

<https://doi.org/10.5281/zenodo.21332790>

Academic vs. Contextual Vocabulary Gains: YouTube Educational Videos as CALL Input for Pakistani ESL Learners



Lubna Khalil	PhD Scholar, Riphah Institute of Language & Literature, Riphah International University, Lahore, Pakistan lubna.khalil1@outlook.com
Dr. Muhammad Islam	Assistant Professor, Department of English Language Teaching & Linguistics, IER, University of the Punjab, Lahore, Pakistan mislam.ier@pu.edu.pk
Fatima Noor	Graduate, Department of English Language Teaching & Linguistics, University of the Punjab, Pakistan fatima.noor.el@gmail.com

Abstract: *This experimental study investigates how YouTube educational videos enhance vocabulary learning among Pakistan's English as a Second Language (ESL) students. Through a one-group pre-test post-test experimental design with 92 undergraduate students at a public university in Lahore, we examined whether Computer-Assisted Language Learning (CALL) materials from educational YouTube channels improve academic and contextual vocabulary knowledge of ESL students. The experimental group received instructions over an eight-week intervention period through carefully selected YouTube videos with subtitles from reputable channels like TED-Ed and BBC Language Learning, followed by vocabulary exercises. Pre-tests and post-tests measured vocabulary gains in academic vocabulary learning and contextual vocabulary understanding separately. Results showed that participant students demonstrated significantly better performance in vocabulary acquisition, with notably higher learning gains in the post-test. The gain in academic vocabulary lexicon was much higher than the contextual vocabulary, indicating YouTube's potential to enhance academic vocabulary knowledge to a greater extent. This research contributes valuable insights to CALL by demonstrating how readily available online resources can be effectively integrated into ESL classrooms to enhance vocabulary learning outcomes, offering practical implications for educators seeking innovative teaching methods in language instruction.*

Keywords: YouTube Educational Videos, ESL Vocabulary Acquisition, CALL, Academic Vocabulary, Contextual Vocabulary

Introduction

Vocabulary knowledge occupies a central position in second language learning (SLA), functioning as the structural foundation on which reading comprehension and broader linguistic competence are built (Nation, 2001a; Schmitt, 2008). Recent empirical work continues to confirm this relationship, demonstrating that vocabulary breadth and depth significantly predict second language (L2)

reading comprehension across learner populations (Allal-Sumoto et al., 2023; Xie & Yeung, 2025). Despite this consensus, learners in English as a Foreign and Second language (EFL/ESL) contexts continue to find active and effective vocabulary acquisition a persistent challenge (Schmitt, 2008; Rasheed et al., 2025). This difficulty signals a need for instructional approaches that move vocabulary from the periphery of the curriculum to its core. One such

approach lies in the affordances of digital technology, which has increasingly reshaped how vocabulary is presented, encountered, and retained in the language classroom (Mohsen et al., 2024; Wei & Fan, 2022).

Since the advent of the twenty-first century, technology integration has become an integral pedagogy in education and language learning. The need is fulfilled by Computer-Assisted Language Learning (CALL), which has emerged as a dynamic tool for second language acquisition (Beatty, 2010). CALL plays a significant role in improving ESL vocabulary and communicative learning through multimedia input (Levy, 1997; Khalil & Islam, 2026). Among the diverse resources that CALL encompasses, online video-sharing platforms have become particularly influential, offering language learners abundant, freely accessible audiovisual input that can be integrated into instruction (Wei & Fan, 2022). YouTube is one of the most popular and accessible platforms, used by millions of people worldwide (Yassin, 2024). Bonk (2008) argues that it is a cultural tool and should be used in educational institutions as a supplement to traditional curricula. Ktoridou et al. (2002) support this idea by highlighting YouTube's potential to deliver language learners with authentic content and context. L2 learners can enhance their language skills to a greater extent when exposed to this audio-visual input. Moreover, interactive YouTube features, such as pause, replay, language-caption options, audio choices, speed, and many more, make it a flexible and user-friendly platform.

The current study is theoretically grounded in Mayer's Cognitive Theory of Multimedia Learning (CTML) as supported by Paivio's (1971) Dual-Coding Theory. The cognitive theory of multimedia learning is based on three cognitive principles of learning: humans process audio and visual information through two different channels, these channels have limited capacity to process information at a time, and human cognition is actively involved in processing information through dual channels (Mayer, 2005). CTML opts for a sensory-modality approach to distinguish between

visually presented materials (images, illustrations, videos) and auditorially presented materials (speech, background sounds, narrations), and a presentation-mode approach to mark the distinction between the construction of pictorial and auditory modalities in human cognition.

The rationale for this theory is based on the dual-coding theory of Paivio (1971). According to this, people learn better when information is presented via both audio (speech) and visual (pictures) modes. This is defined as the "multimedia principle" by Mayer (2005). Sadoski (2005) explains that DCT principles can be really helpful in meaningful vocabulary teaching and learning. Relatively recent scholarship continues to support this theory, using functional brain evidence to demonstrate that mental representations are modality-specific and multimodal (Paivio, 2014).

This study explores the potential of CALL materials available on YouTube educational videos as a pedagogical approach in ESL classrooms. The study also distinguishes between the role of YouTube Educational input in enhancing academic vocabulary knowledge and contextual vocabulary understanding among Pakistani ESL learners.

Research Questions

The study aims to address the following two research questions:

1. To what extent do YouTube educational videos affect ESL students' academic vocabulary learning?
2. To what extent do YouTube educational videos affect the ESL students' contextual vocabulary learning?

