Computer Assisted Language Learning Electronic Journal (CALL-EJ), 25(3), 23-44, 2024

Animated Film as a Linguistic Tool: Enhancing EFL Vocabulary Learning Through Subtitles and Glossing

Hailah Alhujaylan (HAlhujaylan@qu.edu.sa)
Associate Professor
Department of English Language and Literature
College of Languages and Humanities
Qassim University, Saudi Arabia

Abstract

This study aims to explore the impact of entertaining audiovisual material in the form of a full-length animated film on the acquisition of incidental vocabulary by Saudi EFL learners. The study utilized the animated movie *Zootopia* as the audiovisual input for the treatment in three learning conditions: no subtitles (NST), subtitles (ST), and glossed keywords subtitles (GKST); on 127 EFL female participants from a Saudi university. The sample was divided into two sub-samples, each with three groups corresponding to the three learning conditions. Incidental vocabulary learning gains were measured using four tests: meaning recognition, form recognition, meaning recall, and form recall, at three time periods using pre-post-delayed-posttest design. The results showed that groups who watched the animated movie with subtitles and glossed keywords outperformed the no subtitle group in all tests; however, glossed keywords provided a significant improvement to the participants' receptive knowledge of word meaning, along with providing word retention gains in all tests. The study findings may imply that the wide availability of audiovisual resources could be used as an effective linguistic input for EFL learners in class for the purposes of vocabulary learning.

Keywords: audiovisual input; animated movie; glossed keywords; subtitles; vocabulary learning

Introduction

Recent research focusing on English as a second or foreign language learning (ESL/EFL) through audiovisual input has examined the effectiveness of viewing DVD films (King 2002), movies (Webb 2010a; Webb and Rodgers 2009b), documentaries (Peters and Webb 2018), television programs (Vanderplank 1988; Webb 2010a, 20; 2010b), and Netflix series (Fievez et al. 2021) for language learning. The research findings may be varied, but ultimately agreed that employing audiovisual input could improve language comprehension (Peters, Heynen, and Puimège 2016; Vanderplank 1988) and vocabulary knowledge of ESL/EFL learners because they essentially provide a good source of "L2 aural input" (Webb 2010a, 497).

In this technological era, audiovisual inputs like TV series, movies, and cell phones seem to replace book reading for entertainment purposes, although not entirely. Arab News states, 80% of Saudis do not prefer reading in their leisure time (Fayyaz 2013). Therefore, one could presume that they are even less inclined to read academic books for learning purposes. In this regard, audiovisual resources could be an effective accelerator for Saudi EFL learners since they usually encounter serious challenges in learning the English language. The present study proposes the use of audiovisual input to provide a refreshing shift to the dynamics of English language learning in Saudi Arabia, aligning with Vision 2030 goals of employing modern technology in education.

Although previous research has explored L2 incidental vocabulary acquisition using different audiovisual inputs, as cited earlier, few studies have investigated incidental vocabulary learning while viewing an entire film or animated movie. Moreover, considering animated movies as entertaining material, research has paid little attention to using them for pedagogical purposes, and none of the studies have targeted Saudi EFL learners. Building upon this background, the current study addresses this gap and provides a deeper insight into the use of audiovisual input. First, the study examines the impact of English audiovisual input (i.e., an animated movie divided into three segments) augmented with three modes of subtitling on the incidental vocabulary gains on receptive and productive knowledge of Saudi EFL learners. Second, it investigates the incidental vocabulary learning gains from audiovisual input on four different aspects, namely meaning recognition, form recognition, meaning recall, and form recall, which, according to the author's best knowledge, has not been fully addressed in the past research (Peters, Heynen, and Puimège 2016).

Background

Incidental vocabulary learning through audiovisual input

Research suggests that viewing movies for language learning (LL) can motivate learners (Webb 2010c), and may help enhance L2/FL vocabulary knowledge, given the wide availability and popularity of animated movies, cartoons, TV shows, and films among L2 learners. Empirical research shows that watching movies for L2 learning can be a beneficial resource (King 2002; Lee 2015; Webb 2015) since they provide a learning tool for L2 auditory input in foreign settings when learners have restricted opportunities for L2 listening (Webb 2010a). There are several pedagogical reasons why animated films should be encouraged in language classrooms. First, they provide auditory and visual stimuli that can be engaging, relatable, and motivating for L2 learners. Second, L2 viewers can feel and relate their dialogues to the real world (King 2002), which could be a strong reason why L2 learners like viewing animated movies to learn their target language. Third, animated movies can engage learners by providing authentic, comprehensible LL input. However,

as an LL input, movies and TV shows must meet the conditions outlined by Nation (2007) that LL input should be readily accessible in large quantities, entertaining and exciting, familiar to learners, and allow them to build vocabulary from the input due to contextual clues. Moreover, while arousing the viewers' interest, animated movies can consciously and directly involve them in learning L2 vocabulary, grammar, and pronunciation.

When it comes to L2 vocabulary acquisition, most research has looked at how well learners pick up words through reading, extensive reading, or listening to the written content. Other studies assessed the vocabulary learning gains via viewing audiovisual input (Dang, Lu, and Webb 2022; Fievez et al. 2021; Montero Perez, Peters, and Desmet 2017; Peters and Webb 2018; Webb 2010a; Webb and Rodgers 2009b; 2009a). Audio-video streams, TV series, animated movies, and films are the most entertaining sources for L2 learners; however, only a few studies have utilized entertaining audiovisual input for vocabulary learning. Most recently, Fievez et al. (2021) examined the effects of 'entertaining' audiovisual input (i.e., six episodes of a Netflix series) on Dutch EFL speakers. The study was conducted in an out-of-classroom setting, totaling 307 minutes of exposure for 21 days. The findings revealed vocabulary learning gains on form and meaning recall tests. Further, studies have reported that watching animated movies for vocabulary learning should be practiced regularly (Lee 2015), and in a continuous stream (Vanderplank 1990) for the maximum benefit.

Using subtitling and captioning for language learning

Research shows that captioning and subtitling have been generally regarded as a learning support for developing language proficiency (Montero-Perez et al. 2014; Talavan 2010). In addition, many studies have shown that audiovisual input with captioning and subtitling assist L2 learners in comprehending content that might otherwise be beyond their comprehension (Hsieh 2020; Peters, Heynen, and Puimège 2016). A large body of research has examined the effects of using captioning and subtitling on listening skills (Danan 2004) reading comprehension (Rodgers and Webb 2017) and vocabulary learning (Dang, Lu, and Webb 2022; Hsieh 2020; Montero Perez, Peters, and Desmet 2017; Pujadas and Muñoz 2019) development. Much evidence has indicated that subtitling correlates positively with SL/FL learning from the earlier mentioned linguistic aspects. For instance, Danan (2004) noted that subtitles encourage viewers by helping them understand dialogues in audiovisual input. Further, Rodgers and Webb (2017) observed a minor difference in learning scores between *captions* and *no captions* groups, suggesting that captioning may aid L2 learners in the comprehension of episodes with particularly difficult English.

Although most studies indicated that subtitling is helpful for LL and vocabulary learning, other research proposed that subtitles may have some negative aspects as well. For example, it could make the learner too dependable on the text and exhibit laziness instead of putting their effort into comprehending what is said (Danan 2004). Also, subtitles can be distracting for learners, especially if they are not used to watching with subtitles (Vanderplank 1988). Furthermore, it is also possible that subtitles could only assist learners with lower language proficiency. Rodgers and Webb (2017) have a similar opinion that captions help lower-level learners comprehend.

