

The Role of Textually Enhanced Audiovisual Input and Word Class in Vocabulary Learning, Retention and Content Comprehension of Saudi EFL Female Students

Najla Abdulrahman Alqahtani

Prince Sattam bin Abdulaziz University, Saudi Arabia

Correspondence

Email: na.alqahtani@psau.edu.sa

Abstract

The present study aims to identify the role of textually enhanced captions TEC in vocabulary learning, retention, and content comprehension of Saudi EFL female undergraduates. Using a quasi-experimental design, the study allocated ninety-seven Saudi EFL female students to one of three captioning conditions, i.e., TEC, normal captions (NC), and no captions (NoC). This design was followed by immediate form and meaning recognition vocabulary posttests along with a comprehension test. After an essential gap of three weeks, vocabulary posttests were taken to account for any long-term retention. The target words were categorized into four parts of speech (PoS), i.e., verbs, nouns, adverbs, and adjectives. The purpose here is to gauge the word class that is mostly learned by the participants. The dichotomy of quantitative and qualitative analysis attested a positive impact of TEC in the short-term learning of word forms, whereas the advantage of TEC diminished for the long-term retention of forms. Regarding the meaning recognition and comprehension tests, the outcomes indicated that TEC did not aid in either the recognition of word meanings or the comprehension of the video content. Furthermore, it became crystal clear that verbs emerged as the easiest and most effectively learned word class among the students.

ARTICLE HISTORY

Received: 15 April 2026

Revised: 20 July 2026

Accepted: 02 August 2026

KEYWORDS

Textual Enhancement, Captions, Vocabulary, Word Class, Saudi EFL Students

How to cite this article (APA 7th Edition):

Alqahtani, N. A. (2026). The role of textually enhanced audiovisual input and word class in vocabulary learning, retention and content comprehension of Saudi EFL female students. *Language Teaching Research Quarterly*, 55, 155–181. <https://doi.org/10.32038/ltrq.2026.55.08>

Introduction

Input modality has become increasingly dominant as a supportive means of learning. Even though the verbal mode has been the popular input modality in the educational realm, different input modes including animations, videos, graphs, illustrations, and pictures have been incorporated into educational materials in recent years (Mayer & Moreno, 2002). Thereby, creating a new platform to enhance language learning. The Cognitive Theory of Multimedia Learning (CTML), proposed by Mayer (2005), states that people are more likely to learn when text features accompany illustrations. Therefore, video viewing can be utilized as an alternative source for vocabulary learning (Rodgers, 2018; Sydorenko, 2010). Indeed, research has found that authentic input supported by on-screen text can yield marked improvements in vocabulary learning and comprehension. Subtitles are texts that appear on a screen in the viewer's native language, whereas captions are texts related to the audio signals that have been spoken and written in the target audio language (Vandergrift, 2007). Nevertheless, there is no consensus regarding the optimum on-screen language to use. Another supporting condition is textual enhancement (TE), which has been proposed by some researchers to make learners aware of particular language elements by underlining, coloring, or making text bold (Sharwood, 1993). Whereas numerous attempts have been made in the empirical literature to explore how TE influences the incidental learning of words and comprehension while reading, there is a paucity of research investigating how TECs affect incidental vocabulary learning, retention, and content comprehension.

Literature Review

The Role of TEC in Vocabulary Acquisition

Several studies have explored the role of TEC by means of bolding and underlining target words in facilitating vocabulary acquisition (Sharwood, 1993). While reading studies have largely documented the positive role of TE on the learning of vocabulary (Bishop, 2004; Boers et al., 2017; Choi, 2018; Lee & Jung, 2024; Sonbul & Schmitt, 2013; Toomer, 2013), research on multimodal input remains limited. Montero Perez et al. (2014) were among the first researchers to investigate the effect of TEC in an audiovisual context through a study on Flemish learners of French, revealing that the groups who received captioned viewing showed superior performance to the control group in clip association and form recognition tests, though no significant differences were detected in the test of meaning recall. Later studies by Montero Perez et al. (2018) expanded on this by examining L1 gloss captions, finding enhanced performance in form recognition and recall. Similarly, Cintrón-Valentín et al. (2019) demonstrated that TE benefits both vocabulary and grammatical structure learning, although long-term retention remained inconclusive. Hsieh (2020) further examined TEC in low-intermediate Chinese learners and found that TE with L1 gloss improved vocabulary gains but did not significantly impact listening comprehension. Majuddin et al. (2021) and Puimège et al. (2021) extended TE investigations to multiword expressions (MWEs) by revealing that TE increased visual attention and reading time. However, it did not necessarily enhance

recall, echoing findings that attention alone does not guarantee learning. Wu and Yang (2022) further supported the benefits of TEC, particularly when L1 glosses were available. They helped learners establish form-meaning connections. A recent study by Li et al. (2026) investigated the impact of TE bilingual captions among Chinese students and found that TECs were effective in terms of recognizing forms and recalling meanings, but not in meaning recognition. Despite extensive international research, studies on TEC in Saudi Arabia remain scarce. Alshumrani (2019) and Alharthi (2020) explored the influence of various AVI conditions on the incidental acquisition of words. Subsequently, findings showed that captions supported form learning but not necessarily meaning recall. Similarly, Masrai (2020) and Aldukhayel (2021) explored the effect of captioned videos on comprehension, with findings suggesting that captions alone are insufficient for comprehension, particularly for lower-proficiency learners. Collectively, these studies provide empirical support for TE's role in vocabulary learning, though methodological inconsistencies and mixed findings emphasize the need for further research, particularly in underexplored contexts such as Saudi Arabia.

Word Class/Parts of Speech (PoS)

Evidence on the contribution of word class in vocabulary learning is limited, and the findings are still inconclusive (Peters, 2019). Word class has an impact on the learning of L2 vocabulary, with nouns being more predictive to learning gains than other PoS in both intentional i.e. Ellis and Beaton (1993) and Rodgers (1969) along with incidental learning conditions (Horst & Meara, 1999). For instance, earlier research by Ellis and Beaton (1993) investigated German students' vocabulary learning and reported that nouns proved easier to acquire than verbs irrespective of how they were taught. This was attributed to the concreteness or imageability effect, as nouns have greater imageability than verbs. However, in a study by Barclay (2017), the target items were controlled for concreteness and revealed that word class had no effect on vocabulary gains. This implies that concreteness rather than word class plays a facilitative role in learning vocabulary. Concerning incidental vocabulary acquisition, research on this topic remains limited (Peters, 2019). Moreover, Horst and Meara (1999) documented the use of a comic book to acquire lexicon. It led to a positive impact on nouns in contrast to verbs, adverbs, and adjectives. The key findings provided authentic graphical visualization support for the enhancement of nouns in the comic book. This concept was later justified in the picture dictionary. Notably, the visualization was linked to the word's concreteness. Similarly, Van Zeeland and Schmitt (2013) noted the importance of listening in acquiring the word class on-site. Hence, the research made it evident that learning nouns is feasible compared to verbs and adjectives. Likewise, Alharthi (2020) posited that movie streaming gives a hand in acquiring lexicon through the use of part of speech. This study aligns with the existing body of literature where the noun acquiring percentage is much higher than other parts of speech.

The existing research revolves around a context-dependent inconclusive analysis followed by changes in research design, audiovisual input type, methods utilized, and the captioning conditions. In the case of methodological design, the previous literature lacks validity as pre-testing remains absent. These studies failed to address the prior knowledge of EFL/ESL learners. Furthermore, unnecessary variability through pretest and posttest methods is evident, which further affects the potential result reliability. Other studies relied on immediate posttests, which eventually ignored the concept of long-term retention. Additionally, restricting the analysis to the quantitative method only overlooked the in-depth insights into learners' learning exposure. The undermined focus of TEC from AVI, especially in Saudi EFL learners, portrays another significant gap. Plus, the difficulties faced by Saudi learners, whose orthographic system differs from that of English, are not much discussed.

