The Stabilizing Effect of Multimedia on ESL Spelling Retention: A Mixed-Methods Study

Rhio P. Pandayan
Foreign Languages Department, Prathomsuksa Thammasat School, Thailand, and Graduate School, University of Perpetual Help System Dalta, Philippines
E-mail: rhio.pandayan@perpetual.edu.ph   ORCID ID: https://orcid.org/0000-0002-2770-8653

Abstract

Background and Aim: This research was conducted to stabilize the spelling retention of English as a Secondary Language (ESL) learners through a multimedia intervention combining weekly short videos that featured vocabulary spelling, pronunciation, definitions, and photos. This, with notebook writing and teacher-guided practice among grade 5 students.

Materials and Methods: The Action research paradigm guided the implementation of the intervention, encompassing planning, implementation, observation of outcomes, and reflection on results. Using an explanatory sequential design, quantitative data from 20 weeks of spelling scores from weekly assessments were analyzed and then explained further by the qualitative data from student focus groups, interviews, and teacher observation.

Results: Quantitatively, students' spelling performance remained stable with an average score of 7.40 out of 10 (SD = 0.68). Minimally, scores fluctuated, which indicates consistent retention rather than progressive improvement. The stable scores suggest that while the multimedia intervention maintained prior knowledge, additional strategies may be required to foster further improvement. The qualitative themes are varying engagement patterns, different multimedia design preferences, mixed perceived efficacy and learning outcomes, support systems, and content relevance and accessibility.

Conclusion: Multimedia interventions effectively stabilized spelling retention in ESL learners but did not indicate improved performance. This implies that there is a necessity for better optimization of multimedia with the Cognitive Theory of Multimedia Learning (CTML) by syncing text and visuals, reducing text density, and segmenting videos. Support for academically struggling students and exploring engagement strategies should be increased. More implications are for consistently using themed units for relevance, gamification for heightened motivation, and training educators on CTML and hybrid learning methods for proper and consistent application of the intervention. The connection between independent practice and guided support proved crucial to stabilizing the student performance.

Keywords: Action Research, ESL Learners, Multimedia Intervention, Scaffolded Learning, Spelling Retention

Introduction

Spelling retention is one of the foundational skills for English as a Second Language (ESL) learners because it directly influences reading comprehension, writing proficiency, and overall academic success (Reed et al., 2016). However, ESL learners often face unique challenges, such as limited exposure to English outside the classroom and difficulties with orthographic mapping—the ability to connect spoken sounds to written letters (Zitha, 2024). Also, traditional methods of vocabulary instruction, such as rote memorization and flashcards, have been criticized for their lack of engagement and failure to address the multimodal nature of language learning (Mayer, 2009).

In recent years, multimedia tools have emerged as a promising solution to these challenges. According to Mayer's Cognitive Theory of Multimedia Learning (2009), combining visual and auditory stimuli enhances learning by leveraging dual-channel processing. For ESL learners, multimedia interventions such as videos featuring vocabulary, spelling, pronunciation, definitions, and contextual photos can provide a richer, more immersive learning experience. Studies have shown that multimedia glossing with embedded definitions, images, and audio in digital texts can significantly improve vocabulary retention and comprehension (Yunus et al., 2016; Akbulut, 2007). For example, a study by Mohsen and Balakumar (2011) found that ESL learners who used multimedia annotations outperformed their peers in vocabulary tests, with learners reporting higher levels of engagement and motivation.

Despite these promising findings, there is limited research based on integrating multimedia videos with traditional classroom practices, such as notebook writing and teacher-guided practice, for ESL learners. Notebook writing, which involves repeated exposure to words through copying, reinforces orthographic mapping (Graham & Santangelo, 2014). Similarly, Vygotsky (1978, as cited in Van de Pol et al., 2010) expounded that teacher-guided practice provides scaffolding and feedback, which are critical for language acquisition. Combining these methods with multimedia videos could create a synergistic effect that would enhance spelling retention.

Existing studies focus on either multimedia resources like glossed films or only traditional approaches such as notebook writing for ESL learning, but seldom combine the two. While multimedia interventions increase engagement (Teng, 2022) and traditional practice improves recall (Graham & Santangelo, 2014), few studies have looked into how combining them affects spelling and vocabulary stability. Bainbridge (2016) compared apps to paper; however did not create a hybrid model. This study fills that gap by combining weekly CTML-aligned videos alongside notebook exercises and teacher-led scaffolding, demonstrating how their integration can promote retention while stressing hurdles to deeper vocabulary gains.

This action research addresses this gap by investigating the effectiveness of a multimedia intervention that combines weekly short videos with notebook writing and teacher-guided practice. Furthermore, the findings of this research have the potential to inform best practices for ESL instruction, particularly in contexts where access to immersive English environments is limited. By demonstrating the effectiveness of multimedia tools in improving spelling retention, this study contributes to the growing body of literature on technology-enhanced language learning and provides actionable insights for educators seeking to adopt innovative teaching strategies.

Objectives

1. To examine the impact of a weekly video intervention featuring spelling, definitions, and visuals on students' spelling retention.

2. To investigate the qualitative factors, such as engagement and parental involvement, that influence the effectiveness of the video intervention.

Literature Review

Learning English can be difficult because spelling words correctly and using them properly in sentences can often be a challenge to children, especially those who learn ESL. A child is said to master a word when they can spell it, recognize it, and use it correctly both in writing and speaking. The ultimate objective of the English writing system goes beyond spelling and pronunciation, focusing on the ability to convey and understand meaning. English can be challenging because many words that are pronounced the same are spelled differently and carry distinct meanings. Although the fundamentals, such as reading and spelling words, are necessary. Due to the complexity of mastering English, the best way to gain a true understanding is by linking reading comprehension with spelling. The key to fluency in reading and writing in English lies in mastering the connections between letter combinations and the sounds they represent (Alshahrani, 2019).

Spelling is a key concept in this study. Spelling is a crucial foundational skill for individuals to express themselves through written communication, as opposed to oral communication. They further emphasize the importance of mastering spelling for developing reading fluency. Supports this by noting that learning to spell is not only important but essential for acquiring additional academic knowledge (Erion et al., 2009; Kosmac, 2010).

Vocabulary is words in a language, which include both individual words and phrases or groups of words that convey specific meanings, much like single words do. These lexical chunks have a crucial role in communication and in developing students' skills. A student's vocabulary strength directly impacts their ability to use more complex materials, enhancing their communication and understanding of others. The depth of a student's understanding of a word's meaning and usage can range from a shallow level which is simply recognizing and using a word in a basic way, to a deeper level, where they can use the word effectively in a variety of contexts (Carlo et al., 2004; Lessard-Clouston, 2013; Sánchez & Manchón, 2007).