Literature Review

In post-colonial countries, like Pakistan, acquiring an L2 is an inevitable cultural phenomenon. However, the SLA is accompanied by several social, cognitive, and linguistic challenges. Mansourzadeh (2014) argues that it requires several competencies and constituents. One of these important constituents is vocabulary learning. Nation (2001) and Webb

(2019) state that vocabulary is the core of language. Harmer (2001) illustrates its significance with the analogy of grammar as the skeleton and vocabulary as the flesh of language. According to Sadoski (2005), vocabulary learning is more important to literacy than it is acknowledged by teachers and researchers. Various psychometric studies have confirmed that rich vocabulary is a central factor in language learners' reading, decoding, and comprehension skills (Chen & Zhang, 2023; Xie & Yeung, 2025).

Academic Vocabulary

Academic vocabulary is defined as words that are used in a large number of academic texts from many different fields, but are not among the most common words in general-use English (Coxhead, 2000). Words like *analyze*, *formulate*, *allocate*, *implement* do not have to be used in a specific subject area, but will be used in lectures, textbooks, and in scholarly writing across subject areas. Nation (2001) places this type somewhere between the general high-frequency words and the technical, more sophisticated discipline-specific words, and reports that it is richly functional in the context of academic communication. As a third-level ESL learner, this vocabulary is important to students' academic achievement because of their need to read academic texts, follow lectures and write in a corresponding register (Webb & Nation, 2017). While a learner may have a good command of conversational English, lack of the development of academic vocabulary is a concern and thus it is a priority to explicitly teach it in English for Academic Purposes (EAP) (Flowerdew & Peacock, 2001).

Contextual Vocabulary

Contextual vocabulary, on the other hand, is based on the learner's ability to deduce and integrate the meaning of the word based on the text or situation at hand instead of on separate word definitions. Much of this learning is incidental, in that as the learners repeatedly see the words in meaningful input, they gradually acquire the sense and use of the words (Nation, 2001). Audiovisual material is especially appropriate for this process as it provides words

in authentic discourse, visual scenes, and intonation, which serve as effective contextual supports for inference (Reynolds et al., 2022). The need for prior contextual meaning processing in accurate word decoding is also highlighted by Sadoski (2005), whose study showed that knowledge of a word requires more than just form recognition. Contextual vocabulary is necessary for ESL learners as it allows them to understand words as they naturally occur in communication, where they are not necessarily explicitly glossed and meaning is reconstructed from context.

Empirical Studies on YouTube and Vocabulary Learning

A growing body of scholarship has investigated the impact of multimedia resources on the language learning of ESL/EFL students. Researchers argue that a blend of audio and visual resources is the best method to teach vocabulary and other language skills to students. Perez et al. (2018) conducted a research study exploring university students' vocabulary learning through audio-visual input. They experimented with different types of L2 subtitles and found that glossed keyword captions were most useful for vocabulary learning and retention.

Similarly, many studies have explored the use of YouTube videos for students' self-learning in addition to their traditional classroom curricula. Elaoufy (2025) explored the effect of YouTube videos on students' vocabulary acquisition outside the classroom. He concluded that YouTube can be used as an effective tool for student-centered vocabulary learning, noting that students were actively engaged with the platform by pausing, taking notes, interacting with comments, and replaying clips. The study proposed that multimedia resources should be mindfully selected and integrated into the curriculum for a more holistic learning of EFL/ESL students. Likewise, Kaboocha and Elyas (2018) integrated YouTube videos in reading classes of Saudi EFL students. Their research revealed that the group who viewed videos performed significantly better than those who did not, highlighting the visible impact of YouTube as CALL material, on the effective

vocabulary learning of students.

Beyond enhancing language skills, YouTube has also been associated with several multifaceted educational benefits. Wu (2024) investigated the impact of YouTube channels on vocabulary acquisition and self-directed learning of students. Results demonstrated that not only was there a significant increase in vocabulary but also in the self-directed learning motivation of students after the intervention. These studies highlight the importance of multimedia resources, especially YouTube, as a multidimensional tool capable of developing multiple language competencies among students.

A similar idea has been propagated by the research of Silverman & Hines (2009), who investigated the impact of multimedia vocabulary learning on English language learners (ELLs) and non-English language learners (non-ELLs). The findings revealed that multimedia resources notably improve students' vocabulary knowledge and understanding. These results align with dual-coding theory and CTML, which focus on the simultaneous integration of audio and visual modes in language learning for a better understanding of students.

In the Pakistani context, there has been little research on the effectiveness of YouTube as a CALL input for students' vocabulary learning. However, quite recently, Nawaz et al. (2026) did a comparative analysis of the impact of YouTube on vocabulary learning in ELLs and non-ELLs. They found that there was a significant increase in the vocabulary knowledge of ELLs after the intervention, as compared to non-ELLs. The study highlighted the importance of learner differences and the need for designing YouTube educational content that accommodates the needs of different types of learners.

YouTube as a CALL input

In the modern world, learners are exposed to screens and multimedia content at a very young age. From early childhood, parents and teachers use various images as word representations to foster their language learning skills. Existing

scholarship has widely investigated the impact of social and multimedia resources on the learning of dual language users (Simonnet et al., 2025). In the same vein, Teng (2023) observed that EFL learners learn better by the use of educational media that incorporate both verbal and nonverbal codes of human cognition. He highlights that this type of content is relatively multidimensional and helps students learn extensive vocabulary. Moreover, the media has become more affordable, demonstrating potential to be used as an effective tool in the educational domain (Mohsen et al., 2024).

Recent local work has shown that Pakistani ESL learners respond positively to CALL-based instruction and perceive it as a valuable support for communicative practice (Khalil & Islam, 2026), yet its application to vocabulary development through YouTube remains largely unexplored. The empirical evidence reviewed above consistently points to the value of audiovisual and multimedia input for vocabulary learning, yet it raises a deeper question of why such input is so effective. This mechanism is best explained by Dual-Coding Theory, which accounts for how verbal and visual information are jointly processed in human cognition.

