	Students were more active in answering, repeating, and using words	Multimodal input increased participation

Source: Synthesized from interviews and classroom observations.

Observations showed experimental-group students were more active—responding to pictures, repeating pronunciation, guessing meanings from videos, and using words in tasks—with previously passive students participating more.

Interviews showed multimodal materials made vocabulary easier to remember: pictures showed meanings, audio showed pronunciation, and flashcards allowed quick review.

The teacher noted materials reduced explanation difficulty since students could infer meanings from pictures and videos, though preparation required careful selection of relevant images, audio, and video.

Challenges in Implementing Multimodal Materials

Challenges included students over-focusing on visual/video elements, more preparation time than conventional lists, technical issues (audio, projector, internet), and difficulty visually representing abstract words.

This indicates materials need pedagogical control: visual, audio, and video elements should support rather than distract from meaning, and should be combined with active tasks so students use, not just watch, vocabulary.

2. Discussion

The findings show multimodal materials significantly improved vocabulary mastery, with the experimental group outperforming control on posttest and delayed posttest, supporting multi-modal presentation as more effective than single-mode instruction.

This aligns with dual coding theory: verbal-plus-visual encounters create multiple memory traces, so a word paired with a picture, pronunciation model, and context is processed more deeply than text alone (Paivio, 1991; Sadoski, 2005).

It also aligns with multimedia learning theory (Mayer & Moreno, 2003): pictures clarified meaning, audio provided pronunciation, videos created context, and sentences showed usage—working together rather than as decoration.

Strong gains in meaning recognition and contextual use suggest multimodal materials moved students beyond translation-based recognition toward deeper understanding of word use, consistent with Chun and Plass (1996), Akbulut (2007), Yoshii and Flaitz (2002), and Abraham (2008).

Pronunciation gains matter since students may know a meaning yet hesitate to use a word without knowing its pronunciation; audio-supported materials addressed this, consistent with prior audiovisual input research (Montero Perez et al., 2014; Peters, 2019; Peters & Webb, 2018; Sydorenko, 2010; Teng, 2019).

Retention in the delayed posttest suggests repeated, varied exposure across modes consolidates vocabulary knowledge, consistent with prior findings (Ramezanali & Faez, 2019; Rodgers & Webb, 2020; Teng, 2022; Teng, 2023).

Qualitative data deepen this picture: visual anchoring supported meaning, auditory reinforcement supported pronunciation, contextual meaning-making supported use, repeated exposure supported retention, and engagement supported motivation.

However, materials are not automatically effective—some students were distracted by visual or video elements, so teachers must guide attention and select materials for clarity, coherence, signaling, and relevance.

The teacher's role remains central: materials must be mediated through introducing vocabulary, guiding attention to visual clues, modeling pronunciation, and providing practice, or students may only passively consume content.

This has assessment implications: vocabulary assessment should include meaning recognition, spelling, pronunciation, and contextual use rather than translation alone, since spelling improved less and needs added focused practice.

Multimodal instruction should be implemented progressively—starting with simple word-picture-audio combinations before more complex video and dialogue tasks—selecting modes that fit the target word type (pictures for nouns, videos for verbs, examples for abstract words).

The findings also suggest multimodal materials support differentiated instruction, benefiting students with varied learning preferences and proficiency levels through pictures, audio, and contextual tasks.

Overall, multimodal materials' effectiveness lies in integrating verbal, visual, auditory, and contextual supports; when carefully designed and teacher-mediated, they transform vocabulary learning from memorization into meaningful processing.

D. Conclusion

This study examined multimodal learning materials' role in EFL vocabulary mastery through a mixed-method design. The experimental group improved from a pretest mean of 18.72 to a posttest mean of 31.28, versus 18.56 to 24.06 for control, with a significant difference, $t(62) = 6.25$, $p < .001$, and large effect size, Cohen's $d = 1.56$; retention was also better in the experimental group. Qualitative findings showed mastery was supported through visual anchoring, auditory reinforcement, contextual meaning-making, repeated exposure, and engagement, though careful design is needed to avoid distraction and overload. Multimodal learning materials can improve vocabulary mastery when relevant, structured, and aligned with learning objectives. Future research should examine this across grade levels, proficiency levels, vocabulary types, and digital platforms.