In the context of ESL education, it is essential to identify strategies that promote long-term retention, especially in the areas of vocabulary acquisition and spelling. Retrieval practice has emerged as a significant technique in helping students master vocabulary and its correct spelling, both of which are crucial for achieving proficiency in ESL. However, despite its importance, there is a noticeable gap in adapting this technique specifically to the needs of ESL vocabulary learning and spelling. Traditional teaching methods often prioritize input-based approaches, focusing primarily on introducing new vocabulary and its spelling. In contrast, output-oriented techniques that encourage students to actively retrieve vocabulary and its spelling have been shown to result in better long-term retention and deeper comprehension of language concepts (Strong, 2023).

Multimedia can be a great output-oriented technique, spelling when combining both input, like watching and listening, and output, like speaking, writing, and applying vocabulary. Teng (2022) in his study examined the effectiveness of four multimedia input conditions for vocabulary learning and retention. The combination of definitions, word information, and video resulted in the best vocabulary gains and retention. Participants reported positive attitudes toward the dual multimedia mode. The study found that combining video with definitions and word information was more effective than audio-based methods, which caused cognitive overload. Video helped reinforce word comprehension, making it especially useful for long-term retention. Participants also felt more engaged and motivated when using videos, as they provided cultural context and mental images.

Srirabut (2023) conducted a study with grade 3 Thai learners aiming to improve the writing and spelling abilities of the students using multimedia, compare their skills before and after multimedia employment, and assess their satisfaction. The research involved 21 students and used learning plans, multimedia instruction, tests, and a satisfaction questionnaire. Results showed that students' performance exceeded the 80/80 threshold, with significant improvement in writing and spelling. Students also expressed high satisfaction with the multimedia learning activities, with a satisfaction score of 2.92.

Furthermore, Bainbridge (2016) examined whether using multimedia through apps or traditional paper methods was more effective for spelling learning among 18 children aged 9-10. The results showed that while children preferred using technology for reasons like repetition and instant feedback, they performed better when using paper, though the difference was not statistically significant. The method used first appeared to influence performance, suggesting that practice effects from the first method may have impacted the results. Many children reported that using apps was engaging and helpful, but some felt that a combination of both paper and technology would be most beneficial. The study highlights the value of integrating both methods in teaching spelling to leverage the advantages of each approach.

Similarly, the current study focused on using a combination of multimedia tools to reinforce vocabulary learning. In this case, students engage with vocabulary by copying words and their meanings, followed by practicing their spelling through videos. These videos provide an additional layer of learning, offering both visual and verbal cues to support understanding and retention of the

vocabulary. Just as Bainbridge (2016) noted that a combination of paper and technology might be most effective, this study also acknowledged the potential benefits of blending these strategies.

An investigation of Alobaid (2021) on how YouTube's enhanced caption features can foster writing accuracy through multimedia learning. Over five months, ESL learners utilized these tools to compare their written output with target language input, hypothesizing that increased visual salience would improve error recognition and self-correction. Mixed-methods analysis revealed significant post-intervention gains in writing accuracy, with qualitative data highlighting learners' reliance on captions for syntactic and lexical refinement. A moderate-strong correlation emerged between caption usage frequency and accuracy improvement, though limited by sample size. Results affirmed that ICT tools like customizable captions can amplify noticing and language restructuring, positioning YouTube as viable for scaffolding writing development in ESL contexts.

Correspondingly, the current study's multimedia videos featuring synchronized spelling animations, definitions, and visuals can help stabilize retention by directing learner attention to orthographic patterns, though design refinements such as adjustable text salience could further amplify vocabulary gains.

Furthermore, a second-order meta-analysis examined the impact of glosses on second language vocabulary learning, synthesizing results from seven meta-analyses covering 136 studies. The findings indicate a medium overall effect size for using glosses, with beginner-level students benefiting the most, showing a large effect size. Recognition tests yielded higher effect sizes than other vocabulary tests. The study also found that multimedia glosses were more effective in expository texts than in narrative ones, and single-mode glosses outperformed multi-mode glosses (Mahdi, Mohsen, & Almanea, 2024). This highlights the potential value of simplifying video content to avoid cognitive overload. The analysis also suggests beginners benefit most from glosses, making your focus on elementary ESL students particularly fitting. Furthermore, the preference for recognition tests over recall tests and the efficacy of expository texts provide useful insights for shaping your video content and assessment methods.

In brief, the related studies and literature underscore the challenges students face in mastering English spelling and vocabulary, emphasizing the importance of connecting spelling skills with reading comprehension for deeper language understanding. Research claims that multimedia approaches, particularly those combining visual, auditory, and written elements, effectively enhance vocabulary retention and spelling proficiency. Studies by Teng (2022) and Srirabut (2023) demonstrate that video-based learning strategies improve student engagement, comprehension, and long-term memory. Additionally, Bainbridge (2016) stresses the benefits of blending multimedia with traditional writing methods, aligning with this study's approach of pairing vocabulary videos with student notebooks and guided weekly practice. By combining multimedia instruction with active learning strategies like writing and teacher support, this study aimed to enhance students' spelling retention in a structured and effective manner.

Theoretical Framework

Drawing on Mayer's Cognitive Theory of Multimedia Learning (MCTML), the intervention aligned with principles of dual-channel processing, that is, visual plus auditory input, and active learning as depicted in Figure 1 (Malamed, 2024; Mayer, 2009).
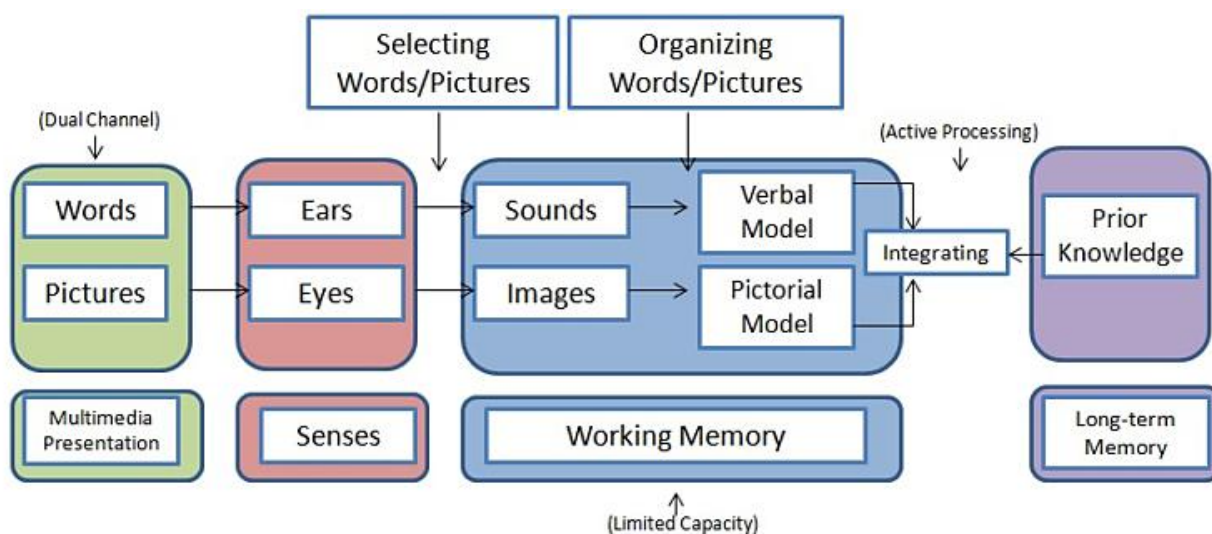
Figure 1. Mayer's Cognitive Theory of Multimedia Learning Model (Malamed, 2024)