The Study

For effective learning of English as a Foreign Language (EFL), authentic linguistic input is necessary. However, Saudi EFL learners face significant challenges due to limited exposure to English in their daily lives, restricting their opportunities for comprehension and practice. With minimal classroom contact hours, these learners struggle to develop even basic vocabulary, hindering their overall language acquisition (Al-Seghayer, 2019; Al Shabab, 2012; Al-Shiekh, 2009). These few contact hours are not enough to develop the four core skills of the learners, i.e., listening, speaking, writing, and reading. This significant gap can be bridged by utilizing Audiovisual Input (AVI) through digital platforms like Television, Netflix, Amazon Prime, YouTube, etc. Durrant et al. (2022) narrates that the integration of TEC will provide a multimodal input by means of graphical representation, reading, and listening exposures.

The current study reveals the efficacy of TEC in facilitating vocabulary acquisition, retention, and content comprehension. A large body of research focuses on the role of captioned AVI on lexicon learning, retention, and content comprehension in both ESL and EFL contexts. However, studies assessing the effects of TEC on Saudi EFL learners remain scarce. This study seeks to fill this gap by examining the predictive capability of TEC in vocabulary learning, retention, and content comprehension, focusing on word recognition and PoS acquisition. Subsequently, the findings are expected to offer valuable insights into optimizing captioned AVI, particularly TEC for language learning.

Research Questions

RQ1: Are there any differences between TEC, NC, and NoC in the short-term learning and long-term retention of words (at the form recognition and meaning recognition levels) of Saudi EFL female undergraduates? And what are their perspectives about them?

RQ2: Are there any differences between TEC, NC, and NoC in the content comprehension of Saudi EFL female undergraduates?

RQ3: Which word class/part of speech (verb, noun, adjective, and adverb) was mostly learned by Saudi EFL female undergraduates while viewing an AVI?

Methods

Research Design

This quasi-experimental study employed a pretest, immediate posttest, and delayed posttest design. The purpose is to compare the role of TEC in vocabulary learning, retention, and content comprehension among Saudi EFL female students. To mitigate the limitations of non-randomization, a “semi-randomization” method was applied. The study includes three intact classes from the English department, with at least 30 students per class. They were randomly distributed across the three captioning conditions. Data collection followed an explanatory sequential mixed-methods approach. This approach was initially selected to align with the research questions and objectives. The quantitative data by means of vocabulary and comprehension tests were expected to assess the association between the study variables (Mackey & Gass, 2021); i.e., whether or not textually enhanced AVI and the word class had an effect on the participants’ vocabulary gains. As far as the qualitative data were concerned, they were employed to further collect in-depth information as well as clarify the quantitative findings via using semi-structured interviews.

Population and Sampling

This study employed a purposive sampling method, a non-probability approach which was selected due to its feasibility for the nature of this research. Probability sampling methods can be costly and problematic and therefore not suitable for L2 research and applied linguistics (Dörnyei, 2007). The participants were deliberately recruited according to particular criteria pertinent to the objectives of the research (Dörnyei, 2007). Participants were female students from the English department, aged 20–23, with similar language learning backgrounds. They had approximately nine years of English study, no exposure to English abroad, and had completed foundational courses, ensuring they possessed the necessary skills to comprehend the study’s AVI. Initially, 113 students participated, but due to dropout and absenteeism, the final sample comprised 97 students (Table 1). To qualify, participants had to demonstrate mastery of the 1K word level of the Updated Vocabulary Levels Test (uVLT) (Webb et al., 2017), scoring at least 24 out of 30 (Schmitt et al., 2001). According to Schmitt et al. (2001), a threshold that demonstrates mastery requires a total of 24 or more to be scored out of 30. Participants who fell behind the 1k level were eliminated from the experiment ($n=10$). The intervention was conducted in a classroom equipped with an overhead projector and speakers, with students divided into smaller groups for optimal video viewing. Participation was voluntary, confidential, and incentivized with extra course credit.

Table 1*An Overview of Participants Per Condition*

Groups	Starting Number of Participants	Participants Below the 1k Level	Participants not Completing the Sessions	Final Number of Participants	Participants Attended the Interview
NC	43	5	4	34	2
TEC	35	2	0	33	2
NoC	35	3	2	30	2
Total	113	10	6	97	6

*Materials and Instruments**Video selection*

The selection of the AVI was guided by essential criteria (Montero, 2014; Majuddin et al., 2021; Alshumrani, 2019; Teng, 2022). Given that participants had mastery of the 1K word level but varied at the 2K word level, a seven-minute TED Talk, *Beautiful New Words to Describe Obscure Emotions*, was chosen. To ensure the applicability of the video, lexical analysis was run using Compleat Lexical Tutor (Cobb, 2007). It showed that 91% of the video's words corresponded to the 1K word level. Wondershare Filmora Video Editor (version 13) was then employed to create normal and enhanced captions. Captioning was only subjected to the exact words of the speakers, but other non-verbal messages such as laughing were not captioned.

Target words selection

To make sure that the selected vocabulary was not familiar to the participants, the video script was examined, and 30 words which were beyond the 2K word family level (COCA; Davies) were identified. Five instructors reviewed the list to confirm these words were unlikely to be known or recently taught. Sixteen words (four nouns, verbs, adjectives and adverbs) were selected. A pilot study verified no prior knowledge of these words (Schmitt, 2010). As well, the form recognition test contained four pseudowords (33% of the target words), as recommended by Vidal (2011) and Nation and Webb (2017). These pseudowords were validated from Waring and Takaki's (2003) research.

Pretests

Prior to the experiment, participants reviewed the information sheet, gave written informed consent, and completed a background questionnaire. Then, two vocabulary pretests in a paper and pencil format were administered (written format of form recognition and meaning recognition tests). Nation (2001) explained that form and meaning recognition are two distinct skills. Although form recognition entails noticing the form, which is perceived as the initial step of learning a word (Sydorenko, 2010), meaning recognition is related to the degree to which learners recognize the right meaning of a certain word (Pulido, 2004). The vocabulary pretests were performed two weeks before the experiment and comprised 16 items in the test's two sections. The form recognition part was in yes/no format whereby participants had to cross out any word that they had already read or heard. To control for potential guessing, four non-words

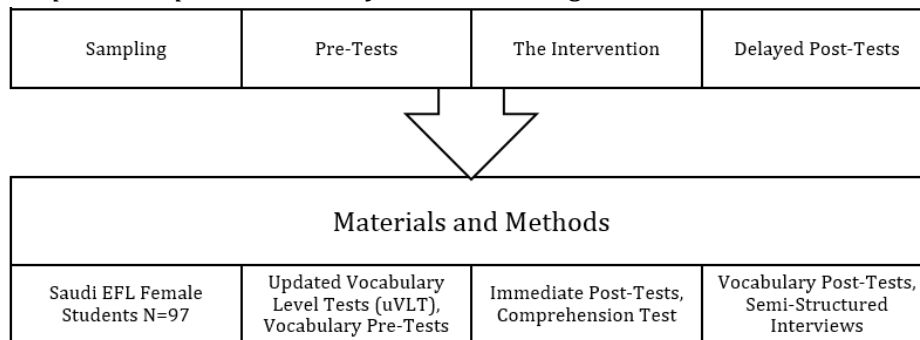
were incorporated into the test, as recommended by Nation and Webb (2011) and Pellicer-Sanchez and Schmitt (2012). Meanwhile, under the meaning recognition test section, the participants were asked to identify the appropriate meaning from four multiple-choice options in English, which included two distractors. In the meaning recognition test, the last option was “I don’t know” as suggested by Rodgers (2013) to control for any guessing.

Posttests

Upon completing the experiment, each participant completed three posttests: a comprehension test consisting of 6 multiple-choice questions assessing participants’ comprehension of the content and two immediate vocabulary posttests, which were the same as those used in the pretests. Three weeks later, the delayed vocabulary posttests were administered. In terms of qualitative data collection, data were gathered through semi-structured interviews due to the flexibility of their structure and procedures. The questions of the semi-structured interview were designed using simple language and were categorized into three main constructs corresponding to each research question: a) participants’ performance on the vocabulary posttests which can ascertain whether the participants’ answers were guessed or learned (Toomer & Elgort, 2013), b) their experiences with captions in promoting vocabulary learning and comprehension and whether captions were a supporting or a hindering tool, c) the word class’s role in vocabulary learning and retention. Following Dörnyei’s (2007) suggestion that interviews must recruit between six to nine participants, this interview was conducted among six participants (two students from each of the experimental groups) on a voluntary basis using simple random sampling method (Figure 1).