Studies have yet to evaluate how various modes of L2 subtitling affect vocabulary acquisition. In this regard, Montero Perez, Peters, and Desmet (2017) adopted a unique approach to examining audiovisual input by employing two enhancement techniques: using various types of subtitling and testing with or without announcement on eight learning conditions. The findings revealed that the group with glossed keyword meanings outperformed all other groups in form recognition and meaning recall tests. Similarly, Hsieh (2020) compared the effects of different caption types on

low-intermediate Chinese EFL learners' listening comprehension and vocabulary knowledge. The study suggested that the caption type significantly impacted vocabulary acquisition more than listening comprehension, with the latter portraying no increase. Moreover, Hsieh (2020) found that captions did not affect the learning of written text/words, suggesting that students may pay closer attention to vocabulary if captioned words are highlighted or glossed in videos.

Glossed caption or subtitles have proved their utility in incidental vocabulary acquisition (Yoshii 2006). "Glossing is when text is enhanced by providing the first language (L1) or L2 meanings of difficult words within the text" (Webb 2010c, 202). Again, regarding glossing in audiovisual input, the research has been limited. However, it sounds intuitive that glossing could potentially have positive effects on vocabulary learning for EFL learners, and it needs further exploration in future research.

To fill the research gap, the present study applied three different modes of subtitles to gauge their effects on incidental vocabulary learning. Unlike previous studies, the audiovisual input used in this study is entertaining: a full animated movie divided into three segments.

Research questions

The study's objective is to investigate how integrating audiovisual input in a Saudi learning context might help EFL learners improve their incidental vocabulary knowledge. Furthermore, it examines the relationship of different subtitling modes with developing receptive and productive aspects of vocabulary knowledge. The current study seeks to provide answers to the following research questions:

RQ1: To what extent does the mode of subtitling affect EFL learners' receptive vocabulary knowledge, as measured by the meaning recognition and form recognition tests?

RQ2: To what extent does the mode of subtitling affect EFL learners' productive vocabulary knowledge, as measured by the meaning recall and form recall tests?

Methodology

Participants

The final study sample consisted of $N = 127$ undergraduate female EFL students at a Saudi University. It should be noted that the final study sample was divided into two sub-samples ($N_1 = 62$, $N_2 = 65$), as discussed in the following sections. The students' ages ranged from 19 y/o to 22 y/o ($M = 20.11$, $SD = 0.946$). According to an investigation, on average, Saudi students have an English vocabulary level of between 1650 and 3000 words when they first enroll in college (Al-Masrai and Milton 2012). By the time they graduate, they typically know roughly between 3000 and 5000 words. Since the participants in this research were midway to their graduation, their vocabulary level was estimated to be at the upper end of the 1650–3000 word range and at the lower end of the 3000-5000 word range, or approximately around 3000 words.

A preliminary survey to gauge the participants' experience of watching movies in the English language suggested that they had little to no experience of watching movies in a classroom setting for learning purposes (89% declared no experience; 7% reported a little experience; 3% stated advanced experience and 1% did not prefer to answer). The data of several participants who could not attend one of the treatment sessions or did not meet the cut-off point for the comprehension test were discarded from the study (See Section 4.3.1). To attribute learning gains

to the treatment, participants were instructed not to view any English-language movies during the research.

Materials

The audiovisual input for the study was an animated movie called *Zootopia (UK Title: Zootropolis)*, a 2016 American computer-animated buddy cop film produced by Walt Disney Animation Studios and released by Walt Disney Pictures. A pilot survey among students of similar backgrounds and academic levels revealed that *Zootopia* was found to be a more popular choice (86%) compared to *Kung Fu Panda 3* (9%) and *Moana* (5%). The author confirmed that none of the final study sample had watched *Zootopia* or its clips before the study. This study used an original animated movie; therefore, the frequency of dialogue words was uncontrollable. The estimated frequency of the target words (TWs) was verified from Lexitutor VocabProfiler (Cobb, T. Compleat Web VP v.2.6 [computer program]. Retrieved on 4 March 2022 from: https://www.lextutor.ca/vp/comp/) within the combined British National Corpus (BNC) and Corpus of Contemporary American English (COCA) 25 k/c category. A list of the first 3,000-word level was compiled, and the TWs of vocabulary were set to be 84 for the study, meticulously extracted from *Zootopia's* script. TWs were selected on the following criteria: not a proper noun, not a preposition, words with self-contained meaning, and a word level of 3K. The word list was approved by three English L1 instructors. Subtitles were presented to the participants in their L2 (English).

Measuring instruments

Comprehension test

After the viewing of three movie segments, all participants immediately took the comprehension test to ensure that they viewed the movie segments attentively and understood the content. Consequently, the test was marked, and a cut-off point of 68% of the total score was set to determine the participants' eligibility for proceeding with the study. Four participants from sub-sample 1, and one from sub-sample 2 did not make the cut-off point and thus, were excluded from the study sample. In addition, six students who missed one or more treatment sessions were sorted out as well, thereby reducing the total number of participants to $N = 127$, whose data was included in the results. The author chose this course of action to further transparentize the results since the participants who did not understand the movie content or missed any sessions could have inflated the results later in the study. However, the results of comprehension tests were not a function of analyses and were restricted to descriptive statistics only.

Vocabulary Tests

Four different types of tests were constructed from the TWs, namely: 1) meaning recognition test; 2) form recognition test; 3) meaning recall test, and 4) form recall test. Among these, sub-sample 1 took recognition tests (meaning and form), and sub-sample 2 took the recall tests (meaning and form).

The tests were designed scrupulously with the assistance of two English L1 speakers with Ph.D. degrees in the subject. The 84 TWs were divided randomly and equally into two sets; set A and set B (*Appendix A, Table A.1*). For example, set A of 42 TWs was given to sub-sample 1 in

the meaning recognition test, and set B for the form recognition test. A similar pattern was used for sub-sample 2. The TWs were divided for two reasons: 1) learners would be prevented from carrying any testing-related acquisition; 2) to deal with testing time constraints. The author ensured that the tests had clear-cut objectives and rubrics; therefore, they would be conveniently markable. Each of the four tests is described in detail in the following sub-sections.

(a). Receptive vocabulary knowledge tests

(1). Meaning recognition test (Sub-sample 1)

The meaning recognition test aimed to assess the participants' word meaning retention. The test comprised of 42 items with four options each, given against the TWs. The participants were required to recognize the correct option and select it. A sample question from the meaning recognition test is given below:

Tick the one from the following options which correctly matches the word meaning as closely as possible:

1. scared

□ relieved □ excited □ unhappy □ frightened

(2). Form recognition test (Sub-sample 1)

For the form recognition test, a yes/no format was constructed. If the participants recognized the word form, they were required to tick yes and vice versa. A third option, not sure, was also introduced to reduce the probability of participants' guessing the correct answer. An example of the question in the form recognition test is as follows:

Are you able to recognize the following word:

1. apologize

□ Yes □ No □ Not sure

(b). Productive vocabulary knowledge tests

(1). Meaning recall test (Sub-Sample 2)

The meaning recall test required the participants to go through the TWs and respond with their meanings. The test primarily focused on recalling the target word meanings and the participants could answer in L1 as well. The reason was to not let the L2 writing skill interfere with testing. The example question is given below:

Write the meaning of the following word. (You may use Arabic)

scared = _____ (Ans: مفزوع/frightened)

(2). Form recall test (Sub-Sample 2)

The form recall test consisted of fill-in-the-blank questions, which the participants had to complete by recalling the target word form. The example question is as follows:

> **Fill in the blanks:**
>
> I must *apologize* for accusing her falsely.