MCTML has three foundational assumptions. The first is based on Dual Coding Theory (Paivio, 1986) that supports multimedia videos leveraging dual channels, in here, by presenting vocabulary is presented through verbal and visual channels. The second assumption is that there is limited capacity of working memory. Hence, in this intervention, short and focused videos prevent cognitive overload by segmenting content into weekly chunks and synchronizing elements by aligning visuals (photos) with spoken or written words. Third is active processing, wherein this intervention promotes active learning through notebook writing and teacher-guided practice.

Table 1. MCTML Cognitive Processes in The Multimedia Videos, Notebook Writing, and Teacher-Guided Practice

Cognitive Process	Application in the Intervention
1. Selecting Relevant Words	Videos emphasize key vocabulary (e.g., bolded text, slow pronunciation).
2. Selecting Relevant Visuals	Photos and animations are intentionally paired with words (e.g., "sculpture" shown as clay sculptures).
3. Building Verbal Models	Definitions and pronunciations are repeated across videos and reinforced in teacher practice.
4. Building Pictorial Models	Consistent use of thematic visuals (e.g., "ocean" paired with waves, fish, and sand imagery).
5. Integrating Models	Notebook writing and guided practice bridge verbal (spelling) and pictorial (photos) knowledge.

Malamed (2024) noted that there are five cognitive processes detailed in Table 1 that are required for multimedia learning, namely: selecting relevant words, selecting relevant visuals, building verbal models, and building pictorial models. Table 1 shows how the intervention in this study directly supports the five cognitive processes.
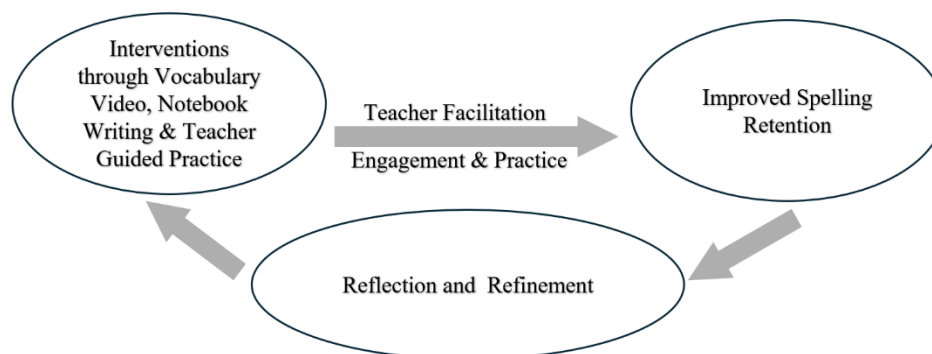
Conceptual Framework

Figure 2. Conceptual Framework

Figure 2 illustrates the process of the study and the relationship of concepts and actions concerning the goal of improved spelling retention. Towards this aim, interventions such as vocabulary video (with words, definitions, photos, and audio), notebook writing activities (students write words and definitions), and weekly teacher guidance (practice and feedback for about 2 hours a week). A regular and quality student engagement and practice, along with strategic and effective teacher facilitation of the intervention, are assumed to result in improved outcomes. As this study followed the flow of action research, it was ensured that quality reflection and refinement recommendations from the student data would be brought to develop an even better intervention.

Methodology

1. Research Design

This study was conducted within the bounds of action research. This paradigm strongly emphasizes taking action to improve practice. Its primary goal is to bring about meaningful changes in teaching methods or strategies. This approach follows a recurring cycle with four key stages: planning, implementing the plan, observing outcomes, and reflecting on results. Insights gained from one cycle inform the next, promoting continuous improvement (Carr & Kemmis, 1986; McNiff, 2013). In this undertaking, it becomes valuable as it could deliver practical improvements that benefit both educators and students directly in learning a second language.

Furthermore, an explanatory sequential design was employed. The data were collected in the quantitative phase before moving on to the qualitative phase (Creswell, 2011). The qualitative phase is used to explain the findings from the first quantitative phase, as well as to explain outliers that are not completely consistent with the obtained data (Abu Sayed Toyon, 2021).

As to the quantitative measure in this study, the weekly spelling scores (20 weeks) were analyzed using descriptive statistics. Meanwhile, the student focus group was used at the same time to gather qualitative data. To collect a range of perspectives, behaviors, and motivations, a focus group was used for it is designed to gain an in-depth understanding of the students' experiences and perceptions through a conversation (Nyumba, et. al, 2018). For exploring the perspectives of primary-level students, it was ensured that age-appropriate language is used and discussion is paced appropriately (*Center for Community College Student Engagement*, 2013).

2. Participants

There were 22 Grade 5 students involved in the weekly spelling test. They are aged 9 to 11 from a government primary school in Pathumtani, Thailand. The number of students is chosen through

convenience sampling since they are the direct students of the researcher. This situation enhanced closed observation and consistent conduct of the intervention.

Action research applies non-probability sampling in its conduct; hence, convenience samples were resorted to. This emphasizes the practicality of sample sizes, which are frequently smaller and chosen purposefully, favoring in-depth exploration and development of a specific context over generalizability (Applied Doctoral Center, 2025).

3. Intervention

The intervention was mainly through weekly videos uploaded to student and parent online groups. This multimedia features word pronunciation in audio, spelling animation, definitions, and contextual photos. For instance, the word "photosynthesis" is paired with an image, a repeated voiceover on how the word is read, and the definition.

These multimedia elements in the intervention were chosen because they align with multimedia learning principles. Mayer's (2017) study revealed that combining visuals and audio supports the multimedia principle, while synchronizing these elements follows the temporal contiguity principle, improving retention. The segmented video format aligns with the segmenting principle and incorporates both auditory and visual elements, leveraging the modality principle, thus enhancing comprehension. These evidence-based strategies support the intervention's elements in improving ESL students' spelling.

Furthermore, vocabulary notebooks were utilized to support the intervention, with students writing the words and definitions completely and accurately. Consequently, a teacher-guided practice happened 2 hours weekly. The guided session included word games, reading and spelling, and word identification through photo or definition, writing the spellings on the board.