Figure 1

Graphical Representation of Research Design



Validity and Reliability

Content validity procedures were used by seeking the feedback of a panel of experts on the tools’ appropriateness. Lawshe (1975) stated that “content validity is used to ensure that each item represented an accurate measure” (p. 75). Accordingly, before approving the experiment’s tools, a random peer check by five experts was used to assess the selected video’s suitability for the students in terms of its content, length, and clarity of the captions. Besides, the experts were also consulted on the list of potential words to

confirm that these words are unknown to the participants and have not been used in the classroom or appeared in their textbooks. As for the comprehension test, the experts have revised the test to assess the correlation between the content of the video and the tested items in addition to validating the structure of the questions and difficulty level. Moreover, the accuracy of the multiple-choices used in the vocabulary meaning recognition test was ensured by validating that all multiple-choices belonged to the same part of speech and were semantically unrelated (Alshumrani, 2019). Finally, the semi-structured interviews' trustworthiness was established by keeping a clear interview protocol, maintaining transparency, and building rapport with participants to foster openness and honesty. A second step to ensure validity was piloting all research tools beforehand so as to cast light on their feasibility for the actual study. Necessary modifications were made prior to their administration.

Scoring and Analyses

For the quantitative analysis, the uVLT, comprehension test, and meaning recognition tests were scored dichotomously, awarding one point per correct response. In the form recognition test, 'Yes' responses to target words signified vocabulary knowledge, whereas 'Yes' responses to non-words indicated guessing. To control for blind guessing, the 'correction for guessing' (cfg) method (Anderson & Freebody, 1983) was applied. The study used SPSS (version 25) for data analysis. A one-way ANOVA was used to compare vocabulary scores, comprehension gains, and word class use across groups. For qualitative analysis, the data from the semi-structured interview were manually analyzed using an inductive thematic approach.

Results

Quantitative Findings

Differences between TEC in contrast to NC and NoC in the vocabulary domain

First, an evaluation was conducted to detect whether the data on vocabulary can be subjected to parametric or non-parametric analysis. Before running the parametric one-way ANOVA test, which is the most achievable test to investigate if the variations observed in the mean scores were statistically significant, several assumptions were tested first. One of these assumptions is the normality of data distribution, which resulted in normally distributed data (Appendix A). Another key assumption underlying parametric ANOVA is homogeneity of variances. To fulfill this assumption, Levene's test, the commonly used test to examine this assumption, was conducted (Field, 2009). It was found that the assumption was met in all the tests except the form recognition immediate posttest (Appendix B). As such, the Welch ANOVA test was employed to identify the differences in the performance of the three groups in the form recognition immediate posttest. For the other tests, which met the assumption of normality and homogeneity of variances, the one-way ANOVA test was run.

Participants' characteristics

The sample size was 97 students, randomly distributed across the three groups: NC ($n = 34$), TEC ($n = 33$), and NoC ($n = 30$). The participants were of a single level (level 5) and were homogeneous in terms of age, gender, native language, nationality, as well as educational level. Given that some learner-related factors, such as the vocabulary size and prior vocabulary knowledge of the target words might influence the performance of the participants on the vocabulary and comprehension tests, it was crucial to decide whether the three experimental groups had any statistically significant differences in these two tests. This was to guarantee that the three groups were comparable in their prior knowledge and vocabulary size, and to ensure that these factors would not play as a confounding variable in the results. Therefore, a between-group comparison was made for the participants' performance on the uVLT and vocabulary pretests.

Participants' vocabulary size

The uVLT was used in a paper and pencil format at the 1k, 2k and 3k word levels to indicate the participants' vocabulary size with 24 out of 30 as a cut-off measure for the level mastery (Schmitt et al., 2001). Based on the descriptive results of the uVLT as detailed in Table 2, all groups showed the highest mastery at the 1K with the percentages falling between 84.84 and 88.57. There were also some variations at the 2k word level, with none reaching the 3k word level. The total mean score for the groups at the 1k was ($M = 25.46$, $SD = 1.860$). Hence, it could be concluded that all participants had mastery of the 1k level, which is necessary to take part in this experiment. This verifies that the participants' vocabulary size was similar across the three groups. Accordingly, this learner-related factor was not further considered in the statistical analyses.

Table 2*Descriptive Statistics of the uVLT*

Word Level	Group	<i>N</i>	<i>M</i>	<i>SD</i>	Percentage
1K	NC	34	25.38	1.79	88.57%
	TEC	33	25.58	1.95	84.84%
	NoC	30	25.43	1.89	86.66%
	Total	97	25.46	1.86	86.69%
2K	NC	34	18.82	4.04	11.43%
	TEC	33	20.18	3.33	15.16%
	NoC	30	19.30	3.40	13.34%
	Total	97	19.43	3.62	13.31%
3K	NC	34	12.44	5.98	0%
	TEC	33	13.45	5.19	0%
	NoC	30	12.40	6.00	0%
	Total	97	12.77	5.69	0%

Participants' prior vocabulary knowledge

To evaluate participants' prior knowledge of the target words, a pretest was conducted two weeks before the experiment. The vocabulary tests' descriptive statistics (Appendix A) showed similar mean scores across the three groups in the form recognition pretest, ranging from 3.37 to 3.97. A similar pattern was found in the meaning recognition pretest where the three groups' mean scores ranged from 3.13 to 3.26. To decide if the mean scores' differences were statistically significant, an ANOVA test was performed. The findings (shown in Table 3) did not reflect significant differences between the three groups in either form recognition ($F(2,94) = 1.101$, $p = .337$) or meaning recognition pretests ($F(2,94) = 0.079$, $p = .924$). Such findings imply that all groups had comparable levels of prior knowledge before starting the actual experiment.

Table 3*One-Way ANOVA for the Vocabulary Pretests*

		Sum of Squares	df	Mean Square	F	Sig.
Form Recognition Pretest	Between Groups	5.776	2	2.888	1.101	.337
	Within Groups	246.554	94	2.623		
	Total	252.330	96			
Meaning Recognition Pretest	Between Groups	.333	2	.167	.079	.924
	Within Groups	198.327	94	2.110		
	Total	198.660	96			

Vocabulary posttests

For the form recognition immediate posttest, there was a substantial increase, with NC reaching a mean score of 9.24 ($SD = 3.115$), TEC at 12.15 ($SD = 1.955$), and NoC at 5.93 ($SD = 2.420$). To establish whether the mean score differences in the form recognition immediate posttest were statistically significant, a Welch ANOVA was conducted. As shown in Table 4, the results indicated statistically significant differences in the three groups' form recognition immediate posttest performance ($F(2, 60.456) = 61.984$, $p < .001$).

Table 4*Welch's ANOVA for the Form Recognition Immediate Posttest*

	Statistic ^a	df1	df2	Sig.
Welch	61.984	2	60.456	.000

a. Asymptotically F distributed.

As shown in Table 5, the post hoc Games-Howell analysis showed several significant differences. The NC group scored much lower than the TEC group ($MD = -2.916$, $p < .001$) and very much higher than the NoC group ($MD = 3.302$, $p < .001$). The TEC group also had a significant score as compared to the NoC group ($MD = 6.218$, $p < .001$).

Table 5*Games-Howell Post Hoc for the Form Recognition Immediate Posttest*

(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.
NC	TEC	-2.916*	.633	.000
	NoC	3.302*	.693	.000
TEC	NC	2.916*	.633	.000
	NoC	6.218*	.558	.000
NoC	NC	-3.302*	.693	.000
	TEC	-6.218*	.558	.000

In terms of long-term retention of word forms, descriptive statistics of the delayed posttest (Appendix A) indicated variance in the performance of the three groups in long-term retention of the forms. The highest mean score was obtained by the TEC group ($M = 7.52$, $SD = 2.68$), and the NC group came right behind ($M = 7.47$, $SD = 2.93$). The NoC group, on the other hand, had the lowest score of all ($M = 4.40$, $SD = 2.22$). As shown in Table 6, the One-way ANOVA indicated statistically significant differences between the three groups. ($F(2,94) = 14.203$, $p < .001$).