Procedure

In this study, participants were subjected to three different learning conditions:

(1) No subtitles (NST) – (*baseline*)
(2) Subtitle (ST)
(3) Glossed Keywords Subtitles (GKST)

The study comprised two experimental and a baseline group for each sub-sample (Figure 1). The ST group watched the movie segments with full subtitles, thereby interacting with the reading ability along with viewing and listening abilities. GKST group watched the movie segments with subtitles that had glossed keywords with meanings, further intensifying the reading aspect. Finally, NST (baseline group) watched the animated movie segments without subtitles.

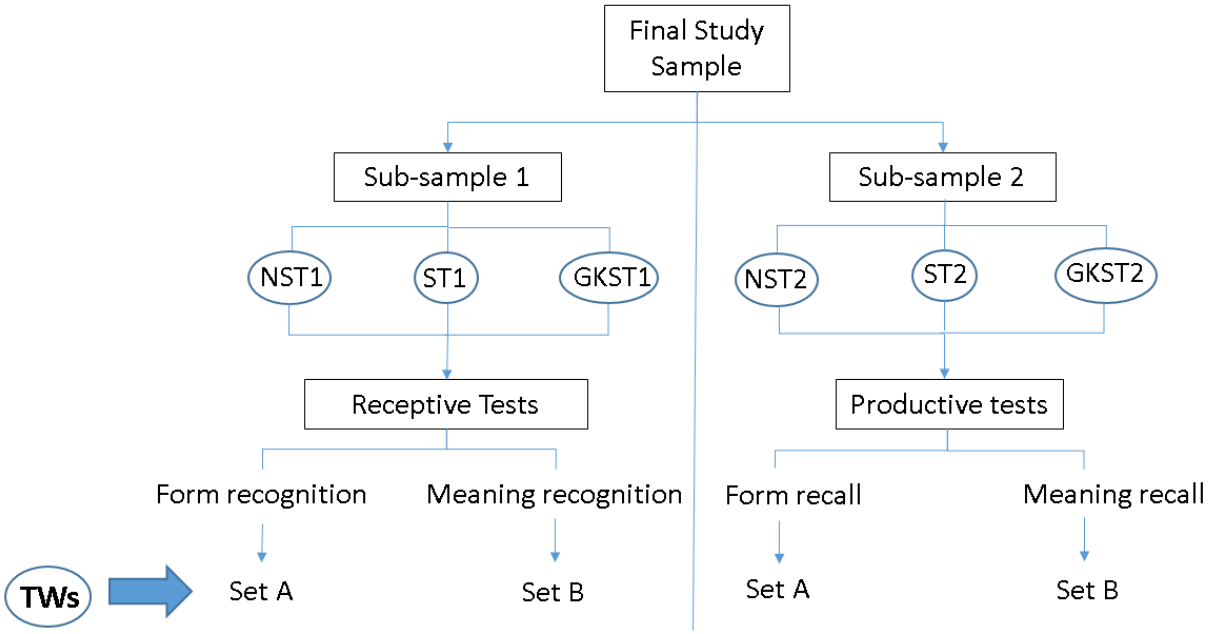

Figure 1. Procedure design

Webb, Newton, and Chang (2012) found that repeated word encounters or repetition could lead to learning gains if the same participants took the recognition and recall tests (form and meaning) with the same TWs, as found in the pilot study. Therefore, the final study sample was segregated into two sub-samples in order to eliminate the variable of repeated word encounter. Another benefit of this segregation was the independent analyses of the participants' learning gains in receptive knowledge (recognition) and productive knowledge (recall). Moreover, the specific

order of conducting recognition and recall tests together for a single sample has proven to be a strong concern in previous studies (Montero Perez, Peters, and Desmet 2017), however, it never posed any issues in this study.

Another precaution to restrict the learnability of TWs was carried out by dividing them into two sets (for details see Section 4.3.2). In any case the participants could not encounter a TW twice, even within a sub-sample, therefore, entirely eliminating the repeated word encounter factor from the study.

During the intervention period, the animated movie was divided into three segments, as per the storyline of the movie; beginning, climax, and ending, as illustrated in Figure 3. Each segment was shown to the participants every week. Rodgers and Webb (2017) recommended a full-length television show (between 22 and 42 minutes) to learners for vocabulary enhancement; therefore, the researcher divided the animated movie into three segments to achieve this viewing length. Another reason was to keep the participants interested in the storyline and to avoid unnecessary information or irrelevant scenes that would be culturally inappropriate to be shown in the classroom.

Three English L1 speakers assessed the movie's linguistic content, and four Saudi EFL instructors deemed the segmented version culturally appropriate for the classroom. For GKST1 and GKST2, Mideo media player played the animated movie. The Mideo media player could provide them with glossed keywords in subtitles, and they could pause the screen to seek up the highlighted keywords' meanings through a pop-up window (Figure 2). In contrast, the NST group had no subtitles, while the ST group had full subtitles but no access to glossed keywords or meanings. The playing speed of the movie segments was set to 0.95x the normal, owing to the fact that 89% of the participants had little to no experience of watching movies in English prior to the study; therefore, they could face difficulty understanding the dialogues at full speed.