Graham et al. (2002) provided empirical evidence that supplemental spelling instruction, like the structured vocabulary notebook activities and teacher-guided practice, can effectively enhance spelling skills and related literacy outcomes. Their study demonstrated that focused spelling interventions — including frequent, targeted practice — lead to significant improvements in spelling, writing fluency, and reading skills. This is parallel to the strategies of the current study, where students engaged in word games, spelling practice, and vocabulary writing under teacher guidance. There is also a reinforced value of dedicating structured instructional time of 2-hour weekly sessions to improve spelling retention.

4. Data Collection

The students' scores in their weekly spelling test were the main numerical data. Before getting the scores, videos were meticulously prepared that are based on their levels and readiness, and relevance to the topics in their subjects. Words, photos, and definitions were chosen and synthesized together in a video, and then added with a teacher's voiceover on how the words are read clearly and correctly. They were then sent to the online groups of students and parents with notes to write the words and their definitions in the notebooks, as well as practice reading and writing the words for mastery and understanding. Videos were sent every Friday and then collected on Mondays of each week. Videos were played in the class for practice. The teachers also used the words for word game practice, board writing, students' sentences, reading chorus of the words and definitions, and other strategies that would allow students to practice the words during weekdays. These teacher-guided practices were conducted twice a week, with 1 hour per session. Spelling tests are then taken by the students on Fridays.

To explore more of the factors affecting the use of this multimedia in the students' spelling retention, students are subjected to an interview through a student focus group. Semi-structured interviews were employed to explore engagement with videos. Students were asked about their engagement in the weekly videos, such as how often they watch the weekly spelling videos, or did they

watch them more than once. Learning strategies and habits were also inquired such as how the students used the videos to prepare for their weekly spelling tests, such as pausing the video to write words down. Through questions asking how they think the videos helped them spell better than before or if they see any parts of the videos confusing, their perceived benefits and challenges were asked. Home support was also inquired about by asking if their parents or guardians watched the videos with them, or how that helped or did not help. The focus groups ended with some open-ended reflections, such as asking them if they could change one thing about the videos to make them better, what would it be?

Data were corroborated with teacher observations. In here, live data were collected with the help of senses of observation in the controlled or naturalistic situations of the occurrence of events - the classroom in this study. The researcher acted as a complete participant observer, participated with the subjects, and had full observation in the classroom (Satapathy, 2019).

To protect participants, confidentiality was ensured through anonymization of all qualitative data, such as pseudonyms replacing student names in transcripts, and recordings and files were stored securely on password-protected devices. Focus group protocols emphasized voluntary participation, allowing students to skip questions or withdraw without penalty. Sessions were conducted in a non-threatening classroom environment by the teacher-researcher who avoided leading questions and prioritized student comfort. To mitigate peer influence, ground rules such as "no laughing at others' answers" were established, and group sizes were limited to 4–6 students. Debriefing conversations reaffirmed that responses would not affect academic evaluations. These safeguards align with ethical guidelines for research involving minors (BERA, 2018; APA, 2020), ensuring respect, dignity, and psychological safety throughout the action research.

5. Data Analysis

Quantitative data consisted of 20 weeks of spelling scores (scale: 0–10), which were analyzed using descriptive statistics (mean, standard deviation, range). Thematic analysis was utilized for the focus group transcriptions. This method involves familiarization with the transcript, coding, topic generation, theme review, theme definition and naming, and writing up. This process is delineated in Table 2. The process centers on iterative refinement aimed at ensuring a comprehensive exploration of the students' perception under investigation (Caulfield, 2023). Initial codes and themes emerged from the data itself, allowing for a rich exploration of patterns, commonalities, and variations in the experiences of students with the intervention (Naeem et al., 2023).

Table 2. Sample Coding Process Leading to the Theme of Engagement Patterns: Frequency and Autonomy

Steps	Details
Coding	*Frequency of Video Use —* "I watch the spelling 4 times a week" (A1); "I watch the video the whole day" (B1); "I watch the video around 2 times" (F1)
	Practice Independence — "I practice by myself and no none assists me" (A4); "My mom helps me" (C3); "My aunt and I practice" (B3)
	Access and Ease of Use — "I don't have any problem accessing the videos" (A6); "I sometimes have a problem accessing the video when the Wi-Fi is off" (D4)
	Video Preferences — "I would prefer the teacher giving the words and meanings on the board" (C6); "I think it is okay if there is no video" (F5)
Topic Generation	*Video Viewing Frequency —* Ranged from watching daily to only once or twice a week.
	Autonomy in Practice — Some students practiced independently, while others relied on family support.

Steps	Details
	Challenges and Preferences — Students requested faster transitions between words, simpler backgrounds, and clearer definitions.
Theme Review	Combined topics revealed a pattern of students' engagement levels varying by practice frequency and autonomy. This aligned with patterns of self-motivation and reliance on support.
Theme Naming	*"Engagement Patterns: Frequency and Autonomy"* — Summarizing varying degrees of student engagement, from frequent independent practice to occasional reliance on family support.

To enhance rigor, triangulation was applied by comparing quantitative trends with qualitative themes, as recommended for mixed-methods action research (Creswell & Plano Clark, 2018). In here, the quantitative spelling scores and qualitative data from focus groups, interviews, and observations were integrated to validate findings and provide a comprehensive understanding of the intervention's impact. The researcher's focus was on the complementarity of the data, where qualitative insights revealed the whys of what the quantitative data confirmed.

Results

1. Quantitative Results

The descriptive statistics shown in Table 3 summarize the central tendency and variability of students' weekly spelling scores over the 20-week intervention. The mean score across all weeks was 7.40 out of 10 ($SD = 0.68$), indicating stable performance with minimal fluctuation. The median score of 7.50 closely aligns with the mean, suggesting a symmetric distribution of scores. The narrow range of 6 to 8 and small standard deviation ($SD = 0.68$) further underscores the consistency in student performance, with no extreme outliers or significant deviations. A 95% confidence interval for the mean (7.08–7.72) confirms that scores remained tightly clustered around the central tendency, reflecting neither improvement nor decline over time.

Table 3. Descriptive Statistics of Weekly Spelling Scores (Scale: 0–10) Across the 20-Week Intervention Period (n = 20 Weeks)

Statistic	Value	Statistic	Value
Mean (SD)	7.40 (0.68)	Minimum	6
Median	7.50	Maximum	8
Range	6 – 8	95% CI for Mean	7.08 – 7.72

Notes: SD = Standard Deviation; 95% CI = Confidence Interval

These findings align with prior studies noting that multimedia interventions often stabilize learning outcomes by reinforcing prior knowledge rather than driving incremental gains (Mayer, 2009). The limited variability ($SD < 1.0$) also suggests that the intervention was uniformly accessible to students, with no subgroup experiencing disproportionate challenges.