Declaration of AI-Assisted Technologies

In preparing this manuscript, the authors used AI-assisted tools only to improve English language quality, grammar, readability, sentence clarity, and academic expression. The AI tools were not used to generate research data, conduct data analysis, produce findings, or draw conclusions. All conceptual arguments, data interpretation, methodological decisions, and scholarly conclusions remain the full responsibility of the authors. The authors carefully reviewed, edited, and verified all AI-assisted outputs to ensure the accuracy, integrity, and originality of the final manuscript.

Bibliography

- Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, 21(3), 199-226. doi: 10.1080/09588220802090246
- Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology*, 40(1), 23-31. doi: 10.1111/j.1467-8535.2007.00800.x
- Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science*, 35(6), 499-517. doi: 10.1007/s11251-007-9016-7
- Bisson, M. J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *The Quarterly Journal of Experimental Psychology*, 68(7), 1306-1326. doi: 10.1080/17470218.2014.979211
- Çekiç, A. (2024). Incidental L2 vocabulary learning from audiovisual input: The effects of different types of glosses. *Computer Assisted Language Learning*, 37(5-6), 1135-1164. doi: 10.1080/09588221.2022.2062004
- Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80(2), 183-198. doi: 10.1111/j.1540-4781.1996.tb01159.x
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Annals of Family Medicine*, 11(3), 213-215. doi: 10.1370/afm.1330

- Hirschel, R., & Fritz, E. (2013). Learning vocabulary: CALL program versus vocabulary notebook. *System*, 41(3), 639-653. doi: 10.1016/j.system.2013.07.016
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26. doi: 10.3102/0013189X033007014
- Lee, H., & Lee, J. H. (2022). The effects of captioned videos on EFL learners' vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, 35(5-6), 1003-1026.
- Li, W., Yu, J., Zhang, Z., & Liu, X. (2022). Dual coding or cognitive load? Exploring the effect of multimodal input on English as a foreign language learners' vocabulary learning. *Frontiers in Psychology*, 13, 834706. doi: 10.3389/fpsyg.2022.834706
- Liu, Y. T., & Todd, A. G. (2014). Dual-modality input in repeated reading for foreign language learners with different learning styles. *Foreign Language Annals*, 47(4), 684-706. doi: 10.1111/flan.12113
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43-52. doi: 10.1207/S15326985EP3801_6
- Montero Perez, M., Peters, E., & Desmet, P. (2014). Is less more? Effectiveness and perceived usefulness of keyword and full captions in L2 listening comprehension and vocabulary acquisition. *ReCALL*, 26(1), 21-43. doi: 10.1017/S0958344013000256
- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45(3), 255-287. doi: 10.1037/h0084295
- Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, 53(4), 1008-1032. doi: 10.1002/tesq.531
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551-577. doi: 10.1017/S0272263117000407
- Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, 90(1), 25-36. doi: 10.1037/0022-0663.90.1.25
- Ramezanali, N., & Faez, F. (2019). Vocabulary learning and retention through multimedia glossing. *Language Learning & Technology*, 23(2), 105-124.
- Rodgers, M. P. H., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *ITL - International Journal of Applied Linguistics*, 171(2), 191-220. doi: 10.1075/itl.18034.rod
- Sadoski, M. (2005). A dual coding view of vocabulary learning. *Reading & Writing Quarterly*, 21(3), 221-238. doi: 10.1080/10573560590949359
- Samur, Y. (2012). Redundancy effect on retention of vocabulary words using multimedia presentation. *British Journal of Educational Technology*, 43(6), E166-E170. doi: 10.1111/j.1467-8535.2012.01320.x

- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363. doi: 10.1177/1362168808089921
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50-73.
- Teng, F. (2019). Maximizing the potential of captions for primary school ESL students' comprehension of English-language videos. *Computer Assisted Language Learning*, 32(7), 665-691. doi: 10.1080/09588221.2018.1532912
- Teng, M. F. (2022). Incidental L2 vocabulary learning from viewing captioned videos: Effects of learner-related factors. *System*, 105, 102736. doi: 10.1016/j.system.2022.102736
- Teng, M. F. (2023). The effectiveness of multimedia input on vocabulary learning and retention. *Innovation in Language Learning and Teaching*, 17(3), 738-754. doi: 10.1080/17501229.2022.2131791
- Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.
- Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO Journal*, 20(1), 33-58.