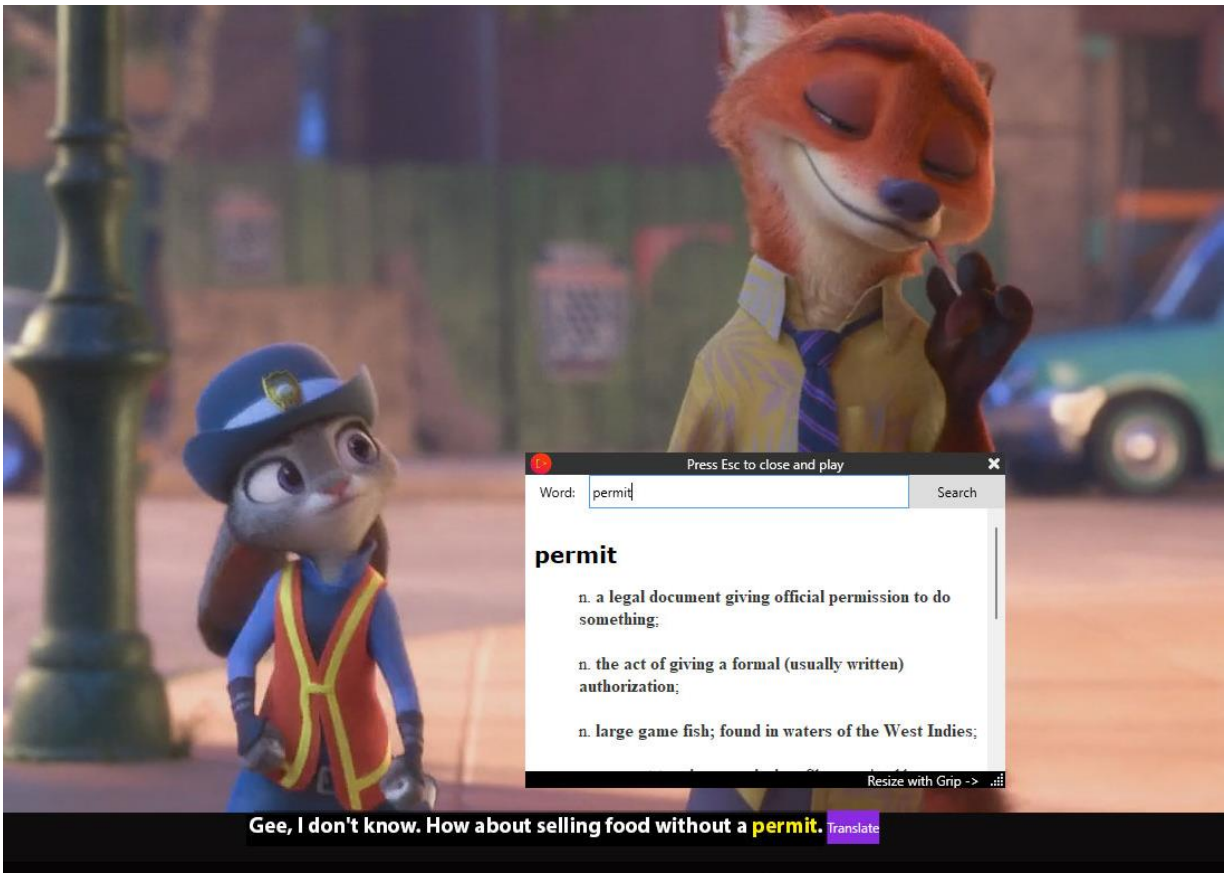

Figure 2. The screenshot from Mideo media player with glossed keyword subtitle

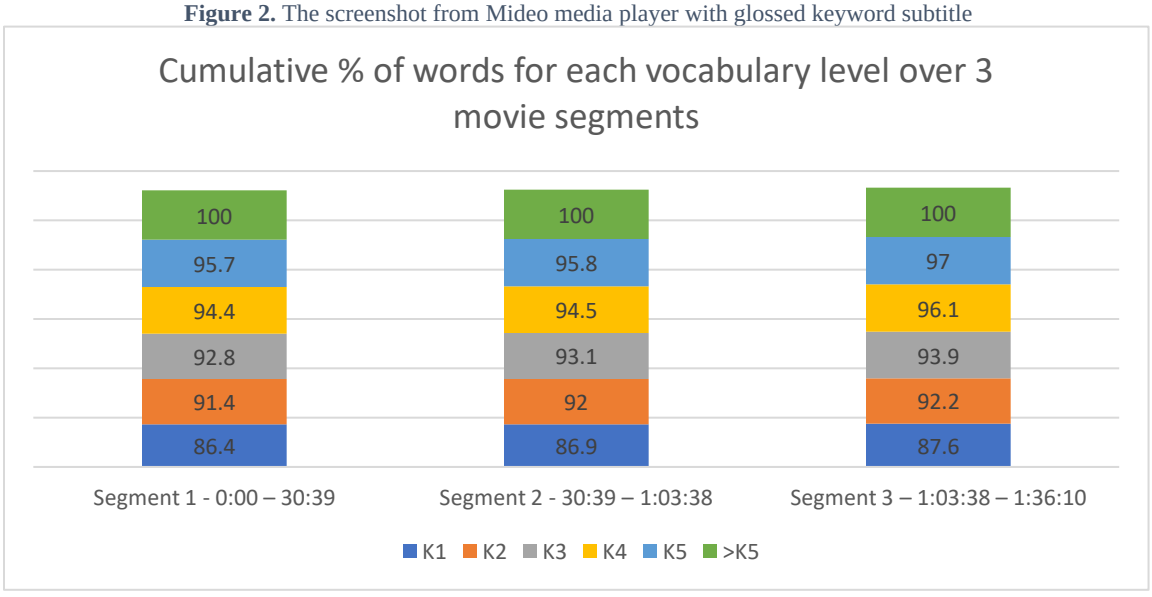

Figure 3. Cumulative % of words for each vocabulary level over 3 movie segments

The research period lasted for twelve weeks. The weekly treatment plan is given in Table 1.

Table 1. Weekly research schedule

Week	Treatment
1	Pretest
2-3	Break
4	Segment 1 viewing
5	Segment 2 viewing
6	Segment 3 viewing
7	Comprehension test
8	Break
9	Immediate posttest
10-11	Break
12	Delayed posttest

Results

Comprehension test

The descriptive statistics for the comprehension test result for sub-sample 1 and sub-sample 2, including mean, SD, and 95% CI are shown in Table 2 and Table 3. The findings show that the participants of both sub-samples sufficiently comprehended the movie content and passed the cut-off point of 17/25. Moreover, one-way ANOVA results revealed that none of the groups showed any significant difference between each other regarding the movie content comprehension for sub-sample 1 [F (2, 59) = 0.23, p = .792], and sub-sample 2 [F (2, 62) = 0.72, p = .491], where α = 0.05.

Table 2. Descriptive statistics for comprehension test (sub-sample 1)

Group	N	Mean	SD	95% CI
NST1	20	19.95	3.762	(18.19, 21.71)
ST1	22	20.73	3.089	(19.36, 22.10)
GKST1	20	20.05	5.060	(17.68, 22.42)

Table 3. Descriptive statistics for comprehension test (sub-sample 2)

Sample	N	Mean	SD	95% CI
NST2	21	20.24	2.931	(18.90, 21.57)
ST2	22	19.50	2.841	(18.24, 20.76)
GKST2	22	20.59	3.418	(19.08, 22.11)

Meaning recognition test

Since there was an unequal number of participants in each group, Levene's test for equality of variances was carried out, and no violation of homogeneity of variances were found for posttest (p = .297), and delayed posttest (p = .383), where the sig. level was set to be 0.05. The descriptive statistics for the meaning recognition test are given in Table 4.

Table 4. Descriptive statistics for meaning recognition test

Group		Pretest		Posttest		Delayed posttest	
	N	Mean	SD	Mean	SD	Mean	SD
NST1	20	15.20	2.375	18.45	3.620	17.60	4.967
ST1	22	14.73	2.746	22.77	4.730	20.36	4.756
GKST1	20	13.75	3.177	26.45	3.316	23.55	5.661

All three groups showed an overall mean score improvement over time. The mean scores increased from pretest to posttest but showed a subsequent decrease in the delayed posttest (Table 4). Furthermore, according to the results in Table 5, significant differences were revealed between the groups over time (F (4, 118) = 7.11, p = .000, η_p^2 = .194), which depicts that the type of subtitle mode significantly affected the word meaning recognition ability of the participants over time.

Table 5. 3 x 3 Mixed ANOVA for meaning recognition test [group (NST1, ST1, GKST1) x time (pretest, posttest, delayed posttest)]

Source	SS	df	MS	F	p	η_p^2
Between Subjects						
Group	521.46	2	260.73	17.56	.000	.373
Error	876.03	59	14.85			
Within Subjects						
Time	2135.49	2	1067.75	61.11	.000	.509
Time x Group	496.61	4	124.15	7.11	.000	.194
Error	2061.89	118	17.47			

*p<0.05 is significant

Furthermore, post-hoc tests based on estimated marginal means (EMM) with Bonferroni adjustment revealed that all groups differed significantly (Table 6) with GKST1 (M = 21.25, SD = .497) scoring significantly higher in the meaning recognition test than ST1 (M = 19.29, SD = .474) and NST1 (M = 17.08, SD = .497) (p < .05). Moreover, ST1 mean score was also significantly higher than NST1 mean score (p < .05). These findings suggest a positive relationship between meaning recognition aspect of vocabulary and the availability of subtitles, with glossed keyword subtitle mode providing even higher improvement compared to the subtitle only mode, as shown by the comparatively steeper gradient of GKST1 compared to ST1 in Figure 4.