Figure 3 illustrates the weekly mean spelling scores over the 20-week intervention, plotted as a line graph to visualize temporal trends. Scores fluctuated moderately between 6 and 8, with a pronounced dip in Week 2 ($M = 6.0$) followed by recovery to baseline levels ($M \approx 8.0$) in Weeks 3–5.
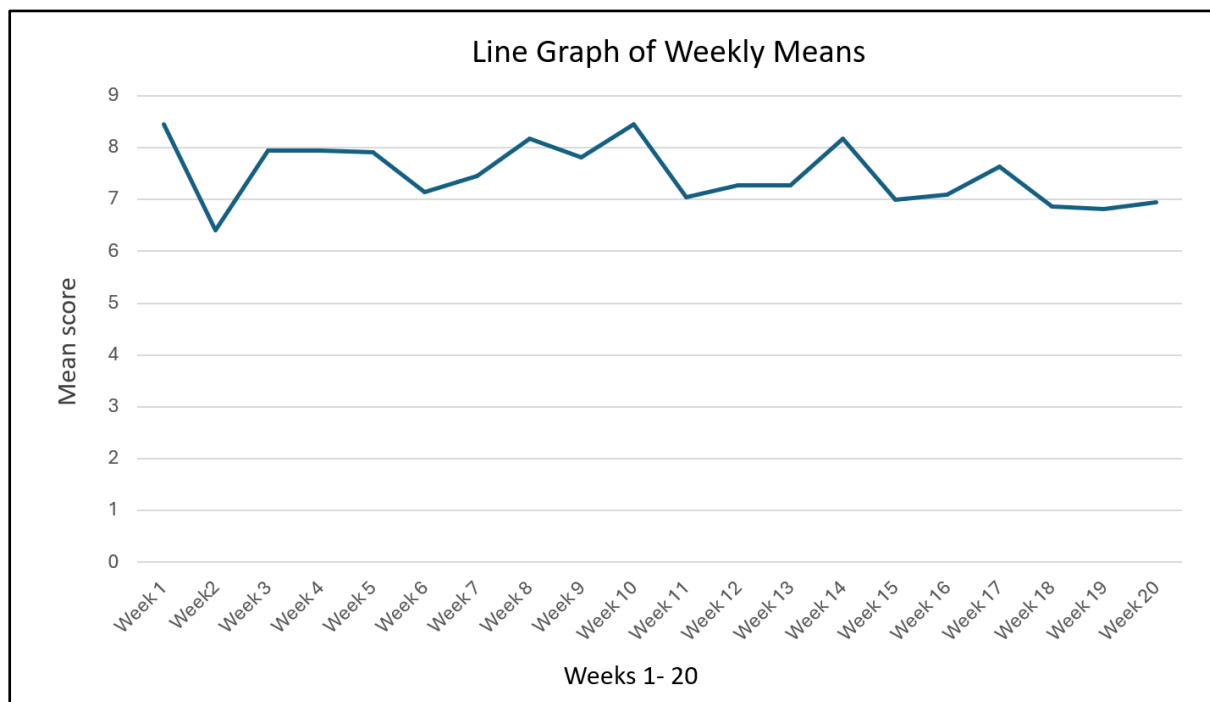
Figure 3. Mean Spelling Scores (Out of 10) Over 20 weeks

Subsequent weeks (6–20) exhibited minor oscillations around the overall mean of 7.40, reflecting a plateau effect. The Week 2 decline may reflect initial adjustment challenges to the intervention, as students acclimated to video-based learning, while the rebound in Weeks 3–5 aligns with qualitative reports of increased engagement after iterative refinements to video content, like adding slow-motion pronunciation guides. The absence of a sustained upward trend, despite weekly video exposure, underscores the need for complementary strategies such as gamification to amplify retention (Yunus et al., 2016). This graph contextualizes the quantitative stability within the intervention's timeline, providing a visual narrative of consistency punctuated by transient variability.

Linking this to the qualitative data, the Week 2 decline (M = 6.0) can be associated with student critiques of slow pacing and desynchronized visuals/text in early videos. This, also, in the engagement patterns of the passive learners who struggled to adapt, as noted in teacher observations, leading to disengagement and lower scores. As to Week 3- 5 recovery, the rebound to baseline (M ≈8.0) corresponded to video refinements and increased teacher scaffolding with the guided practice sessions. The plateau in the graph reflected saturation effects among autonomous learners and persistent barriers for others. While some students maintained high engagement, others critiqued mismatched visuals, which fractured retention pathways.

2. Qualitative Themes

Below are the findings structured by emergent themes from thematic analysis, supported by direct quotations and aligned with theoretical frameworks (Braun & Clarke, 2006; Mayer, 2009).

2.1 Engagement Patterns: Frequency and Autonomy

Students exhibited varied engagement levels with the weekly videos. There were some students engaged deeply, with Student A watching videos *"4 times a week"* and Student B dedicating *"the whole day on Sunday."* The autonomous students achieved scores near the upper range of around 8/10. Others were showing passive engagement like Student C (*"two times a week"*), reported sporadic use, while Student F critiqued pacing (*"wait 3–5 seconds for texts"*). Furthermore, some students manifested non-engagement. A subset (Group Responses 9–11) admitted saying *"I don't watch the video"* or *"don't*

practice," which emphasizes motivational gaps. These passive and non-engaged learners scored closer to 6/10 but were buoyed by teacher-guided practice, preventing further decline. This scenario of variance can be explained by Self-Determination Theory (Ryan & Deci, 2000), where autonomy, such as Student A's self-practice, and scaffolding, like the parental support, influence engagement.

Furthermore, teacher observations noted that 5 to 6 students required repeated prompting to review spelling videos during class breaks. Also, two students avoided retrieving their notebooks, reflecting passivity and resistance. A transferred student with English proficiency 30 % below grade level struggled despite remediation efforts, underscoring gaps in differentiated instruction.

There was also observed strategic cramming where some students optimized their effort by reviewing only 1 or 2 days before Friday tests. A high-performing student admitted, *"I watch the videos on Thursday night because I remember better then,"* achieving 98% accuracy. This is aligned with the desirable difficulty theory (Bjork & Bjork, 2020), where shorter intervals between study and testing enhance recall but may limit long-term retention.

2.2 Multimedia Design Preferences

Students emphasized visual simplicity and pacing. As to visual appeal, images were praised by them. *"The most liked parts are the image,"* Student C expressed. However, there were also comments, such as text-heavy definitions, that caused disengagement. The student also articulated, *"Don't like the writing."* Some students responded about the pacing of the video, that slow transitions frustrated them (*"I must wait 3 to 5 seconds"* – Student F; Group Response 12).