Dual Coding Theory

The dual coding theory, as proposed by Pavio (1971), served as a supporting framework for this research. This theory suggests that human cognition is processed by two distinct yet interconnected codes: verbal (speech) and non-verbal (visual). The verbal code is used for learning words, while the non-verbal code serves as mental imagery for those words. When words are taught with visual aids, animations, and images, human cognition understands and retains them better.

Sadoski (2005) asserts that words can simply be stated as "verbal labels" for concepts. Concrete ideas, such as "tree", can easily be processed as mental imagery, while many abstract ideas, such as "happiness", need descriptive explanations to create a mental representation. The verbal and nonverbal codes of human cognition work together to understand words and their meanings. Sadoski et al. (1991) argue that the

DCT view of words and concepts can be compared to many other single-coded theories, such as schema theory. Schema theory sees human cognition as having data structures that are responsible for understanding words by the use of generic abstract concepts, already present in memory (Rumelhart, 1980). However, these single-coded theories do not distinguish between abstract and concrete concepts, nor do they answer why concrete concepts are easier to understand than abstract ones (Sadoski, 2005).

Dual-Coding theory posits that words, their meanings, and even sometimes their pronunciation are determined by the range of connections they have with various verbal and nonverbal codes. This understanding of meaning is heavily influenced by the context. Sadoski (2005) has underscored the importance of context in language learning. Although word reading is primarily verbal in nature, understanding its meaning largely depends on nonverbal representations. While Dual-Coding Theory establishes the foundational principle that verbal and nonverbal information are processed through two distinct cognitive systems, it is Mayer's Cognitive Theory of Multimedia Learning that extends this premise into a model of instructional design. Building on the dual-channel assumption, CTML further specifies how learners select, organize, and integrate words and images during multimedia learning.

Cognitive Theory of Multimedia Learning

The cognitive theory of multimedia learning, as presented by Mayer (2005), explains how people learn language by the use of words and pictures, while they have two separate channels to process this verbal and nonverbal information. It is further based on the idea that each channel has limited capacity to process the information, and meaningful learning depends on the active involvement of human cognition. CTML is rooted in the idea that people learn better when they are exposed to both verbal (words) and nonverbal (visual) modes. However, all the instructing methods that use both words and pictures are not equally effective (R. E. Mayer, 2024). The definition of words and pictures, as used by CTML, is vast. Words can be used by

any mode of communication, written (book-based or computer-based), or spoken (speech, narration, direct or indirect). Likewise, pictures can be static (photos, illustrations) or dynamic (videos, animations). This definition of words and pictures is broad enough to include all the communication models used across the globe (Mayer, 2005).

Assumptions of CTML

Camp et al. (2021) argue that the development of the cognitive theory of multimedia learning (CTML) is based on various existing theories in cognitive psychology, such as dual-coding theory by Paivio (1971) and the multi-stage model of memory by Baddeley (1986). The following are the three cognitive assumptions on which CTML is based:

a) Dual Channels: This assumption of CTML is currently most related to Paivio's dual-coding theory. It states that humans process verbal/spoken information through the auditory channel and visual/pictorial information through the visual channel. Both these channels work separately but are interconnected. It is opposed to the designer's conception of human cognition, according to which humans possess a single channel to process all the information, regardless of its modality (Mayer, 2005).

b) Limited Capacity: The second assumption of CTML focuses on the limited capacity of each channel to process the presented information at a time.

c) Active Processing: According to Mayer (2005), humans are active processors of the information presented to them. Human cognition actively gets involved with the information to make a coherent representation of it. Conflicting to this is the idea of humans being passive processors of knowledge. This idea proposes that human cognition records as much information at a time as possible in the memory to retrieve later.

Five Cognitive Processes in CTML

Human minds are active processors of knowledge, transforming information into comprehensible mental representations/modalities. Some of the typical

ways in which knowledge is constructed include comparing, sorting, generalizing, and classifying information (R. E. Mayer, 2024). Five major cognitive processes are used when learners are engaged in active processing of knowledge (verbal and nonverbal).

1. Selecting relevant words: When words (spoken or written) reach the auditory or visual channel, the human mind actively starts selecting only the relevant and meaningful information.
2. Selecting relevant images: When visuals are presented to the eyes, human cognition selects the most relevant images.
3. Organizing selected words: After the selection of words, humans organize those words in their mental lexicon.
4. Organizing selected images: Selected images by visual channel are actively organized in the brain to create a mental representation.
5. Integrating verbal and pictorial modes into each other: Both words and pictures are coherently integrated into each other to develop the full message. The prior memory/schema is used to create the context of the information.

Research on the application of YouTube in language classrooms has shown significant findings that encourage its use with systematic planning. Most of the present literature has explored the effect of YouTube on the overall vocabulary learning of students. However, studies separately investigating the effectiveness of YouTube in academic and contextual vocabulary teaching and learning are very limited. Evidence regarding the impact of YouTube educational videos on higher education students also remains limited in the existing scholarship. Although there has been significant research on the effectiveness of YouTube as a CALL resource for language learning of students, few scholars have examined its impact on Pakistani ESL students. The global research following this research problem has investigated EFL learners, and

limited research is available on the vocabulary learning of ESL students.

Therefore, the current study sought to explain the impact of YouTube as CALL input on higher education ESL students in Pakistan, supported by Paivio's Dual-Coding theory and Mayer's Cognitive theory of Multimedia learning. This research also contributes to the existing literature by distinguishing between academic vocabulary acquisition and contextual vocabulary understanding, thereby providing a more nuanced perspective on multimedia integration in language learning.

Research Methodology

This research employed a one-group quantitative experimental design with a pre-test post- test design.