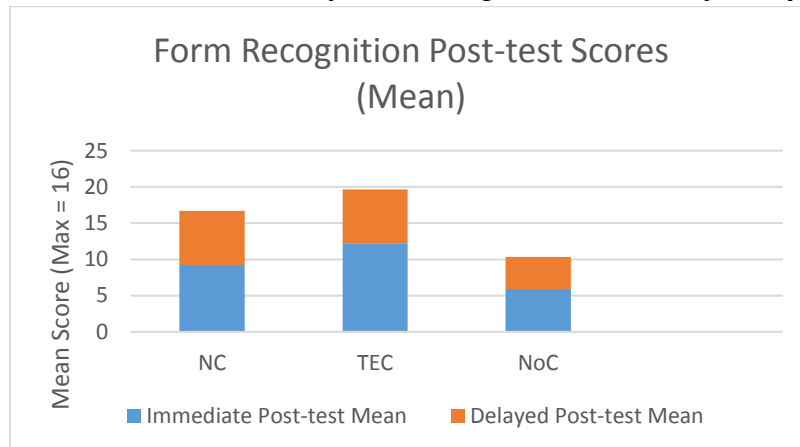
Table 6*One-Way ANOVA for the Form Recognition Delayed Posttest*

		Sum of Squares	Df	Mean Square	F	Sig.
Form Recognition Delayed Posttest	Between Groups	198.211	2	99.105	14.203	.000
	Within Groups	655.913	94	6.978		
	Total	854.124	96			

A post hoc comparison using the Tukey HSD test was carried out to assess the pairwise dissimilarities between groups in the form recognition delayed posttest. As displayed in Table 7, participants in the NC group scored significantly higher than those in the NoC group ($MD = 3.071$, $p < .001$). Similarly, participants in the TEC group scored significantly higher than those in the NoC group ($MD = 3.115$, $p < .001$). However, no significant difference was found between the NC group and the TEC group ($MD = 0.045$, $p = .997$) (Figure 2).

Table 7*Tukey HSD Post Hoc for the Form Recognition Delayed Posttest*

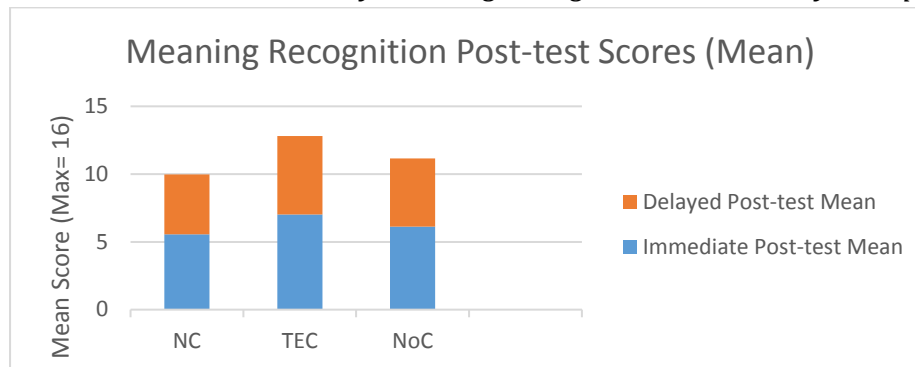
Dependent Variable	(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.
Form Recognition Delayed Posttest	NC	TEC	-.045	.646	.997
		NoC	3.071*	.662	.000
	TEC	NC	.045	.646	.997
		NoC	3.115*	.666	.000
	NoC	NC	-3.071*	.662	.000
		TEC	-3.115*	.666	.000

Figure 2*Mean Scores on Vocabulary Form Recognition Posttests by Group*

In terms of meaning gains, the descriptive statistics of the meaning recognition posttests (Appendix A) indicated differences in performance across the three conditions. Initially, the groups scored similarly on the pretest: NC ($M = 3.26$, $SD = 1.729$), TEC ($M = 3.15$, $SD = 1.253$), and NoC ($M = 3.13$, $SD = 1.306$). However, remarkable improvements were revealed in the immediate posttest, with the TEC group ($M = 7.03$, $SD = 2.651$) outperforming the NC ($M = 5.56$, $SD = 2.218$) and the NoC groups ($M = 6.13$, $SD = 2.543$). Despite some decline in scores during the delayed posttest, the TEC group ($M = 5.79$, $SD = 3.189$) still exhibited higher scores compared to the NC ($M = 4.41$, $SD = 2.017$) and NoC groups ($M = 5.03$, $SD = 2.953$). One-way ANOVA was run to examine the role of captions on meaning recognition in both immediate and delayed posttests (Table 8). In the immediate posttest, the ANOVA test resulted in a slightly significant effectiveness of captions on meaning recognition ($F(2,94) = 3.003$, $p = .054$). Figure 3 suggests that the variations in mean scores between the groups (NC, TEC, NoC) were notable but did not quite reach the conventional level of statistical significance ($p < .05$). Similarly, the delayed posttest did not show a statistically significant effect of captions on meaning recognition ($F(2,94) = 2.096$, $p = .129$).

Table 8*One-Way ANOVA for the Meaning Recognition Posttests*

	Sum of Squares	df	Mean Square	F	Sig.	Sum of Squares
Meaning Recognition Immediate Posttest	Between Groups	36.728	2	18.364	3.003	.054
	Within Groups	574.819	94	6.115		
	Total	611.546	96			
Meaning Recognition Delayed Posttest	Between Groups	31.778	2	15.889	2.096	.129
	Within Groups	712.717	94	7.582		
	Total	744.495	96			

Figure 3*Mean Scores on Vocabulary Meaning Recognition Posttests by Group**Differences between TEC in contrast to NC and NoC in content comprehension*

The descriptive statistics for the content comprehension test (Table 9) provided an overview of the participants' performance in the test across the three groups. The overall mean score ($M = 3.99$, $SD = 1.271$) suggests that the participants had a moderate level of comprehension, which implies that they understood the content to a satisfactory extent. Participants in the TEC group had the highest mean score ($M = 4.09$, $SD = 1.284$), indicating that TEC may aid in better content comprehension. The NoC group followed with a mean score of 4.00 ($SD = 1.259$), showing a comparable level of comprehension to the TEC group, albeit slightly lower. The NC group had the lowest mean score ($M = 3.88$, $SD = 1.297$) of all groups.

Table 9*Descriptive Statistics of the Content Comprehension Test*

	Group	N	Mean	Std. Deviation	Skewness	Kurtosis
Content Comprehension Test	NC	34	3.88	1.297	.054	-.763
	TEC	33	4.09	1.284	-.179	-.213
	NoC	30	4.00	1.259	-.444	-.134
	Total	97	3.99	1.271	-.167	-.503

Before running inferential statistics, an evaluation was conducted to determine the suitability of parametric or non-parametric analyses. The assumption of normality of data distribution as well as the homogeneity of variances (Table 10) were fulfilled, indicating the suitability of conducting the parametric one-way ANOVA test.

Table 10*Test of Homogeneity of Variances for the Content Comprehension Test*

		Levene Statistic	df1	df2	Sig.
Content Comprehension Test	Based on Mean	.030	2	94	.970
	Based on Median	.001	2	94	.999
	Based on Median and with adjusted df	.001	2	93.721	.999
	Based on trimmed mean	.028	2	94	.972

A one-way ANOVA test was run to examine group differences in the content comprehension test. The ANOVA results (Table 11) revealed that the groups did not differ significantly, $F(2,94) = 0.223$, $p = 0.800$. (Figure 4).