The negative comments from the students manifest violations of Mayer's (2009) *temporal contiguity principle* – desynchronized visuals/text - increased extraneous cognitive load. Although visuals and text are presented simultaneously in the video, frustrations with waiting 3–5 seconds for transitions disrupted the flow of information and increased cognitive load. There was a delay between visuals and text/narration, making it harder for learners to connect the two.

Quantitatively, these flaws likely suppressed potential gains, as even motivated learners such as Student A, expressing 9.5/10 perceived efficacy, faced avoidable cognitive barriers.

2. 3 Perceived Efficacy and Learning Outcomes

There were manifestations of high and low efficacy of the intervention, resulting in mixed perceptions of the intervention's effectiveness. For instance, Student A rated videos *"9.5/10"*, crediting them for test preparation, while Student D noted *"7/10"* usefulness. Moreover, there was observed skepticism, such as Student F dismissing the videos, expressing that *"I can still spell without them."* Further criticism arose from inconsistencies in design alignment, as noted by another student who observed a disconnect between video content and assessments: *"The photos in the videos aren't consistently used in the tests,"* highlighting mismatches that created confusion.

These discrepancies in perceived efficacy align with Mayer's (2009) Coherence Principle from CTML, which asserts that learning is optimized when multimedia elements are directly relevant to instructional goals and free of extraneous or inconsistent material. When applied effectively, coherence strengthens retention by reducing cognitive load and fostering clear mental connections between visuals, text, and concepts. In this study, the high ratings from students like A likely reflect instances where the intervention adhered to this principle, such as videos that tightly synchronized photos, definitions, and narrated examples to create cohesive as well as memorable learning episodes.

Teacher observation pointed out last-minute efficacy where cramming correlated with high scores for some, qualitative data revealed this group relied heavily on rote memorization rather than deeper understanding. Also, an equity barrier is observed with the transferred student's challenges *("I don't know these words"),* presenting misalignment between the intervention and varying proficiency levels, exacerbating disengagement.

2.4 Support Systems: Scaffolding vs. Independence

Parental involvement is complementary factors as revealed by students such as B, C, and D relying heavily on familial support, with Student D noting, *"My mom helps me practice,"* a reflection of Vygotsky's (1978) concept of social scaffolding, where learning is mediated through collaborative interactions with more knowledgeable others. Conversely, Student A and participants in Group Responses 4–8 emphasized self-driven strategies, such as *"practice by myself alone,"* highlighting intrinsic motivation and autonomy aligned with self-determination theory (Ryan & Deci, 2000). On the other hand, for resistant learners, scaffolding was critical. Observations noted a few students illustrating dependency on external motivation.

These findings suggest that a blended approach, integrating guided parental or teacher support with opportunities for independent practice, may optimize learning outcomes. By balancing external scaffolding with autonomy, diverse learner preferences can be catered to, fostering both confidence in collaborative settings and self-efficacy in solo endeavors.

2.5 Content Relevance and Technical Accessibility

There are critical considerations for content relevance and technical accessibility in designing multimedia interventions. *"Math words this week, science next"* is a response that suggests a desire for contextually secure learning that bridges classroom subjects and multimedia content. Students in Group Responses 13–14 advocated for themed vocabulary aligned with their academic needs. This is a principle supported by Bransford et al. (2000), emphasizing the value of situating instruction in meaningful contexts. However, technical barriers emerged as a significant hurdle, exemplified by Student D's struggle with inconsistent connectivity (*"Wi-Fi is off"*). This situation features equity concerns in digital access (Selwyn, 2017). To address these two challenges, the intervention could integrate thematically organized vocabulary. For example, domain-specific terms for math, science, or social studies, in order to boost relevance while offering offline accessibility, such as downloading the MP4 files to mitigate connectivity disparities. This dual strategy would not only align with learners' curricular needs but would also ensure equitable participation, particularly for students in unstable digital environments.

Table 4. Synthesis of Themes

Theme	Key Insight	Representative Quote
Engagement	High variability is driven by autonomy/support.	*"I watch the whole day on Sunday."* – Student B
Design	Visuals favored; pacing/text critiqued.	*"Wait 3–5 seconds for texts."* – Student F
Efficacy	Mixed perceptions are tied to content alignment.	*"Photos don't help…not used in tests."* – Student F
Content Relevance	Demand for themed, practical vocabulary.	*"Words usable in daily conversation."* – Group

Table 4 shows the themes carrying three critical implications for optimizing multimedia interventions in ESL learning. First, design refinement is essential where there is a need to simplify visuals, accelerate pacing, and ensure that multimedia content directly aligns with assessments can address frustrations like Student F's critique of *"waiting 3–5 seconds for texts"* and mismatches between video content and tests. Second, equity measures, such as providing offline access, such as

downloadable videos, are vital to overcoming technical barriers like connectivity, ensuring all learners can engage with materials regardless of infrastructure limitations. Third, contextual alignment—integrating themed, practical vocabulary (e.g., *"words usable in daily conversation"*). This can boost relevance and retention, as advocated by Bransford et al. (2000). Together, these adjustments create a more inclusive, coherent, and engaging learning environment, bridging gaps between multimedia design, learner needs, and real-world application to enhance spelling retention and conceptual mastery among ESL students.

Discussion

The qualitative findings illuminated critical aspects, some are subtleties, in how ESL learners interacted with the multimedia intervention, contextualizing the quantitative stability in spelling scores and revealing why significant improvements were elusive at some points. Students' engagement variability from deep autonomy to passive or non-engagement aligns with Self-Determination Theory (Ryan & Deci, 2000). As observed by the teacher, the students with deep autonomy are those who are performing well in the class, especially in the subjects of English 4 Skills and Conversation, with 90% above average grade. Surprisingly, the students in groups 9-11 who have remarks of not watching the video are also in the group of performing students. This implies that the level of difficulty of vocabulary could be easy for them. Autonomy-driven learners thrived, while others required scaffolding, such as parental or home support. Students whose mothers and guardians are active in the parents' group and school-parent linkages are also the same parents who are supportive of the students' practice. These students perform well in class and on the spelling tests. This echoes Vygotsky's (1978) emphasis on social mediation as parents and guardians assist learners at home. Hence, there is a need for interventions that balance independent practice with guided support to accommodate diverse learner needs.

The engagement spectrum from autonomous learners to resistant students mirrors Self-Determination Theory's continuum of motivation. Autonomous learners, driven by intrinsic motivation, thrived within the intervention's structure, achieving stable, high scores. Conversely, students requiring external prompts operated under controlled motivation, which links to shallow, compliance-driven engagement. From here, there is a need to nurture autonomy, such as with student choice in video pacing, competence shown in tiered difficulty levels, and relatedness like peer collaboration, to transform passive compliance into enduring self-regulation.