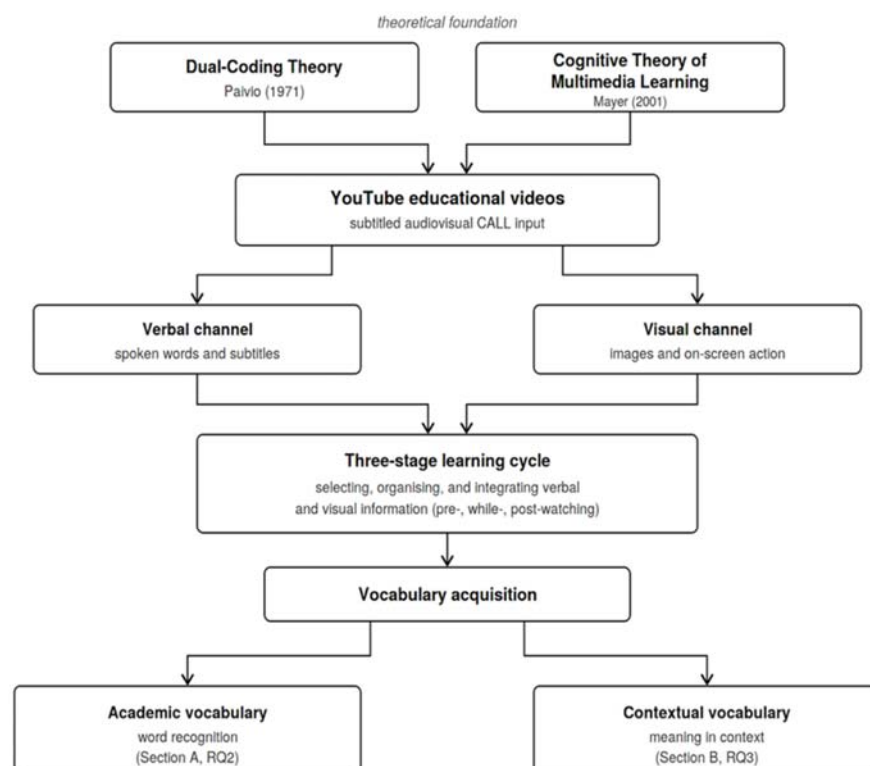


Figure 1

Conceptual Framework of the Study (Developed by the authors based on Paivio (1971) and Mayer (2001))

Building on this theoretical foundation, the conceptual framework of the present study specifies how these cognitive principles apply to the use of YouTube educational videos for vocabulary development. As shown in Figure 1, subtitled YouTube videos serve as audiovisual CALL input that simultaneously activates the verbal channel, through spoken words and subtitles, and the visual channel, through images and on-screen action. This dual input is processed through a three-stage learning cycle in which learners select, organize, and integrate the incoming information across pre-watching, while-watching, and post-watching activities. The framework proposes that this process leads to vocabulary acquisition, examined in the study through two distinct outcomes: academic vocabulary, measured as word recognition (Section A), and contextual vocabulary, measured as understanding of meaning in context (Section B). Overall vocabulary gain is assessed from pre- to post-intervention.

Participants and context of the study

110 ESL undergraduates in the 2nd semester at a public university in Lahore, Pakistan, were involved in the research. The participants' ages ranged from 18 to 21 years. After data cleansing, 92 valid responses were analyzed. The remaining responses were excluded due to missing data and incomplete participation. The UVLT (updated vocabulary levels test) (Webb et al., 2017) measured the students' vocabulary knowledge at an intermediate level of the English language.

Data Collection Tools

Both pre- and post-tests, consisting of 30 items, were divided into two sections. Section A (15 items) was designed to test the academic word recognition level of the students. Section B (15 items) tested the students' ability to understand vocabulary in context. The difficulty level of

both pre- and post-test was maintained approximately the same, but the item order was changed to reduce the practice effect on students' performance. The instrument's content validity was established through evaluation by two subject-matter experts, who reviewed the test items for relevance and representativeness of the target vocabulary.

The pre-test was conducted before the intervention to assess the vocabulary knowledge of participants. The post-test was conducted after the intervention to compare the difference between students' vocabulary knowledge before and after the intervention.

Intervention design

The experimental group watched meticulously selected YouTube videos with subtitles and follow-up vocabulary exercises. The researchers designed the vocabulary exercises according to

the vocabulary level of the participants. The intervention consisted of 16 sessions over the period of 8 weeks, with 2 hours of sessions per week.

The CALL input was collected from various YouTube channels such as TEDx Talks, TED-ED, BBC Language Learning, and Crash Course. A total of 7 YouTube videos were selected based on their vocabulary level and engaging content. Video duration ranged from 12 to 21 minutes. Before playing the videos, it was ensured that they were audible and had clear and readable English subtitles. All the selected videos were educational and academic in nature rather than being merely entertaining. Figure 2 presents an overview of the selected video themes, sources, and duration of the targeted vocabulary in each video.