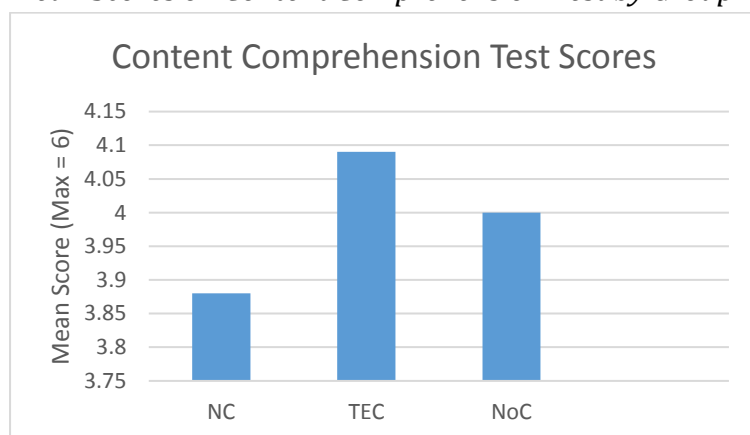
Table 11

One-Way ANOVA for the Content Comprehension Test

		Sum of Squares	df	Mean Square	F	Sig.
Content Comprehension Test	Between Groups	.733	2	.367	.223	.800
	Within Groups	154.257	94	1.641		
	Total	154.990	96			

Figure 4

Mean Scores on Content Comprehension Test by Group



Word class learning

The data in this investigation were obtained using the results of the participants in the form and meaning recognition immediate posttests. Frequencies were computed of all the word classes that were correctly answered during each vocabulary immediate posttest. Notably, the analysis was done only in the immediate posttest as it was considered enough to address the third research question. The word classes' descriptive statistics, as illustrated in Table 12, were summarized as follows. In the form test, verbs had the highest performance among all word classes ($M = 3.00$, $SD = 0.97$). Nouns followed ($M = 2.69$, $SD = 1.00$) then Adjectives ($M = 2.35$, $SD = 0.89$), and finally adverbs which had the lowest scores ($M = 1.32$, $SD = 0.99$). In the meaning test, the scores were generally lower than those of forms indicating better learning of the word forms than meanings. Verbs again had the highest mean score ($M = 2.36$, $SD = 0.83$), followed by nouns ($M = 1.92$, $SD = 0.99$), adjectives ($M = 1.36$, $SD = 0.86$), and finally adverbs with the lowest mean score ($M = 0.65$, $SD = 0.71$).

Table 12*Descriptive Statistics of the Word Classes*

Vocabulary Test	Word Class	N	Mean	Std. Deviation	Skewness	Kurtosis
Form	Verbs	97	3	0.968	-.773-	.008
	Nouns	97	2.69	1.004	-.225-	-.746-
	Adjectives	97	2.35	0.890	.058	-.746-
	Adverbs	97	1.32	0.985	.185	-.977-
	Verbs	97	2.36	.831	.122	-.498-
Meaning	Nouns	97	1.92	.986	.035	-.337-
	Adjectives	97	1.36	.856	-.063-	-.707-
	Adverbs	97	.65	.708	.622	-.790-

Before running the one-way ANOVA test, the key assumptions of the test were checked. The assumption of normality of data distribution as well as the homogeneity of variances (Table 13) were met, which implies the suitability of utilizing the parametric one-way ANOVA test.

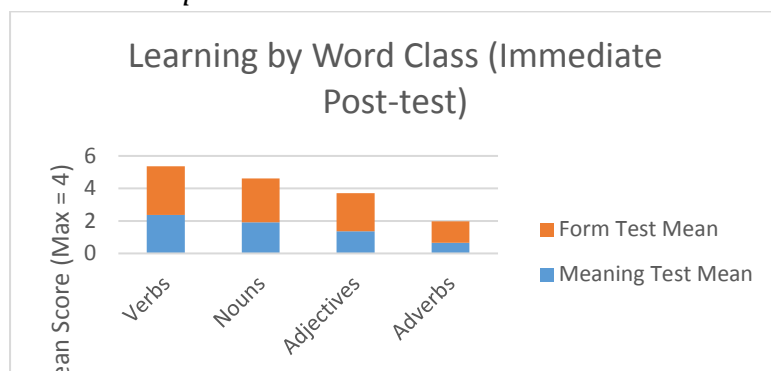
Table 13*Test of Homogeneity of Variances for the Word Classes*

		Levene Statistic	df1	df2	Sig.
Form	Based on Mean	1.491	3	384	.216
	Based on Median	.638	3	384	.591
	Based on Median and with adjusted df	.638	3	383.405	.591
	Based on trimmed mean	1.147	3	384	.330
Meaning	Based on Mean	1.292	3	384	.277
	Based on Median	.658	3	384	.578
	Based on Median and with adjusted df	.658	3	366.487	.578
	Based on trimmed mean	1.296	3	384	.275

The findings of one-way ANOVA as displayed in Table 14 and Figure 5 indicated statistically significant differences between all word classes in the form recognition $F(3, 384) = 55.817, p < .001$, and the meaning recognition tests ($F(3, 384) = 73.080, p < .001$).

Table 14*One-Way ANOVA for the Word Classes*

		Sum of Squares	df	Mean Square	F	Sig.
Form	Between Groups	155.196	3	51.732	55.817	.000
	Within Groups	355.897	384	.927		
	Total	511.093	387			
Meaning	Between Groups	158.814	3	52.938	73.080	.000
	Within Groups	278.165	384	.724		
	Total	436.979	387			

Figure 5*Mean Scores per Word Class in Immediate Posttests*

For the form recognition test, a Tukey HSD post hoc test was conducted to assess the pairwise comparisons between different word classes. As reported in Table 15, all pairwise comparisons revealed statistically significant differences. Looking at verbs compared with the other word classes, it was clear that the mean difference between verbs and nouns was statistically significant ($MD = 0.31, p < .001$). Similarly, significant differences were found between verbs and adjectives ($MD = 0.65, p < .001$), and between verbs and adverbs ($MD = 1.68, p < .001$). Nouns also differed significantly from adjectives ($MD = 0.34, p < .001$), and adverbs ($MD = 1.37, p < .001$). For adjectives, they were significantly different from adverbs ($MD = 1.03, p < .001$).

Table 15*Tukey HSD Post Hoc for the Word Classes in the Form Recognition Test*

Dependent Variable	(I) Word Class	(J) Word Class	Mean Difference (I-J)	Std. Error	Sig.
Form	Verbs	Nouns	.31*	.13824	.015
		Adjectives	.65*	.13824	.000
		Adverbs	1.68*	.13824	.000
	Nouns	Verbs	-.31*	.13824	.015
		Adjectives	.34*	.13824	.008
		Adverbs	1.37*	.13824	.000
	Adjectives	Verbs	-.65*	.13824	.000
		Nouns	-.34*	.13824	.008
		Adverbs	1.03*	.13824	.000
	Adverbs	Verbs	-1.68*	.13824	.000
		Nouns	-1.37*	.13824	.000
		Adjectives	-1.03*	.13824	.000

For the meaning recognition test, findings of the Tukey HSD post hoc pairwise comparisons (Table 16) showed statistically significant differences between all pairwise comparison groups with $p < .05$ across all word classes. Significant differences were found between verbs and nouns ($MD = 0.44, p = .002$), verbs and adjectives ($MD = 1.00, p < .001$), and verbs and adverbs ($MD = 1.71, p < .001$). Nouns also showed significant differences when compared to adjectives ($MD = 0.56, p < .001$) and adverbs ($MD = 1.27,$

$p < .001$). Finally, adjectives were significantly different from adverbs ($MD = 0.71$, $p < .001$).

Table 16

Tukey HSD Post Hoc for the Word Classes in the Meaning Recognition Test

Dependent Variable	(I) Word Class	(J) Word Class	Mean Difference (I-J)	Std. Error	Sig.
Meaning	Verbs	Nouns	.44*	.12221	.002
		Adjectives	1.00*	.12221	.000
		Adverbs	1.71*	.12221	.000
	Nouns	Verbs	-.44*	.12221	.002
		Adjectives	.56*	.12221	.000
		Adverbs	1.27*	.12221	.000
	Adjectives	Verbs	-1.00*	.12221	.000
		Nouns	-.56*	.12221	.000
		Adverbs	.71*	.12221	.000
	Adverbs	Verbs	-1.71*	.12221	.000
		Nouns	-1.27*	.12221	.000
		Adjectives	-.71*	.12221	.000

Qualitative Findings

Initially, interviewees were asked about their performance in the vocabulary posttests to ensure their answers were gained and not guessed. At the beginning of the interview, they were shown a piece of paper including the 16 target words with 4 highlighted words (one word from each word class). Interviewees were asked whether they remembered these words and to illustrate their meaning in either Arabic or English. As shown in Table 17, all six interviewees showed excellent recognition and understanding of these words. However, only one participant from the NC group failed to provide meanings for the three words and was only able to provide meaning for the first word “*published*”. Hence, it could be inferred that captions, particularly TEC, had positively impacted vocabulary learning.