Looking at the quantitative stability and qualitative critiques of slow pacing and desynchronized text/visuals, and Mayer's (2009) CTML, violations of the temporal contiguity principle and coherence principle disrupted dual-channel processing were observed. This forced learners to expend cognitive resources on reconciling disjointed inputs rather than encoding spelling patterns. Teacher observations underscored these results, with the students who crammed videos days before tests likely compensated for poor retention by leveraging short-term memory, a strategy that masks deeper encoding failures. This aligns with CTML's assertion that superficial engagement with multimedia yields fragile, context-bound knowledge. While the transferred students' struggles reflect a breach of the pre-training principle, without prior familiarity with vocabulary, the intervention's pacing and density overwhelmed their limited working memory capacity. Theoretical implication suggests that while CTML-compliant design stabilized retention through dual-channel reinforcement, inconsistencies in pacing and coherence stifled meaningful gains. Future interventions must prioritize segmenting, consistent synchronization of text-visual-audio alignment, and pre-training in the form of vocabulary priming for struggling learners to reduce cognitive load.

Mismatches between video content and assessments breached the coherence principle, fracturing retention pathways. As observed and practiced, the test of spelling used the words' definitions' and images together or separately. Sometimes, presentation of a combined image and text of the spelling test yields better results for students' spelling recall. These issues likely diluted the intervention's efficacy. Reasonably, inconsistent design impeded the integration of verbal and pictorial models, explaining the stagnant quantitative scores.

Equity and relevance further moderated outcomes. A connectivity barrier was observed. This amplified inequities. Also, there were demands for consistent and not so variegated themed vocabulary. This highlighted the gap between generic content and contextual relevance (Bransford et al., 2000). Students' call for practical, curriculum-aligned terms reflected the need for anchored instruction that can bridge the classroom and real-world application.
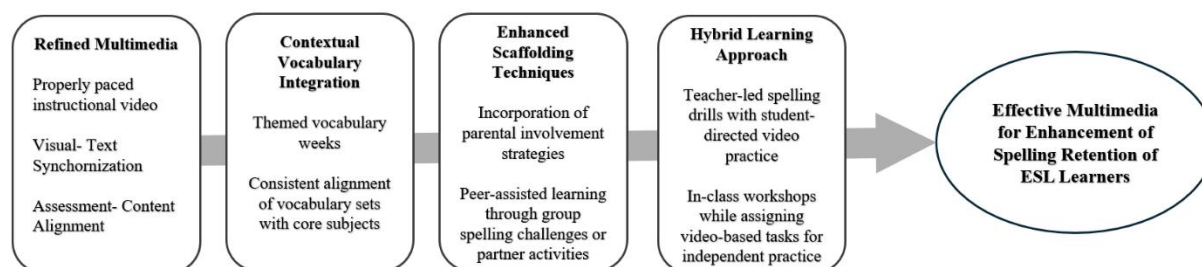
Knowledge Contribution

Figure 4. Effective Multimedia for Enhancement of Spelling Retention (ESL)

Figure 4 represents the synergistic plan for enhancing spelling retention. The enhanced outcome is achieved through four significant elements revealed in this study, namely Refined Multimedia, Contextual Vocabulary Integration, Enhanced Scaffolding Technique, and Hybrid Learning Approach.

The first element can be achieved by accelerating visual transitions through adjusting pacing in instructional videos to eliminate frustrating delays. Visual-text synchronization will also ensure visuals appear simultaneously with corresponding text or narration to align with Mayer's Temporal Contiguity Principle. As well, assessment-content alignment is necessary by incorporating key video visuals into spelling assessments to reinforce knowledge transfer.

Contextual vocabulary integration is key and can be done by introducing themed vocabulary weeks to provide meaningful context for new words. Alignment of vocabulary sets with core subjects. For instance, developing classroom projects where students use newly learned vocabulary in sentence-building activities or creative storytelling. In the practice of this intervention, vocabulary was extracted from the topics of different subjects that they encountered in previous lessons or will encounter in the following lessons. Although the vocabulary was based contextually on subjects that students learned, results of the study suggest that there must be uniformity in choosing the subjects where words are sourced such as a week having Math and Science vocabulary, and the following set of words about Arts and Music, or they may be equal number of words from each of the core subjects.

Enhanced scaffolding techniques are effective and attainable through collaborative approaches, such as incorporating parental involvement strategies in creating take-home spelling kits and providing parents with downloadable video resources and multilingual guides for better home support. It is also deemed essential to have peer-assisted learning through group spelling challenges or partner activities.

Finally, the hybrid learning approach is doable by combining teacher-led spelling drills with student-directed video practice to accommodate diverse learning preferences. For instance, using in-

class workshops for hands-on activities such as spelling puzzles, while assigning video-based learning tasks for independent practice. All these elements are not cyclical. These are deemed integrative and can be applied and observed in the process of implementing interventions such as multimedia to enhance the spelling and vocabulary of ESL learners.

Recommendation

1. Multimedia design can be optimized by aligning with CTML principles. In doing this, delays between text and visuals (temporal contiguity) must be eliminated, and assessment-photo consistency for coherence. For example, embedding key video visuals into spelling tests. Also, it is beneficial to reduce text density, use dynamic visuals, and segment videos into shorter, interactive chunks to sustain engagement.

2. In the students' writing in their notebooks and the teacher-guided practice, there should be closer attention to students perceived to have the most difficulty in writing and engaging. Other strategies can be explored by further research to enhance their engagement and hasten practice.

3. To ensure equity and accessibility, it is recommended to advise students to download the video from an online application to mitigate connectivity gaps. Also, offering multilingual guides for parents to strengthen home scaffolding in low-resource settings would be a helpful step.

4. Enhanced contextual relevance is important and can be suggestively applied by developing themed units (e.g., science or math vocabulary every week) that align with curricular goals and practical use. Although the vocabularies used in this study come from different subjects each week, students tend to prefer a single subject or field where vocabulary can be extracted for uniformity.

5. Integrating gamification, such as word quests, to motivate autonomous learners and reinforce retention will be beneficial.

6. Professional development on both topics will be beneficial for learners and teachers alike. Topics reflective of the study are on CTML-based design to avoid cognitive overload and leverage dual-channel processing effectively. Also, on the topic of Hybrid Learning Strategies, which blends teacher-guided practice with self-paced video modules to cater to both scaffolded and autonomous learners.

Conclusion

This study shifts ESL discourse by proving that hybrid methods, such as multimedia and traditional practice, stabilize retention, but not growth. From here, the research cautions against over-reliance on technology, urging designers to prioritize CTML rigor as well as equity. The use of multimedia interventions entails strategic refinement so as to drive meaningful gains. While the videos provided foundational reinforcement through dual-channel processing, design inconsistencies, equity barriers, and contextual misalignment limited their transformative potential. The interplay of autonomy and scaffolding further accentuated the need for adaptable, learner-centered approaches.