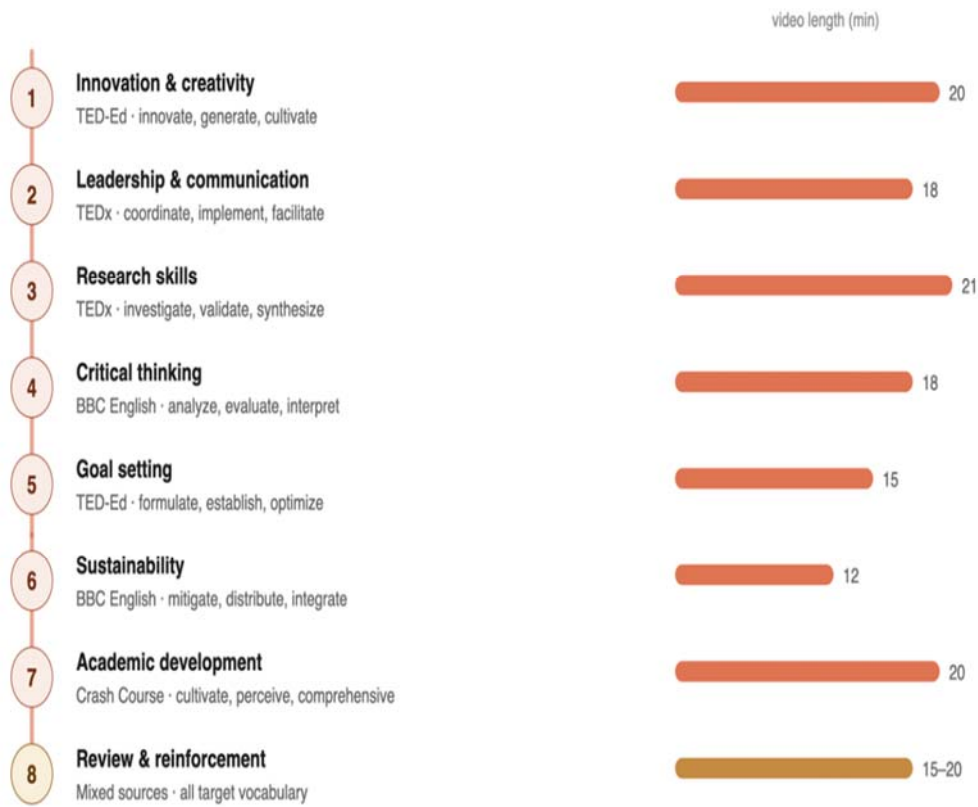


Figure 2

YouTube-CALL intervention

Participants were exposed to approximately 8-12 lexical items during each intervention week.

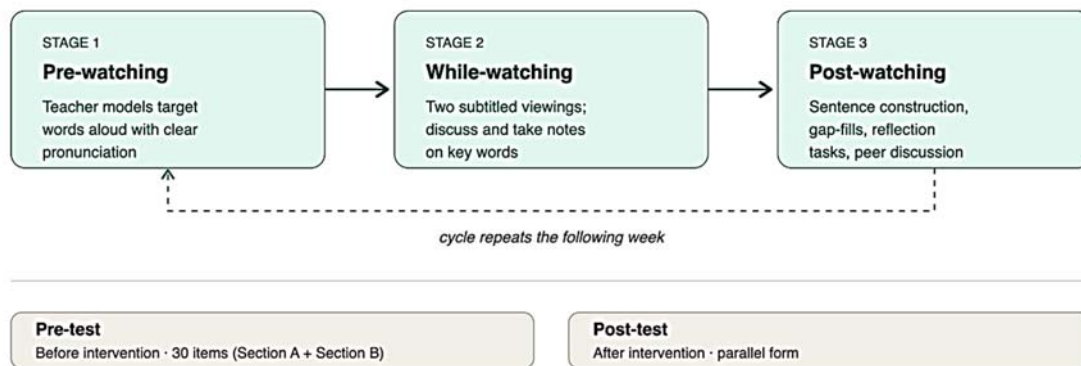


Figure 3

Weekly vocabulary learning cycle- 3 stages

A three-stage vocabulary learning cycle was implemented to ensure that students actively get involved with the words and learn their contextual use (Figure 3). The learning cycle consisted of pre-watching, while-watching, and post-watching phases.

In the pre-watching stage, the instructor introduced the topic to the students by using the targeted vocabulary words with clear pronunciation and intonation. The teacher intentionally emphasized the target words to help students make them part of their subconscious minds while watching the videos. In the while-watching phase, students watched the videos with English subtitles. After the first watching session, the teacher discussed some of the vocabulary words used in the video. In the second watching phase, students took notes of important words and their context. The videos were played at a normal pace, without any unnecessary pauses. The post-watching phase included vocabulary retention activities such as sentence constructions, gap-filling activities, reflection tasks, and peer discussions.

Data Analysis Procedure

IBM SPSS (Statistical Package for the Social Sciences) was used to analyze the research data. After conducting the pilot testing, the reliability of the pre- and post-intervention tests was checked. Reliability analysis was based on

Cronbach's alpha (α) value, which helped researchers assess the internal consistency of the test items. Once the data were fully cleansed and organized, descriptive statistics such as mean value and standard deviation were computed. A paired sample t-test was conducted to compare the mean difference between pre- and post-test results for the participants. Descriptive statistics and paired sample t-tests were also conducted for Section A (academic vocabulary learning) and Section B (contextual vocabulary understanding) separately. This step contributed significantly to the results of the research study, as it helped researchers examine which part of students' vocabulary was most improved, followed by the YouTube videos as a CALL input intervention.

Ethical Considerations

The researchers adhered to the ethical considerations to the best of their ability. All study participants were informed about the research objectives, and informed consent was obtained from the participants. The researchers ensured the anonymity and confidentiality of the data collected from the participants. Participation was voluntary, and participants were allowed to withdraw from the study at any stage.

Results and Discussion

This section presents the findings of the study

alongside their interpretation, organized around two research questions. The overall results revealed substantial gains in students' vocabulary knowledge following the **Table 1**

Descriptive Statistics and Effect Size Analysis

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>Cohen's d</i>
<i>Pre-test</i>	92	13.271	4.530	-
<i>Post-test</i>	92	27.630	5.117	-
<i>Gain</i>	92	14.358	7.034	2.041

Table 1 results indicated a significant improvement in the post-test mean score (27.63) compared with the pre-test mean score (13.27). To determine the magnitude of the intervention's impact, Cohen's d was calculated. Effect size analysis further confirmed the findings with a value of 2.04. This extraordinarily high value indicates that the intervention demonstrated both statistical and

intervention. Total mean scores of the participants in the pre- and post-tests were calculated using IBM SPSS v28.

practical significance. Although the absence of a control group limits causal interpretations, a very large effect size suggests that the intervention likely played a significant role in improving students' vocabulary knowledge. These descriptive statistics of the experimental data confirmed that YouTube can be used as an effective tool for meaningful vocabulary learning in ESL classrooms.