Table 6. Post-hoc tests for meaning recognition test

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
NST1	ST1	-2.20*	.69	.006	-3.86	-.55
	GKST1	-4.17*	.70	.000	-5.86	-2.48
ST1	NST1	2.20*	.69	.006	.55	3.86
	GKST1	-1.96*	.69	.016	-3.61	-.31
GKST1	NST1	4.17*	.70	.000	2.48	5.86
	ST1	1.96*	.69	.016	.31	3.61

*. The mean difference is significant at the .05 level.

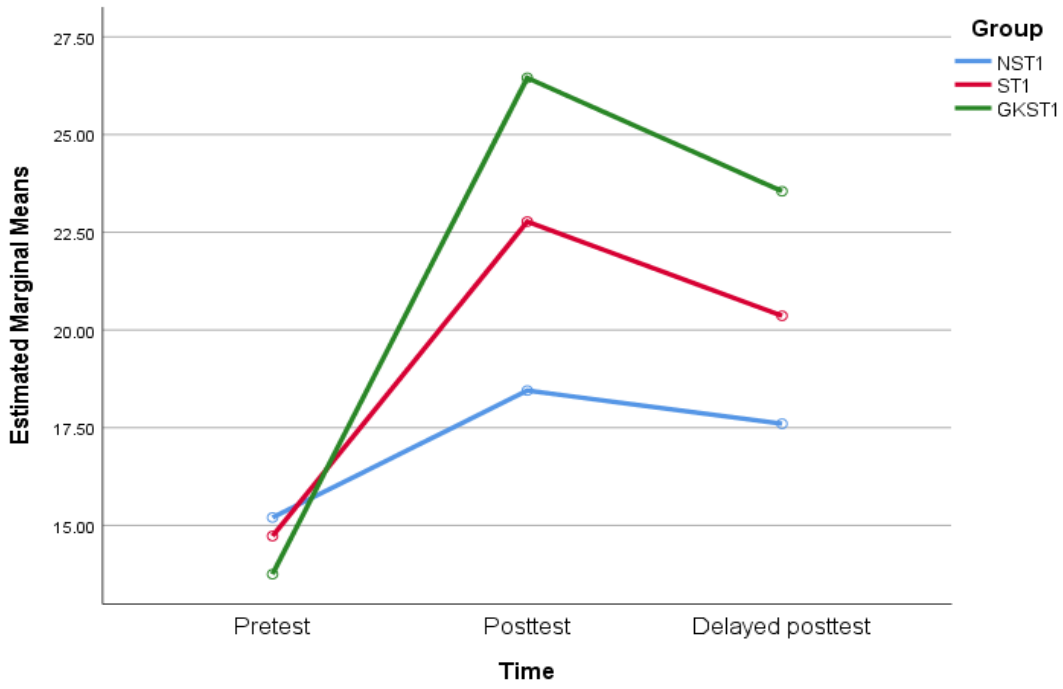

Figure 4. EMM over time for all groups (meaning recognition test)

Form recognition test

Levene's test for equality of variances revealed no violation of homogeneity of variances for posttest (p = .399), and delayed posttest (p = .335) where the sig. level was set to be 0.05. The descriptive statistics for the form recognition test are given in Table 7.

Table 5. Descriptive statistics for the form recognition test

Group		Pretest		Posttest		Delayed posttest	
	N	Mean	SD	Mean	SD	Mean	SD
NST1	20	19.05	3.620	22.80	5.952	21.75	4.723
ST1	22	19.45	6.300	27.73	7.291	26.59	5.688
GKST1	20	18.95	7.193	29.85	6.098	29.40	6.468

From the pretest to the posttest, the mean scores rose for all groups; but the delayed posttest revealed a consequent decline in the mean scores (Table 7). As far as the differences between groups' mean scores were concerned, they were not significant over time (F (4, 118) = 2.41, p = .053, η_p^2 = .076), as shown by the insignificant *time x group* interaction term (Table 8), indicating that the type of subtitle mode could not considerably influence the participants' form recognition aspect of vocabulary over the course of three testing points, even though the main effects were found to be significant for both *group* and *time*.

Table 8. 3 x 3 Mixed ANOVA for form recognition test [group (NST1, ST1, GKST1) x time (pretest, posttest, delayed posttest)]

Source	SS	df	MS	F	p	η_p^2
Between Subjects						
Group	749.58	2	374.79	12.07	.000	.290
Error	1831.29	59	31.04			
Within Subjects						
Time	2163.19	2	1081.60	27.58	.000	.319
Time x Group	378.26	4	94.57	2.41	.053	.076
Error	4628.05	118	39.22			

*. The mean difference is significant at the .05 level.

In addition, post-hoc tests based on estimated marginal means (EMM) with Bonferroni adjustment revealed that both ST1 (M = 24.59, SD = .686) and GKST1 (M = 26.07, SD = .719) differed significantly from the NST1 (M = 21.20, SD = .719) at p < .05 (Table 9). However, ST1 mean score did not differ significantly compared to that of GKST1 mean score (p = .305), although GKST1 displayed better mean score improvement. When combined with the line graph in Figure 5, these results indicate that the addition of glossed keyword in the subtitles enhanced the GKST1 participants' form recognition abilities, albeit not significantly compared to the subtitle only mode for ST1.

Table 9. Post-hoc tests for form recognition test

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
NST1	ST1	-3.39*	.99	.003	-5.78	-1.00

ST1	GKST1	-4.87*	1.02	.000	-7.31	-2.42
	NST1	3.39*	.99	.003	1.00	5.78
GKST1	GKST1	-1.48	.99	.305	-3.87	.91
	NST1	4.87*	1.02	.000	2.42	7.31
	ST1	1.48	.99	.305	-.91	3.87

*. The mean difference is significant at the .05 level.

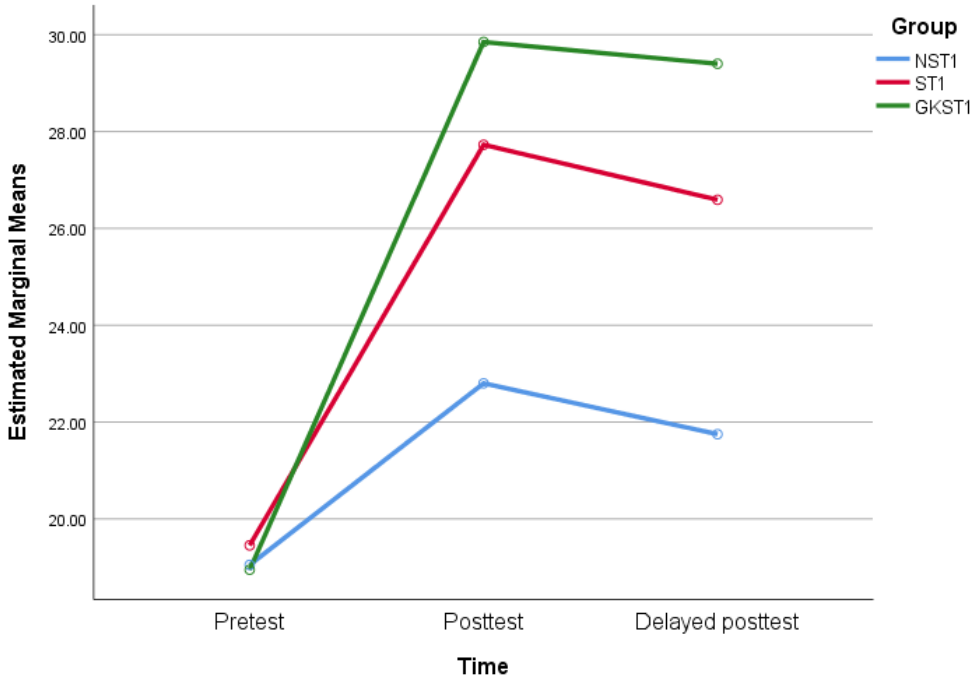

Figure 5. EMM over time for all groups (form recognition test)

Meaning recall test

Levene's test for equality of variances revealed no violation of homogeneity of variances for posttest (p = .573), and delayed posttest (p = .512) where the sig. level was set to be 0.05. The descriptive statistics for the meaning recall test are given in Table 10.