Table 17

Qualitative Findings of the Interviewees' Performance on the Vocabulary Posttest

Interviewee's Referent Number and Group	St 1 (NC)	St 2 (NC)	St 3 (TEC)	St 4 (TEC)	St 5 (NoC)	St 6 (NoC)
Interviewees' Answers	yes. Published means ينشر Sorrows means معاناة Hypothetical means نظري Essentially means truly	yes, the word published means ينشر. Sorrows um I don't know. Hypothetical I don't know it . essential physical, maybe!	um, Published means ينشر Sorrows means like deep sadness Hypothetical means نظريه Essentially means basic	Of course! Published means ينشر sorrows احزان Hypothetical افتراضي Essentially اساسي	published ينشر sorrows means ابتلاء or something like that, essentially like basic. Hypothetical I don't know!	Yes, published is ينشر sorrows احزان hypothetical تجريبي essentially, important.
Score	3/4	1/4	4/4	4/4	3/4	4/4

Positive attitudes and familiarity towards captions were shown by most interviewees (Table 18). However, some interviewees said that they were familiar with both captions and subtitles. Still, they utilized captions for learning (St 3 and St 6), whereas subtitles were used for entertainment purposes or difficult content only (St 3 and St 4). In contrast, only one student (St 5) expressed no preference for either captions or subtitles as she relied on listening to the audio only to test her language.

Table 18

Qualitative Findings of the Interviewees' On-screen Language Preference

Referent	Codes	Themes	Category
St 1	- Preference for captions		
St 3, St 6	- Captions for learning purposes		
St 2, St 6	- Shift in preference from subtitles to captions	- Captions' preference	On-screen language preference
St 4	- Captions are a must in video viewing		
St 1	- Familiarity with caption		
St 2, St 6	- Old preference for subtitles		
St 3	- Subtitles for entertainment purposes	-Subtitles' preference	
St 3	- Subtitles preference for language learning		
St 4	- Subtitles for difficult contents only		
St 5	- No text preference	-No text preference	

For the interviewees' experiences with captioned/subtitled AVI (Table 19), they were asked several questions addressing this topic. In terms of captioning effects, most of the interviewees claimed that captions facilitated both vocabulary learning (St 1, St 2, St 5, St 6) and content comprehension (St 1, St 2, St 5). Further, they revealed that captions improved their concentration (St 1) and that they can compensate for native speakers' fast speech (St 1, St 2, St 5). The perceived advantages of captions were also confirmed for the TEC. Nevertheless, captions are not without disadvantages. Two interviewees confirmed the distracting aspect of the captions (St 3, St 4) and justified that captions cause more focus on the written text rather than the content and audio, which eventually led to distraction. Additionally, the non-verbal cues presented in the videos, such as transcription to sounds of a door, wind, etc can cause distraction. Regarding Arabic subtitles, different views were elicited from the interviewees. Whereas one interviewee stated that subtitles increase learning and attention (St 3), subtitles were perceived as a hindering tool to vocabulary learning and retention (St 4 and St 5), because of the absence of word forms.

Table 19*Qualitative Findings of the Interviewees' Experiences on captioned/subtitled videos*

Referent	Codes	Themes	Category
St 1, St 4, St 6	- Simultaneous watching and reading		
St 1			
St 1	- Concentration aid		
St 1, St 2, St 5, St 6	- Understanding enhancement		
St 1, St 6	- Vocabulary learning facilitation		
St 1, St 2	- Lack of distraction with captions		
St 1, St 2, St 5	- Content comprehension		
St 2, St 4	- Compensation for fast speech	- Pros and cons of captions	Experience on captioned/s subtitled videos
St 3	- Contribute to spelling learning		
St 4	- TE attention attraction		
St 4, St 5, St 6	- TE Single viewing is not enough		
	- Captions' role in displaying words' forms		
St 3, St 4	- Distraction with captions		
St 3, St 4	- Captions lead to more reading than video watching		
St 3	- Non-verbal captioning cause distraction		
St 3	- Subtitles increase learning		
St 3	- Subtitles increase attention		
		- Pros and cons of subtitles	
St 4, St 5	- Subtitles hinder learning		
St 4	- Subtitles' absence of word form display		
	- Subtitles hinder vocabulary retrieving		
St 4			

Regarding the word class learning, interviewees illustrated different opinions towards each word class, provided justification, and explained the challenges they encountered in learning these word classes. As shown in Table 20, it was clear from the interviewees' responses that nouns were regarded as the easiest word class, followed by verbs, adjectives, and finally adverbs, which ranked as the most difficult word class. The explanations provided by the interviewees were insightful in providing more understanding of how Saudi EFL female undergraduates perceive word classes and the challenges they encountered in learning them. The focus on nouns and verbs in textbooks, frequency of usage in daily life, and their clarity made them easier to learn and interpret. In contrast, the confusing aspect of adjectives and adverbs and their infrequent usage made them much harder for the students to use and comprehend.

Table 20*Qualitative Findings of the Interviewees' Learning to the Word Classes*

Referent	Codes	Themes	Category
St 1, St 3, St 5, St 6 St 2, St 5 St 3, St 4, St 6 St 1 St 1, St 5 St 2, St 4, St 6	- Verbs and nouns are the easiest - Verbs are easy - Nouns are the easiest - Adjectives are the most difficult - Adverbs are easy - Adverbs are the hardest	-The easiest word class	Word classes learning
St 1 St 1 St 5	- Nouns and verbs' frequency of encounter - Educational focus on nouns and verbs - Clarity of nouns and verbs	-The most difficult word class	
St 1, St 3, St 6 St 1 St 4 St 4	- Nouns' association with objects - Nouns' Imaginability - Nouns' frequency of usage - Nouns' significance in meaning delivery	- Reasons for preference	
St 2, St 6 St 2 St 5	-Verbs' frequency of usage -Verbs' functional significance -Verbs ease of interpretation	- Challenges with word classes	
St 1 St 1 St 1	- Insufficient adjectives in mind - Adjectives' comprehension difficulty - Adjectives' usage difficulty		

Discussion

The findings of the short-term learning of word forms demonstrated the influence of TE on incidental form learning. In addition, it was revealed that TE improved the participants' noticing of word forms, leading to better results in the immediate posttest. Such results were inconsistent with most of the studies found in the previous literature (Hsieh, 2020; Montero Perez et al., 2014; Wu & Yang, 2022). This could be attributed to the restricted cognitive capacity to simultaneously process and store ongoing transient information while viewing videos where the multimodality of input may have overly burdened EFL learners who have poor linguistic proficiency, thereby hindering their active processing (Wu & Yang, 2022). This aligns with the CLT (Sweller, 1994) which claims that working memory is constrained with limited capacity. Regarding the retention of word forms, all experimental groups detected a decline in their delayed posttest with no statistical significance found between the NC and the TEC groups. The findings of Majuddin et al. (2021) and Cintrón-Valentín and García-Amaya (2021) aligned with the current study's findings where they found that captioning in general, and not TE, aids in immediate vocabulary learning, but its benefits for long-term retention may be limited. It is worth mentioning that the literature acknowledged that forgetting is a normal part of the learning process (Schmitt, 2010; Waring & Takaki, 2003). As reported by previous studies (Hulstijn, 2003; Waring & Takaki, 2003), delayed posttests would normally have a decline in scores, and students need to encounter the target words repeatedly in various settings.