Table 2

Paired-Samples t-test analysis of Pre- and Post-test results

Variable Pair	MD	SD	t	df	p-value
Pre vs Post	-14.358	7.034	-19.578	91	.000

A paired sample t-test was conducted to analyse the difference between ESL students' vocabulary knowledge at pre- and post-intervention stages. Table 2 results indicated an absolute mean difference of 14.3 between the two stages, which means that students' average scores in the post-test were more than double compared to the pre-test. The 2-tailed significance value revealed that the difference between students' pre- and post-test scores was statistically significant,

($t(91) = -19.57, p < 0.00$).

These findings underscore the importance of YouTube educational videos as an effective CALL input for ESL students' vocabulary acquisition. The results are also supported by Dual Coding Theory, which suggests that learning is enhanced when verbal and visual information are processed simultaneously. Also, Mayer's CTML points out that audio and visual information, when carefully selected, organized,

and integrated, can immensely enrich the learning process.

The findings are also consistent with Kabooaha and Elyas (2018), Wu (2024), and Elaoufy (2025), who reported positive effects of YouTube-based instruction on vocabulary development. The convergence of findings across different contexts strengthens the argument that YouTube educational videos constitute a valuable tool for enhancing vocabulary learning in ESL settings.

RQ1: To what extent do YouTube educational videos affect ESL students' academic vocabulary learning?

Academic vocabulary refers to words occurring frequently and across a wide range of academic disciplines, distinct from both everyday high-frequency words and discipline-specific technical terms (words like *analyze*, *formulate*, *implement*, *allocate*) (Coxhead, 2000; Nation, 2001). To assess the gain in students' academic

vocabulary, a paired sample t-test was conducted between pre-test and post-test responses of section A. This section consisted of fifteen items, including vocabulary words such

as accumulation, allocate, formulate, incorporate, and implement, etc. Students also completed follow-up exercises in each session, relevant to these vocabulary words.

Table 3

Paired-Samples t-test analysis of Pre- and Post-test academic vocabulary learning

Variable Pair	MD	SD	t	df	p-value
Sec A (Pre vs post)	-9.195	3.384	-26.059	91	.000

The results in Table 3 revealed that there was a statistically significant gain in students' post-test scores ($M=13.8$, $SD=2.7$) as compared to the pre-test ($M=4.6$, $SD=1.5$). The absolute mean difference between the scores at two stages was 9.19, indicating a 3x gain in post-test scores. The effect size of this section was calculated by Cohen's d, which resulted in a substantially high value of 2.71.

Repeated verbal and visual input resulted in a stronger lexical representation of academic words in students' cognition. This phenomenon

is explained by Samudra (2019) as "Multimedia may be an appropriate platform to provide L2 vocabulary instruction to DLLs, using clear verbal and visual scaffolds to support vocabulary learning and deliver ample and repeated exposure to new words." Many recent studies also support these findings by pointing out that YouTube educational content not only improves students' vocabulary knowledge but also helps them become independent language learners (Sadoski, 2005; Marulis & Neuman, 2013; Wu, 2024).

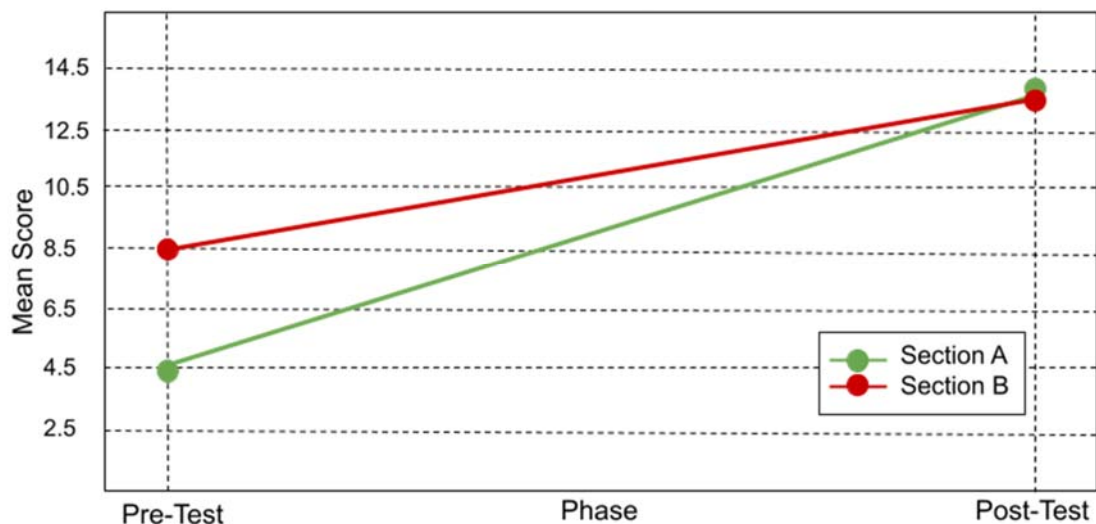


Figure 3: Mean Score Comparison of Section A and Section B at Pre- and Post-Test Stages

Figure 3 shows the mean difference between participants' scores at pre-test and post-test stages for Section A (academic vocabulary learning) and Section B (contextual vocabulary understanding) separately.

RQ2: To what extent do YouTube educational

videos affect the ESL students' contextual vocabulary learning?