Table 6. Descriptive statistics for meaning recall test

Group		Pretest		Posttest		Delayed posttest	
	N	Mean	SD	Mean	SD	Mean	SD
NST2	21	9.48	2.272	12.33	3.039	11.19	3.763
ST2	22	10.27	4.355	19.45	6.486	17.64	4.510
GKST2	22	10.05	2.081	18.82	5.114	21.23	4.800

The results portray that NST2 and ST2 showed an overall mean score improvement over time, with a slight drop from posttest to delayed posttest. However, GKST2 mean score did not follow this pattern, since its mean score increased from the posttest to the delayed posttest (Table 10). The reason for this finding could be attributed to the stronger word meaning retention due to the glossed keyword subtitles, which is considered in detail in the Section 6. Moreover, as shown by the results in Table 11, there were significant differences over time across the groups (F (4, 124) = 6.54, p = .000, η_p^2 = .174), indicating that the type of subtitle mode had a definite effect on the participants' recall of the word meaning.

Table 11. 3 x 3 Mixed ANOVA for meaning recall test [group (NST2, ST2, GKST2) x time (pretest, posttest, delayed posttest)]

Source	SS	df	MS	F	p	η_p^2
Between Subjects						
Group	1199.32	2	599.66	55.85	.000	.643
Error	665.64	62	10.74			
Within Subjects						
Time	2030.61	2	1015.30	45.97	.000	.426
Time x Group	577.59	4	144.40	6.54	.000	.174
Error	2738.51	124	22.08			

*p<0.05 is significant

In addition, ST2 (M = 15.79, SD = .403) and GKST2 (M = 16.70, SD = .403) differed significantly (p < .05) compared to NST2 (M = 11.00, SD = .413), as shown by the post-hoc tests based on estimated marginal means (EMM) and Bonferroni adjustment, given in Table 12. However, no significant difference was visible between ST2 and GKST2 (p = .256). These results highlight that even though the participants who watched the animated movie with glossed keyword subtitles displayed better word meaning retention, they did not perform statistically better over time in the meaning recall test compared to those who received the audiovisual input in subtitle only mode, as shown by similar gradients of GKST2 and ST2 until posttest in Figure 6.

Table 12. Post-hoc test for meaning recall test

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
NST2	ST2	-4.79*	.58	.000	-6.17	-3.40
	GKST2	-5.70*	.58	.000	-7.08	-4.31
ST2	NST2	4.79*	.58	.000	3.40	6.17
	GKST2	-.91	.57	.256	-2.28	.46
GKST2	NST2	5.70*	.58	.000	4.31	7.08
	ST2	.91	.57	.256	-.46	2.28

*. The mean difference is significant at the .05 level.

Figure 6. EMM over time for all groups (meaning recall test)

Form recall test

Levene's test for equality of variances revealed no violation of homogeneity of variances for posttest (p = .935), Delayed posttest (p = .250) where the sig. level was set to be 0.05. Table 13 shows the descriptive statistics for the form recall test.

Table 7. Descriptive statistics for form recall test

Group		Pretest		Posttest		Delayed posttest	
	N	Mean	SD	Mean	SD	Mean	SD
NST2	21	14.76	5.009	17.24	6.236	15.24	3.974
ST2	22	13.73	3.042	21.68	5.939	19.68	6.357
GKST2	22	13.36	2.498	22.09	5.079	21.82	4.717

The mean scores of all three groups increased from pretest to posttest but showed a subsequent decrease in the delayed posttest (Table 13). Furthermore, according to the results in Table 14, significant differences were revealed between all the groups over time (F (4, 124) = 4.36, p = .002, η_p^2 = .123), which depicts that the type of subtitle mode exhibited a significant effect on the form recall aspect of the participants' vocabulary over time.

Table 14. 3 x 3 Mixed ANOVA for form recall test [group (NST2, ST2, GKST2) x time (pretest, posttest, delayed posttest)]

Source	SS	df	MS	F	p	η_p^2
Between Subjects						
Group	396.48	2	198.24	7.91	.001	.203
Error	1554.00	62	25.06			
Within Subjects						
Time	1460.25	2	730.12	30.53	.000	.330
Time x Group	417.07	4	104.27	4.36	.002	.123
Error	2965.52	124	23.92			

Like all of the previous tests except meaning recognition, groups with subtitles and glossed keyword subtitles, i.e., ST2 (M = 18.36, SD = .616) and GKST2 (M = 19.09, SD = .616) differed significantly ($p < .05$) compared to NST2 (M = 15.75, SD = .631), as shown by the post-hoc tests based on estimated marginal means (EMM) and Bonferroni adjustment, given in Table 15. However, no significant difference was visible between ST2 and GKST2 (p = .683). But again, GKST2 showed the best performance in the delayed posttest compared to the other groups, pointing to a possible better retention of word form. These findings suggest that although viewers who received glossed keyword subtitles showed improvement in word form retention, they did not fare statistically better than viewers of the audiovisual input in subtitle-only mode over time in the form recall test, as evidenced by the roughly comparable gradients of GKST2 and ST2 until posttest in Figure 7.

Table 15. Post-hoc test for form recall test

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
NST2	ST2	-2.62*	.88	.012	-4.74	-.50
	GKST2	-3.34*	.88	.001	-5.46	-1.23
ST2	NST2	2.62*	.88	.012	.50	4.74
	GKST2	-.73	.87	.683	-2.82	1.37
GKST2	NST2	3.34*	.88	.001	1.23	5.46
	ST2	.73	.87	.683	-1.37	2.82

*. The mean difference is significant at the .05 level.

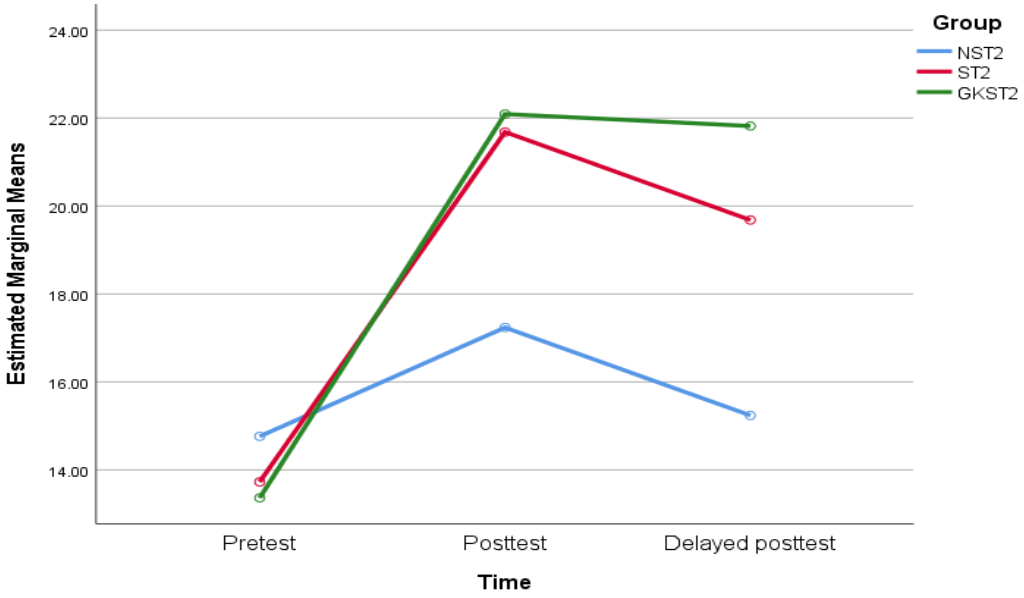

Figure 7. EMM over time for all groups (form recall test)

Discussion

This study examined the effects of audiovisual input with different subtitling modes on EFL learners' vocabulary. The quantitative results showed that participants who watched animated movie segments with subtitles and glossed keywords subtitles scored better on receptive and productive vocabulary measures than the no subtitle group.