Different findings regarding the learning of words' meanings were reached. Although the TE group scored slightly higher than the other groups in the meaning recognition tests, the three groups did not differ significantly. This result was in agreement with the findings of Hsieh's (2020) and Montero Perez et al. (2014) but in contrast with Wu and Yang (2022). One plausible explanation is that participants were focusing their attention on the novel words shown in the caption without noticing the TE itself (Montero Perez et al., 2014), and that captions can introduce additional information that may overwhelm learners. This overload was explained by the CLT (Sweller, 1994), which states that working memory has a finite processing capacity. Crucially, the quantitative findings of this question were further supported by the qualitative results. Although interviewees from the three groups were able to successfully retrieve the meanings of most of the target words, interviewees from the TEC group slightly outscored the other groups in recognizing words' forms. Further, participants' attitudes were positive regarding the use of captions when they watch a video. Specifically, TE was also reported positively by two interviewees from the TEC group who said that it increased their attention but stated that a single viewing is not enough. This supports the CTML (Mayer, 2005). Conversely, other interviewees thought that captions can distract them due to the multimodality nature of the AVI which provides support to the CLT (Sweller, 1994).

Regarding the second research question, the three groups' outcomes did not differ significantly, indicating a similar effect of the two caption conditions (TEC and NC) and the NoC condition on video comprehension, which corroborates the previous research findings (Montero Perez et al., 2014; Hsieh, 2020) but contradicts other studies (Majuddin et al., 2021). Such findings could be attributed to the cognitive load where multiple information sources were displayed in the video, distracting learners from fully comprehending its content (Bird & Williams, 2002). Interestingly, positive attitudes were shown by most of the participants regarding the impact of captions in facilitating comprehension of the video's content. Nonetheless, captions were reported to be disadvantageous by some interviewees as explained by the CLT (Sweller, 1994) which could account for the non-significant differences found in the three groups' content comprehension test.

The third research question indicated that verbs are the most easily and effectively learned word class by Saudi EFL female undergraduates, followed by nouns and adjectives, with adverbs being the least learned. This contrasts with previous research (Alharthi, 2020; Ellis & Beaton, 1993; Horst & Meara, 1999) that highlighted the advantage of nouns over the other PoS. Furthermore, interviews revealed a preference for nouns over other PoS due to their concreteness and frequent usage, despite the quantitative data showing a preference for verbs. This suggests that while verbs may be easier to recall in tests, nouns are foundational in practical language use. Hence, it could be concluded that both nouns and verbs are the most favored PoS by the Saudi EFL female

undergraduates, followed by adjectives and finally adverbs as the least favored word class.

Conclusion

The present study puts forward valuable implications that EFL teachers and learners should be aware of. First, teachers should raise the awareness of their students towards the usefulness of authentic videos viewing to increase their vocabulary size. Additionally, they should incorporate authentic videos into their lessons and become proficient in using readily available software to add enhanced captions that can optimize their students' learning experience. Incidental and intentional learning of vocabulary must be perceived as complementary, with both being crucial to the gradual process of vocabulary acquisition (Nation, 2001). Since TE alone did not significantly improve word-meaning acquisition or content comprehension, explicit form-meaning activities (e.g., Szudarski, 2012) and translation exercises (e.g., Laufer & Girsai, 2008) are recommended. Besides, teachers are recommended to adopt the "staged video approach" (Danan, 1992) in which the text amount is gradually decreased while viewing AVI. In the Saudi EFL context, incorporating TE videos into university curricula may help address learners' limited vocabulary knowledge. Also, instructions should emphasize verbs and nouns while providing targeted support for adjectives and adverbs.

The study has some limitations that need to be acknowledged. First, purposive sampling was used with a relatively small sample of female Saudi EFL learners from a single university, thereby restricting its generalizability. Research involving larger and more diverse samples from different educational contexts and linguistic backgrounds is recommended. Second, the study examined a limited set of target words and assessed only receptive vocabulary knowledge using written measures. Future research should investigate broader aspects of vocabulary learning, including productive knowledge and spoken word acquisition, using multiple assessment methods. Finally, the short duration of the instructional videos may not reflect long-term learning. Longitudinal studies using longer video materials and process-oriented methods, such as think-aloud protocols or eye-tracking techniques, could provide deeper insights into how textually enhanced captions support vocabulary learning over time.

ORCID

 <https://orcid.org/0000-0002-8115-8250>

Publisher's Note

The claims, arguments, and counter-arguments made in this article are exclusively those of the contributing authors. Hence, they do not necessarily represent the viewpoints of the authors' affiliated institutions, or EUROKD as the publisher, the editors and the reviewers of the article.

Acknowledgements

The author gratefully acknowledges Prince Sattam bin Abdulaziz University for funding this research through project number PSAU/2025/02/34564. Sincere appreciation is also extended to all participants for their valuable time, cooperation, and contribution to this study.

Funding

This research was funded by Prince Sattam bin Abdulaziz University through project number PSAU/2025/02/34564.

CRedit Authorship Contribution Statement

Najla Abdulrahman Alqahtani: Conceptualization, Methodology, Data Curation, Data Analysis, Reviewing, Editing, Supervising and Project Administration

Generative AI Use Disclosure Statement

Throughout the development of this research, the author utilized ChatGPT (OpenAI) and Grammarly for proofreading and language editing. The author revised, edited, and is fully accountable for the content of this research and confirms that it complies with the journal's ethical guidelines and standards of academic integrity.

Ethics Declarations

World Medical Association (WMA) Declaration of Helsinki–Ethical Principles for Medical Research Involving Human Participants

The research was conducted in accordance with established ethical principles. Participants were informed about the study's purpose and their role in it. Informed consent was secured from each participant before data collection. Participants were also informed that all personal information provided would be handled confidentially, and that their participation was entirely voluntary.

Competing Interests

The author declares that there are no financial or personal relationships that could have influenced the work presented in this article.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Aldukhayel, D. (2021). The effects of captions on l2 learners' comprehension of vlogs. *Language Learning & Technology*, 25(2), 178-191. <https://doi.org/10.125/73439>
- Alharthi, T. (2020). Can adults learn vocabulary through watching subtitled movies? An experimental corpus-based approach. *International Journal of English Language and Literature Studies*, 9(3), 219-230. <https://doi.org/10.18488/journal.23.2020.93.219.230>
- Al-Seghayer, K. (2019). Unique challenges Saudi EFL learners face. *Studies in English Language Teaching*, 7(4), 509-510. <https://doi.org/10.22158/selt.v7n4p490>

- Al Shabab, O. S. (2012, April). First person domain: Threshold mental lexicon and Arab learners of English. In *Proceedings of the Second Symposium on English Language Teaching in KSA: Realities and Challenges* (pp. 141-208). Riyadh.
- Al-Shiekh Mubarak, G. (2009). *Vocabulary range and frequency in first intermediate girls' English textbook in Saudi Arabia: Learners' achievement and strategies*. [Unpublished master's thesis]. King Faisal University.
- Alshumrani, H. (2019). *L2 incidental vocabulary learning and retention through different modalities of audio-visual input*. [Unpublished doctoral dissertation]. University of Southampton.
- Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Hutson (Ed.), *Advances in reading/language research: A research annual* (pp. 231-265). JAI Press.
- Barclay, (2017, March). *The effect of word class and word length on the decay of lexical knowledge* [Conference presentation]. AAAL 2017, Portland, OR, United States.
- Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 227-248). John Benjamins.
- Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, 27(2), 448-469. <https://doi.org/10.1111/ijal.12141>
- Choi, S. (2018). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, 21(3), 403-426. <https://doi.org/10.1177/1362168816653271>
- Cintrón-Valentín, M., García-Amaya, L. & Ellis, N. C. (2019). Captioning and grammar learning in the L2 Spanish classroom. *The Language Learning Journal*, 47(4), 439-459. <https://doi.org/10.1080/09571736.2019.1615978>
- Cintrón-Valentín, M. C., & García-Amaya, L. (2021). Investigating textual enhancement and captions in L2 grammar and vocabulary: An experimental study. *Studies in Second Language Acquisition*, 43(5), 1068-1093. <https://doi.org/10.1017/S0272263120000492>
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, 11(3), 38-63. <https://doi.org/10.64152/10125/44117>
- Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning*, 42(4), 497-527. <https://doi.org/10.1111/j.1467-1770.1992.tb01042.x>
- Davies, M. (2008). *The corpus of contemporary American English*. <http://www.english-corpora.org/coca/>
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Durrant, P., Siyanova-Chanturia, A., Kremmel, B., & Sonbul, S. (2022). *Research methods in vocabulary studies*. John Benjamins.
- Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, 43(4), 559-617. <http://dx.doi.org/10.1111/j.1467-1770.1993.tb00627.x>
- Field, A. (2009). *Discovering statistics using SPSS*. Sage Publications.
- Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology*, 23(2), 84-104.
- Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *Canadian Modern Language Review*, 56(2), 308-328. <http://dx.doi.org/10.3138/cmlr.56.2.308>
- Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, 33, 567-589. <https://doi.org/10.1080/09588221.2019.1577898>
- Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29(4), 694-716. <https://doi.org/10.1093/applin/amn018>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Lee, M., & Jung, J. (2024). Effects of textual enhancement and task manipulation on L2 learners' attentional processes and grammatical knowledge development: A mixed methods study. *Language Teaching Research*, 28(4), 1552-1571. <https://doi.org/10.1177/13621688211034640>
- Li J., Yuan D., & Gao W. (2026). The effects of textual enhancement and individual differences on L2 incidental vocabulary learning via bilingual subtitles. *Journal of Language and Education*, 12(1), 146-159. <https://doi.org/10.17323/jle.2026.26989>

- Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. Cambridge University press.
- Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, 12(1), 107–119. [https://doi.org/10.1016/S0959-4752\(01\)00018-4](https://doi.org/10.1016/S0959-4752(01)00018-4)
- Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: the role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, 43(5), 985–1008. <https://doi.org/10.1017/S0272263121000036>
- Masrai, A. (2020). Can L2 phonological vocabulary knowledge and listening comprehension be developed through extensive movie viewing? The case of Arab EFL learners. *International Journal of Listening*, 34(1), 54–69. <https://doi.org/10.1080/10904018.2019.1582346>
- Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning and Technology*, 18(1), 118–141. <https://doi.org/10.64152/10125/44357>
- Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, 31(2), 1–26. <https://doi.org/10.1080/09588221.2017.1375960>
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle Cengage Learning.
- Nation, I. S. P., & Webb, S. (2017). *How vocabulary is learned*. Oxford University Press.
- Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, 53(4), 1008–1032. <http://doi.org/10.1002/tesq.531>
- Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring Yes–No vocabulary tests: Reaction time vs. nonword approaches. *Language Testing*, 29(4), 489–509. <https://doi.org/10.1177/0265532212438053>
- Puimège, E., Montero Perez, M., & Peters, E. (2021). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study. *Second Language Research*, 39(2), 71–92. <https://doi.org/10.1177/02676583211049741>
- Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity. *Language Learning*, 54(3), 469–523. <http://doi.org/10.1111/j.1467-9922.2007.00415.x>
- Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, 12(1), 107–119. [https://doi.org/10.1016/S0959-4752\(01\)00018-4](https://doi.org/10.1016/S0959-4752(01)00018-4)
- Rodgers, M. P. (2018). The images in television programs and the potential for learning unknown words. *International Journal of Applied Linguistics*, 169(1), 191–211. <https://doi.org/10.1075/itl.00012.rod>
- Rodgers, T.S. (1969). On measuring vocabulary difficulty. An analysis of item variables in learning Russian English vocabulary pairs. *IRAL*, 7(4), 327–343. <https://doi.org/10.1515/iral.1969.7.4.327>
- Rodgers, M. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. [Unpublished doctoral dissertation]. Victoria University.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88. <https://doi.org/10.1177/026553220101800103>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, 15(2), 165–179. <https://doi.org/10.1017/S0272263100011943>
- Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, 63(1), 121–159. <https://doi.org/10.1111/j.1467-9922.2012.00730.x>
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50–73. <https://doi.org/10.64152/10125/44214>
- Szudarski, P. (2012). Effects of input enhancement and explicit instruction on L2 acquisition of formulaic sequences. *International Journal of Applied Linguistics*, 22(3), 315–338. <https://doi.org/10.1111/j.1473-4192.2012.00312.x>
- Teng, M. F. (2022). Incidental L2 vocabulary learning from viewing captioned videos: Effects of learner-related factors. *System*, 105, Article 102736. <https://doi.org/10.1016/j.system.2022.102736>
- Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, 69(2), 405–439. <https://doi.org/10.1111/lang.12335>

- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40(3), 191–210. <https://doi.org/10.1017/S0261444807004338>
- Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, 41(3), 609–624. <https://doi.org/10.1016/j.system.2013.07.012>
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219–258. <https://doi.org/10.1111/j.1467-9922.2010.00593.x>
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2), 130–163. <http://hdl.handle.net/10125/66776>
- Wu, H., & Yang, X. (2022). Effectiveness of textually enhanced captions on Chinese High-school EFL learners' incidental vocabulary learning. *Porta Linguarum Interuniversity Journal of Foreign Language Didactics*, (38), 209–228. <https://doi.org/10.30827/portalin.vi38.23511>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated vocabulary levels test: developing and validating two new forms of the VLT. *International Journal of Applied Linguistics*. 168(1), 34–70. <https://doi.org/10.1075/itl.168.1.02web>

Appendix A

Descriptive Statistics of the Vocabulary Tests

Vocabulary Test	Group	N	Mean	Std. Deviation	Skewness	Kurtosis
Form Recognition Pretest	NC	34	3.74	1.601	-.103-	-1.097-
	TEC	33	3.97	1.591	-.047-	-1.137-
	NoC	30	3.37	1.671	-.297-	.266
	Total	97	3.70	1.621	-.160-	-.592-
Form Recognition Immediate Posttest	NC	34	9.24	3.115	.138	-.791-
	TEC	33	12.15	1.955	-.199-	-.758-
	NoC	30	5.93	2.420	.224	-.271-
	Total	97	9.21	3.565	-.140-	-.932-
Form Recognition Delayed Posttest	NC	34	7.47	2.926	-.119-	-.910-
	TEC	33	7.52	2.682	-.075-	-.571-
	NoC	30	4.40	2.222	.304	-.440-
	Total	97	6.54	2.983	.109	-.807-
Meaning Recognition Pretest	NC	34	3.26	1.729	.088	-.377-
	TEC	33	3.15	1.253	-.101-	-.372-
	NoC	30	3.13	1.306	-.761-	.295
	Total	97	3.19	1.439	-.097-	-.048-
Meaning Recognition Immediate Posttest	NC	34	5.56	2.218	.297	-.555-
	TEC	33	7.03	2.651	.361	-.176-
	NoC	30	6.13	2.543	-.133-	-.625-
	Total	97	6.24	2.524	.262	-.277-
Meaning Recognition Delayed Posttest	NC	34	4.41	2.017	-.016-	-.768-
	TEC	33	5.79	3.189	.538	.662
	NoC	30	5.03	2.953	.111	-.534-
	Total	97	5.07	2.785	.494	.579

Note. Maximum score is 16.

Appendix B*Test of Homogeneity of Variances for the Vocabulary Tests*

		Levene Statistic	df1	df2	Sig.
Form Recognition Pretest	Based on Mean	.053	2	94	.948
	Based on Median	.040	2	94	.961
	Based on Median and with adjusted df	.040	2	89.744	.961
	Based on trimmed mean	.052	2	94	.949
Form Recognition Immediate Posttest	Based on Mean	3.806	2	94	.026
	Based on Median	3.703	2	94	.028
	Based on Median and with adjusted df	3.703	2	84.977	.029
	Based on trimmed mean	3.751	2	94	.027
Form Recognition Delayed Posttest	Based on Mean	1.308	2	94	.275
	Based on Median	1.310	2	94	.275
	Based on Median and with adjusted df	1.310	2	91.655	.275
	Based on trimmed mean	1.324	2	94	.271
Meaning Recognition Pretest	Based on Mean	2.127	2	94	.125
	Based on Median	1.783	2	94	.174
	Based on Median and with adjusted df	1.783	2	86.336	.174
	Based on trimmed mean	2.127	2	94	.125
Meaning Recognition Immediate Posttest	Based on Mean	.830	2	94	.439
	Based on Median	.804	2	94	.451
	Based on Median and with adjusted df	.804	2	92.939	.451
	Based on trimmed mean	.844	2	94	.433
Meaning Recognition Delayed Posttest	Based on Mean	2.587	2	94	.081
	Based on Median	2.413	2	94	.095
	Based on Median and with adjusted df	2.413	2	85.327	.096
	Based on trimmed mean	2.665	2	94	.075