YouTube supplies limitless opportunities to maximize learning for language students and "has the potential to anchor instruction in such rich learning contexts" (Bonk, 2008, p. 10). Using YouTube videos considerably improves

students' recognition of authentic language content and context. It enhances their ability to comprehend and produce correct language

Table 4

Paired-Samples t-test analysis of Pre- and Post-test contextual vocabulary understanding

Variable Pair	MD	SD	t	df	p-value
Sec B (Pre vs post)	-5.163	5.248	-9.436	91	.000

Table 4 presents a paired-samples t-test for Section B, which measured vocabulary improvement in context. The results indicate that students gained a mean score of 5.1 after the intervention. The pre-test analysis of section B yielded a mean score of 8.5 ± 4.4 , whereas in the post-test, the students scored considerably higher ($M=13.7$, $SD=2.6$). Effect size analysis resulted in a Cohen's d value of 0.984, which is considered a notably large effect in experimental research.

Research on the application of YouTube in language classrooms has shown significant findings that encourage its use with systematic planning. Studies presented by (Perez et al., 2018) and Kabooha & Elyas (2018) reveal that students who are exposed to both audio and visual modes of information perform significantly better than those who are not. The results are supported by Dual-Coding theory as well, which explains that people determine the meaning of words by a range of connections (mainly verbal and nonverbal). These meanings are influenced by the context in which information is presented. Sadoski (2005) has also highlighted the importance of context in language learning, according to whom accurate decoding of information cannot take place without prior contextual understanding of words.

Overall, the results of the study show a consistent and theoretically coherent pattern. YouTube educational videos, when integrated through a structured three-stage learning cycle, produced substantial vocabulary gains overall, yet the effect was markedly stronger for academic vocabulary than for contextual vocabulary. This divergence suggests that the explicit, repeated, and instructor-mediated

structures, improving their overall language skills (Ktoridou et al., 2002).

exposure to academic target words fostered deeper lexical encoding than the comparatively incidental processing involved in contextual understanding. These findings align with Dual-Coding Theory and Mayer's CTML, both of which hold that learning is enhanced when verbal and visual information are deliberately selected, organized, and integrated. While the absence of a control group warrants caution in attributing these gains solely to the intervention, the convergence of large effect sizes across all three measures strengthens the case for YouTube as a viable CALL resource in ESL vocabulary instruction.

Study Limitations

There are some limitations to be noted. Firstly, the study was limited to higher education ESL learners in Pakistan. The duration of the intervention was quite brief and a delayed post-test to check vocabulary retention was not carried out. The one-group pretest-posttest design makes it difficult to draw any causal conclusions, as other variables such as pre-existing vocabulary knowledge may have influenced the results. Also, a researcher-designed test of specific target words was used to determine learners' vocabulary level, which does not generalize to productive vocabulary use or to other words. These limitations could be overcome in future research by introducing a control group, a longer intervention period and a delayed post-test to assess retention.

Conclusion

The study aimed at investigating the impact of YouTube videos as CALL input on the vocabulary learning of ESL students in Pakistan. Based on the theoretical foundations of Mayer's Cognitive Theory of Multimedia Learning

(2001) and Paivio's Dual-Coding theory (1971), the study explored how audio and visual inputs support ESL students in academic and contextual vocabulary learning. The research design employed a one-group experiment with pre- and post-tests. Results were analyzed using a paired sample t-test, and the effect size was calculated by Cohen's d. Analysis revealed that students scored significantly higher in the post-test than in the pre-test, indicating a substantial gain in vocabulary knowledge following the intervention. Furthermore, results highlighted a statistically significant improvement in the academic vocabulary section and the contextual vocabulary. This indicated that using YouTube as a source of multimedia input was particularly useful for developing students' academic lexical knowledge.

Future studies should employ a control group alongside a longer intervention period and a delayed post-test to establish causal claims regarding vocabulary retention and to overcome the limitations of the one-group pre-test post-test design employed in this study. Additionally, subsequent research could extend this line of inquiry to EFL contexts and diverse proficiency levels across South Asian higher education settings, enabling broader generalizability of the findings related to YouTube-based CALL instruction for both academic and contextual vocabulary development.

References

- Allal-Sumoto, T. K., Miyoshi, K., & Mizuhara, H. (2023). The effect of productive vocabulary knowledge on second language comprehension. *Frontiers in Psychology*, 14, 1049885. <https://doi.org/10.3389/fpsyg.2023.1049885>
- Baddeley, A. (1986). *Working Memory* (Vol. 11). Oxford University Press.
- Beatty, K. (2010). *Teaching and Researching Computer-Assisted Language Learning* (2nd ed). Longman.
- Bonk, C. J. (2011). *YouTube anchors and enders: The use of shared online video content as a macrocontext for learning*. *Asia-Pacific Collaborative Education Journal*, 7(1), 13–24.
- Camp, G., Surma, T., & Kirschner, P. A. (2021). Foundations of Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 17–24). Cambridge University Press. <https://doi.org/10.1017/9781108894333.004>
- Chen, T., & Zhang, D. (2023). Different aspects of vocabulary depth knowledge in L2 Chinese reading comprehension: Comparing higher- and lower-proficiency readers. *Foreign Language Annals*, 56(3), 786–806. <https://doi.org/10.1111/flan.12687>
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213. <https://doi.org/10.2307/3587951>
- Elaoufy, H. (2025). Social media for vocabulary learning: Exploring YouTube's effectiveness in EFL contexts – A study of video-based language learning. *International Journal of Science and Research Archive*, 15(3), 1122–1136. <https://doi.org/10.30574/ijrsra.2025.15.3.1863>
- Flowerdew, J., & Peacock, M. (2001). *Research Perspectives on English for Academic Purposes* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524766>
- Harmer, Jeremy. (2001). *The practice of English language teaching* (3rd ed.). Longman.
- Kabooha, R., & Elyas, T. (2018). The Effects of YouTube in Multimedia Instruction for Vocabulary Learning: Perceptions of EFL Students and Teachers. *English Language Teaching*, 11(2), 72. <https://doi.org/10.5539/elt.v11n2p72>
- Khalil, L., Islam, M., Lodhi, A. S., & Sagheer, I. (2026). Developing AI Literacy Through a Holistic Framework among Pakistani University Students. *Journal*