Cohesive design, equitable access, and curriculum-anchored content to bridge the gap between technology and pedagogy may be prioritized by future research. By concentrating and solving issues in these dimensions, educators can truly transform multimedia tools from supplementary aids into catalysts for deeper language acquisition. Further research should explore longitudinal impacts of these refinements and their scalability across diverse socioeconomic contexts, ensuring that innovation in ESL education remains inclusive and cognitively grounded.

The intervention's success in stabilization paradoxically revealed multimedia's ceiling – without addressing cognitive load and motivation, tools sustain, but cannot transform learning. Future research can break this ceiling by testing gamified CTML tools, offline delivery models, and

proficiency-tiered content. Future research can also pair these with longitudinal hybrid studies to map pathways from stability to mastery.

References

Abu Sayed Toyon, M. (2021). View of explanatory sequential design of mixed methods research: Phases and challenges. *Business and Social Science*. https://www.ssbfnet.com/ojs/index.php/ijrbs/article/view/1262/971

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. https://doi.org/10.1007/s11251-007-9016-7

Alobaid, A. (2021). ICT multimedia learning affordances: Role and impact on ESL learners' writing accuracy development. *Heliyon*, *7*(7), e07517. https://doi.org/10.1016/j.heliyon.2021.e07517

Alshahrani, H. A. (2019). Strategies to improve English vocabulary and spelling in the classroom for ELL, ESL, EO, and LD students. *International Journal of Modern Education Studies*, *3*(2). https://doi.org/10.51383/ijonmes.2019.41

American Psychological Association. (2020). *Ethical principles of psychologists and code of conduct*. https://www.apa.org/ethics/code

Applied Doctoral Center. (2025). Action research resource. http://resources.nu.edu/c.php?g=1013605&p=8464648

Atapathy, S. (2019). Observation as a tool for collecting data. *International Journal of Multidisciplinary Educational Research*, *8*(5[1]), 152. https://www.researchgate.net/publication/371475537

Bainbridge, F. (2016). Apps-olutely brilliant: Can technology improve the spelling ability of primary school children? *Language Studies Working Papers*, *7*, 36–49. https://www.reading.ac.uk/elal/-/media/project/uor-main/schools-departments/elal/lswp/lswp-7/elal_6_bainbridge.pdf

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn*. National Academy Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

British Educational Research Association [BERA]. (2018). *Ethical guidelines for educational research* (4th ed.). https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018

British Educational Research Association [BERA]. (2018). *Ethical Guidelines for Educational Research* (4th ed.).

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., & White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, *39*(2), 188–215.

Carr, W., & Kemmis, S. (1986). *Becoming critical: Education, knowledge and action research*. Falmer Press. https://doi.org/10.4324/9780203496626

Caulfield, J. (2023, June 22). How to do thematic analysis. *Scribbr*. https://www.scribbr.com/methodology/thematic-analysis

Center for Community College Student Engagement. (2013). *Students speak—Are we listening? A focus group toolkit*. The University of Texas at Austin, Community College Leadership Program. https://www.ccsse.org/docs/SSBTN_FG_Handouts.pdf

Creswell, J. W. (2011). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.

Erion, J., Davenport, C., Rodax, N., Scholl, B., & Hardy, J. (2009). Cover-copy-compare and spelling: One versus three repetitions. *Journal of Behavioral Education*, *18*, 319–330.

Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing*, *27*(9), 1703–1743. https://doi.org/10.1007/s11145-014-9517-0

Graham, S., Harris, K. R., & Chorzempa, B. F. (2002). Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology*, *94*(4), 669–686. https://doi.org/10.1037/0022-0663.94.4.669

Kosmac, A. M. (2010). *Cover, copy, and compare: The analysis of a self-managed intervention to increase spelling accuracy across task and time* [Doctoral dissertation, University of Cincinnati]. ProQuest Dissertations and Theses Global.

Lessard-Clouston, M. (2013). *Teaching vocabulary*. TESOL International Association.

Mahdi, H. S., Mohsen, M. A., & Almanea, M. (2024). Multimedia glosses and second language vocabulary learning: A second-round meta-analysis. *Acta Psychologica*, *248*, 104341. https://doi.org/10.1016/j.actpsy.2024.104341

Malamed, C. (2024). Applying Richard Mayer's Cognitive Theory Of Multimedia Learning. *The eLearning Coach*. https://theelearningcoach.com/learning/learning-theory-and-multimedia/

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, *33*(5), 403–423. https://doi.org/10.1111/jcal.12197

McNiff, J. (2013). *Action research: Principles and practice* (3rd ed.). Routledge. https://doi.org/10.4324/9780203112755

Mohsen, M. A., & Balakumar, M. (2011). A review of multimedia glosses and their effect on L2 vocabulary acquisition in CALL literature. *The Turkish Online Journal of Educational Technology*, *10*(2), 247–258.

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods*, 1–18. https://doi.org/10.1177/16094069231205789

Nyumba, T., Wilson, K., Derrick, C. J., & Mukherjee, N. (2018). The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution*, *9*(1), 20–32. https://doi.org/10.1111/2041-210X.12860

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.

Reed, D. K., Petscher, Y., & Foorman, B. R. (2016). The contribution of vocabulary knowledge and spelling to the reading comprehension of adolescents who are and are not English language learners. *Reading and Writing*, *29*(4), 633–657. https://doi.org/10.1007/s11145-015-9619-3

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sánchez, A., & Manchón, R. M. (2007). Research on second language vocabulary acquisition and learning: An introduction. *International Journal of English Studies*, *7*(2), vii–xvi.

Selwyn, N. (2017). *Education and technology: Key issues and debates* (2nd ed.). Bloomsbury Publishing.

Srirabut, P. (2023). The development of writing and spelling abilities of PrathomSuksa 3 students using multimedia. *Journal for Developing the Social and Community*, *10*(2), 195–206. https://so03.tci-thaijo.org/index.php/rdirmu/article/view/262335

Strong, B. (2023). Retrieval, repetition, and retention: Unveiling vocabulary acquisition strategies for ESL learners. *Art Human Open Access Journal*, *5*(3), 185–190. https://doi.org/10.15406/ahoaj.2023.05.00206

Teng, M. F. (2022). The effectiveness of multimedia input on vocabulary learning and retention. *Innovation in Language Learning and Teaching*. https://doi.org/10.1080/17501229.2022.2131791

Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

Vygotsky, L. S. (1978). *Mind in Society*. Harvard University Press.

Yunus, M. M., Salehi, H., & John, D. S. A. (2016). Using visual aids as a motivational tool to enhance students' interest in reading literary texts. *Advances in Language and Literary Studies, 7*(1), 1–7.

Zitha, I. (2024). Orthographic challenges encountered by fourth-grade English language learners in a rural-based South African primary school. *Forum for Linguistic Studies*, *6*(6), 84–93. https://doi.org/10.30564/fls.v6i6.6972