RQ1: To what extent does the mode of subtitling affect EFL learners' receptive vocabulary knowledge, as measured by the meaning recognition and form recognition tests?

To answer RQ1, the results showed that significant differences were visible between the three groups of the study for the meaning recognition test over time due to different modes of subtitling ($p < .05$), with the posthoc tests confirming the best performance of participants who received audiovisual input with glossed keyword subtitles. On the other hand, form recognition test results depicted an insignificant *time x group* interaction ($p > .05$), which highlights that the type of subtitle mode could not significantly affect the participants' form recognition aspect of vocabulary throughout the duration of three testing points. However, since the *p*-value is very close to the significance level of 0.05, it could be argued that had there been more statistical power for this study, a significant p-value could have been achieved. Future research could explore this domain with a greater sample size. In line with this speculation, post-hoc tests did show significant differences between the no-subtitle group and groups who received subtitles in either subtitle-only mode or glossed keyword subtitle mode. The findings show that subtitles appear to improve receptive vocabulary, as measured by meaning and form recognition tests.

The study's findings support previous research suggesting a positive association between subtitled audiovisual input and vocabulary acquisition, such as Peters, Heynen, and Puimège's (2016) study concluded. For the present research, one possible explanation of higher learning gains in the meaning recognition test may be the combination of audiovisual input and the on-screen text presented in subtitles. Research indicates that the learning material provided through the combination of animation and words could be more understandable for acquiring new words (Rodgers and Webb 2017; Vanderplank 1988), which logically yields better results. The learners

receive information from various modes (verbal/visual) and can connect them better than presented in one mode. Paivio (1990) sees this combination as verbal (words) and nonverbal referents (objects, pictures, events, etc.). Learning could easily occur if the association between these two becomes more robust and comprehensible (as in the case of glossed keywords subtitles). This combination may "lead to the conscious focusing on the form (especially on the correct word form)" (Vanderplank 1988, 276) of unfamiliar words, as shown by the results in the post hoc tests of form recognition test in this study. Furthermore, this combination might allow learners to concentrate better on novel or hard words, which they could not be able to acquire otherwise. In fact, Rodgers and Webb (2017) say that while reading captions, learners apply their reading skills, which may enhance their language comprehension.

Despite the mean score increment for all groups, it is possible that subtitles with glossed keywords may have helped GKST1 score higher than ST1 and NST1 over time. Thus, GKST1 participants answered 30 out of 42 questions correctly in the form recognition posttest, on average, while ST1 did 28 words correctly. These findings align with Montero Perez, Peters, and Desmet (2017), indicating that learners become more attentive when seeing glossed words in the subtitles. Given that, GKST1 had access to the meaning of the glossed keywords, and they could have involuntarily acquired a better understanding of what was being said. Hence, they showed an improvement in the posttests and delayed posttests. Also, they could see the application of the target word in the context by viewing the respective images in the animated movie segments. Still, the inclusion of glossed keywords did not significantly improve over the subtitles-only regime for form recognition. This finding lines up with Peters, Heynen, and Puimège (2016) because they did not draw a conclusive result on form recognition aspect of word knowledge. Also, a recent study by Dang, Lu, and Webb (2022) reported learning gains, although insignificant, in the form recognition test through various input modes such as reading, listening, reading while listening, viewing, and viewing with captions.

Interestingly, the delayed posttest means for GKST1 were higher than all other groups in both meaning and form recognition tests. These findings support previous research showing that glossing in any format improves vocabulary learning (Watanabe 1997). The finding above may be explained by the fact that learners pay more attention to the glossing while taking in the scene's visual information, making it easier to remember the target words later on, even if they do not fully comprehend them. Furthermore, if it is in the form of visuals, as an animated movie for example, it could "greatly strengthen or enhance the verbal message" (Vanderplank 1990, 223).

RQ2: To what extent does the mode of subtitling affect EFL learners' productive vocabulary knowledge, as measured by the meaning recall and form recall tests?

Regarding RQ2, the results clearly show that the participants' vocabulary recall was significantly affected by the subtitle mode in both meaning recall and form recall aspects (p <.05). Previous research findings indicate that the groups viewing videos without subtitles scored lower than those with subtitles (Montero-Perez et al. 2014; Vanderplank 1988), which is similar to the results obtained through the present study, i.e., NST2 group showed lower performance than ST2 and GKST2 groups in both meaning recall and form recall tests. The overall improvement of ST2 and GKST2 can be summarized with Danan's (2004) finding that if the audiovisual input is at the same proficiency level as the learner, it is possible that word learning could take place. In the present study, the vocabulary level of the words was set according to the vocabulary size of the participants. This is an important finding since it explains why word learning is possible in the present study's context.

GKST2 seemed to portray a strong knowledge retention in both meaning recall (M = 21.23, SD = 4.800) and form recall aspects (M = 21.82, SD = 4.717). Vanderplank (1988) believed that rich audio video input may motivate learners for stronger word retention, and subtitles could assist learners in acquiring new words consciously, as Canning-Wilson (2000) asserted. Similarly, Peters, Heynen, and Puimège (2016) stated that on-screen aids have the potential to improve form learning for EFL learners. Relating these to the results of this study, there seems to be a significant boost in word retention owing to the keyword glossing. The audiovisual input was "enriched" by adding glossed keywords, and it possibly drew GKST2 learners' attention to those words (Nation 2001), creating a more powerful word meaning association for language use. Dang, Lu, and Webb (2022) considered learners' attention to the input as an essential condition in incidental vocabulary learning. It is to be noted that glossed keyword subtitling might have facilitated the productive aspect of vocabulary knowledge. Learners' conscious effort may result in stronger retention and recall of the lexicon, as observed in the case of GKST2. Moreover, the present findings corroborate the results of Montero Perez, Peters, and Desmet's (2017) study that the glossed captioned group improved most vocabulary learning gains on meaning recall tests. Similar are the findings of Fievez et al. (2021) that glossed captions can positively impact vocabulary learning.

On the other hand, NST2 showed the least learning gains on form recall and meaning recall tests compared to ST2 and GKST2. Van Zeeland and Schmitt's (2013) study suggests that learners could build a better connection of form-meaning through verbal and written input. Expanding on the previous studies, the present study indicated that NST2 might not be able to build a form-meaning link in the absence of subtitling. Another explanation of the lower NST2 gains could be the higher processing demand of audiovisual input as Peters, Heynen, and Puimège (2016) mentioned. NST2 did not have any learning aid other than the on-screen images used in the animated movie and they also could not go back to reading the text to understand the word form and meaning in the context. Therefore, the processing load of the lexicon might be higher than ST2 and GKST2 participants, who had texts and glossed key words as on-screen aids respectively.