- of *Political Stability Archive*, 4(3), 724–742.
<https://journalpsa.com.pk/index.php/JPSA/article/view/664>
- Ktoridou, D., Zarpetea, P., & Yiangou, E. (2002). Integrating Technology in EFL: VII Conference of Pedagogiki Eteria of the University of Cyprus. *Proceedings of VII Conference of Pedagogiki Eteria of the University of Cyprus*, A, 437–407.
- Levy, M. (1997). *Computer-Assisted Language Learning: Context and Conceptualization*. Oxford University Press.
- Mansourzadeh, N. (2014). *A Comparative Study of Teaching Vocabulary through Pictures and Audio-visual Aids to Young Iranian EFL Learners*. 24(1), 47–59.
- Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (1st ed., pp. 31–48). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511816819.004>
- Mayer, R. E. (2024). The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review*, 36(1), 8.
<https://doi.org/10.1007/s10648-023-09842-1>
- Mohsen, M. A., Althebi, S., Alsagour, R., Alsalem, A., Almudawi, A., & Alshahrani, A. (2024). Forty-two years of computer-assisted language learning research: A scientometric study of hotspot research and trending issues. *ReCALL*, 36(2), 230–249.
<https://doi.org/10.1017/S0958344023000253>
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language* (1st ed.). Cambridge University Press.
<https://doi.org/10.1017/CBO9781139524759>
- Nawaz, H. F., Khan, A., & Saeed, M. (2026). IMPACT OF YOUTUBE VIDEOS UPON LEARNING VOCABULARY FOR NON- NATIVE SPEAKERS OF ENGLISH: A COMPARATIVE ANALYSIS. *Journal of Social Sciences Development*, 73.
<https://doi.org/10.53664/JSSD/05-01-2026-06-73-86>
- Paivio, A. (1971). *Imagery and Verbal Processes*. Holt, Rinehart and Winston.
- Paivio, A. (2014). *Mind and Its Evolution*. Psychology Press.
<https://doi.org/10.4324/9781315785233>
- Perez, M. M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, 31(1–2), 1–26.
<https://doi.org/10.1080/09588221.2017.1375960>
- Rasheed, M. H., Tugral, A., Ahmed, K., Bibi, R., Ali, S., & Ashraf, R. S. (2025). Enhancing Vocabulary Acquisition in Pakistani ESL Learners Through Video Games: An Experimental Study. *Forum for Linguistic Studies*, 7(7).
<https://doi.org/10.30564/fls.v7i7.9161>
- Reynolds, B. L., Cui, Y., Kao, C.-W., & Thomas, N. (2022). Vocabulary Acquisition through Viewing Captioned and Subtitled Video: A Scoping Review and Meta-Analysis. *Systems*, 10(5), 133.
<https://doi.org/10.3390/systems1005033>
- Rumelhart, D. E. (1980). Schemata: The Building Blocks of Cognition. In *Theoretical Issues in Reading Comprehension*. Routledge.
- Sadoski, M. (2005). A Dual Coding View of Vocabulary Learning. *Reading & Writing Quarterly*, 21(3), 221–238.
<https://doi.org/10.1080/10573560590949359>
- Sadoski, M., Paivio, A., & Goetz, E. T. (1991). Commentary: A Critique of Schema Theory in Reading and a Dual Coding Alternative. *Reading Research*

- Quarterly, 26(4), 463.
<https://doi.org/10.2307/747898>
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363.
<https://doi.org/10.1177/1362168808089921>
- Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of English-language learners and non-English-language learners in pre-kindergarten through second grade. *Journal of Educational Psychology*, 101(2), 305–314.
<https://doi.org/10.1037/a0014217>
- Simonnet, E., Loiseau, M., & Lavoué, É. (2025). A Systematic Literature Review of Technology-Assisted Vocabulary Learning. *Journal of Computer Assisted Learning*, 41(1), e13096.
<https://doi.org/10.1111/jcal.13096>
- Teng, M. F. (2023). The effectiveness of multimedia input on vocabulary learning and retention. *Innovation in Language Learning and Teaching*, 17(3), 738–754.
<https://doi.org/10.1080/17501229.2022.2131791>
- Webb, S. (2019). Helping students become autonomous learners of vocabulary. *TESOL Connections*, 1–4.
- Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, 168(1), 33–69.
<https://doi.org/10.1075/itl.168.1.02web>
- Wei, R., & Fan, L. (2022). On-Screen Texts in Audiovisual Input for L2 Vocabulary Learning: A Review. *Frontiers in Psychology*, 13, 904523.
<https://doi.org/10.3389/fpsyg.2022.904523>
- Wu, C. P. (2024). Enhancing Vocabulary Proficiency and Self-Directed Learning through YouTube: A Study in an EFL Context. *English Language Teaching*, 17(7), 109.
<https://doi.org/10.5539/elt.v17n7p109>
- Xie, Q., & Yeung, S. S.-S. (2025). How is vocabulary involved in second language reading comprehension? A study in Chinese-English bilingual children. *PLOS One*, 20(10), e0334529.
<https://doi.org/10.1371/journal.pone.0334529>
- Yassin, B. (2024). Enhancing English as a Foreign Language (EFL) learning in Saudi Arabia: The academic contribution of YouTube in EFL learning and cultural awareness. *Frontiers in Education*, 9, 1451504.
<https://doi.org/10.3389/feduc.2024.1451504>