Limitations and Conclusion

The author acknowledges the study's limitations. First, the listening skills were not tested in the comprehension test. Second, the researcher could have examined the learning of vocabulary chunks from the animated movie. However, due to the short time and limited resources, it was practically not possible for the researcher. The study's findings may have implications for the instructors and curriculum designers. More research is needed to explore the effects of entertaining audiovisual input on the different aspects of vocabulary knowledge.

Building on the previous research, this study investigated the effectiveness of viewing audiovisual input with different modes of subtitling on incidental vocabulary acquisition in a Saudi learning context. This study contributes to filling the existing literature gap by its rich findings that through audiovisual entertaining material, Saudi EFL learners can acquire various aspects of vocabulary in a short time. The present study's findings indicate that the inclusion of subtitles in an animated movie can lead to increased vocabulary learning gains. The participants viewing the animated movie with subtitles and glossed keywords subtitles were able to make significant gains over those who viewed it with no subtitles in terms of receptive and productive aspects of vocabulary knowledge. However, glossed keywords provided a significant improvement in the meaning recognition aspect only. EFL learners' vocabulary recognition and recall aspect can be positively affected if culturally appropriate subtitled films or animated movies are selected and implemented as in-class learning input.

References

Al-Masrai, Ahmed, and James Milton. 2012. “The Vocabulary Knowledge of University Students in Saudi Arabia.” *TESOL Arabia Perspectives* 19 (3): 13–19.

Canning-Wilson, Christine. 2000. “Practical Aspects of Using Video in the Foreign Language Classroom.” *The Internet TESL Journal* 6 (11). http://iteslj.org/Articles/Canning-Video.html.

Danan, Martine. 2004. “Captioning and Subtitling: Undervalued Language Learning Strategies.” *Meta* 49 (1): 67–77. https://doi.org/10.7202/009021ar.

Dang, Thi Ngoc Yen, Cailing Lu, and Stuart Webb. 2022. “Incidental Learning of Collocations in an Academic Lecture Through Different Input Modes.” *Language Learning* 72 (3): 728–64. https://doi.org/10.1111/lang.12499.

Fayyaz, Ali. 2013. “Most People Prefer Not to Read, Survey Shows | Arab News.” 2013. https://www.arabnews.com/news/471781.

Fievez, Isabeau, Maribel Montero Perez, Frederik Cornillie, and Piet Desmet. 2021. “Promoting Incidental Vocabulary Learning through Watching a French Netflix Series with Glossed Captions.” *Computer Assisted Language Learning* 36 (1–2): 26–51. https://doi.org/10.1080/09588221.2021.1899244.

Hsieh, Yufen. 2020. “Effects of Video Captioning on EFL Vocabulary Learning and Listening Comprehension.” *Computer Assisted Language Learning* 33 (5–6): 567–89. https://doi.org/10.1080/09588221.2019.1577898.

King, Jane. 2002. “Using DVD Feature Films in the EFL Classroom.” *Computer Assisted Language Learning* 15 (5): 509–23. https://doi.org/10.1076/call.15.5.509.13468.

Lee, Shinwoong. 2015. “Lexical Demand of Animation Movies and Vocabulary Learning Through Watching Them: A Corpus-Driven Approach.” *STEM Journal* 16 (1): 23–47. https://doi.org/10.16875/stem.2015.16.1.23.

Montero Perez, Maribel, Elke Peters, and Piet Desmet. 2017. “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.

Montero-Perez, Maribel, Elke Peters, Geraldine Clarebout, and Piet Desmet. 2014. “Effects of Captioning on Video Comprehension and Incidental Vocabulary Learning.” *Language, Learning and Technology* 18 (1): 118–41.

Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759.

Nation, Paul. 2007. “The Four Strands.” *Innovation in Language Learning and Teaching* 1 (1): 2–13. https://doi.org/10.2167/illt039.0.

Paivio, Allan. 1990. *Mental Representations: : A Dual Coding Approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001.

Peters, Elke, Eva Heynen, and Eva Puimège. 2016. “Learning Vocabulary through Audiovisual Input: The Differential Effect of L1 Subtitles and Captions.” *System* 63 (December): 134–48. https://doi.org/10.1016/j.system.2016.10.002.

Peters, Elke, and Stuart Webb. 2018. “INCIDENTAL VOCABULARY ACQUISITION THROUGH VIEWING L2 TELEVISION AND FACTORS THAT AFFECT LEARNING.” *Studies in Second Language Acquisition* 40 (3): 551–77. https://doi.org/10.1017/S0272263117000407.

Pujadas, Geòrgia, and Carmen Muñoz. 2019. “Extensive Viewing of Captioned and Subtitled TV Series: A Study of L2 Vocabulary Learning by Adolescents.” *The Language Learning Journal* 47 (4): 479–96. https://doi.org/10.1080/09571736.2019.1616806.

Rodgers, Michael P. H., and Stuart Webb. 2017. “The Effects of Captions on EFL Learners’ Comprehension of English-Language Television Programs.” *CALICO Journal* 34 (1): 20–38. https://doi.org/10.1558/cj.29522.

Talavan, Noa. 2010. “Subtitling as a Task and Subtitles as Support: Pedagogical Applications.” In *New Insights into Audiovisual Translation and Media Accessibility: Media for All 2*, edited by Jorge Díaz Cintas, Anna Matamala, and Josélia Neves, 285–89. New York, NY: Rodopi. https://doi.org/10.1163/9789042031814.

Van Zeeland, Hilde, and Norbert Schmitt. 2013. “Incidental Vocabulary Acquisition through L2 Listening: A Dimensions Approach.” *System* 41 (3): 609–24. https://doi.org/10.1016/j.system.2013.07.012.

Vanderplank, R. 1988. “The Value of Teletext Sub-Titles in Language Learning.” *ELT Journal* 42 (4): 272–81. https://doi.org/10.1093/elt/42.4.272.

Vanderplank, R. 1990. “Paying Attention to the Words: Practical and Theoretical Problems in Watching Television Programmes with Uni-Lingual (CEEFAX) Sub-Titles.” *System* 18 (2): 221–34. https://doi.org/10.1016/0346-251X(90)90056-B.

Watanabe, Yuichi. 1997. “INPUT, INTAKE, AND RETENTION: Effects of Increased Processing on IncidentalLearning of Foreign Language Vocabulary.” *Studies in Second Language Acquisition* 19 (3): 287–307. https://doi.org/10.1017/S027226319700301X.

Webb, Stuart. 2010a. “A Corpus Driven Study of the Potential for Vocabulary Learning through Watching Movies.” *International Journal of Corpus Linguistics* 15 (4): 497–519. https://doi.org/10.1075/ijcl.15.4.03web.

Webb, Stuart. 2010b. “Pre-Learning Low-Frequency Vocabulary in Second Language Television Programmes.” *Language Teaching Research* 14 (4): 501–15. https://doi.org/10.1177/1362168810375371.

Webb, Stuart. 2010c. “Using Glossaries to Increase the Lexical Coverage of Television Programs.” *Reading in a Foreign Language* 22 (01): 201–21.

Webb, Stuart. 2015. “Extensive Viewing: Language Learning through Watching Television.” In *Language Learning beyond the Classroom*, edited by D. Nunan and J. C. Richards, 159–68. New York: Routledge.

Webb, Stuart, Jonathan Newton, and Anna Chang. 2012. “Incidental Learning of Collocation: Incidental Learning of Collocation.” *Language Learning* 63 (1): 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x.

Webb, Stuart, and M. P. H. Rodgers. 2009b. “The Lexical Coverage of Movies.” *Applied Linguistics* 30 (3): 407–27. https://doi.org/10.1093/applin/amp010.

Webb, Stuart, and Michael P. H. Rodgers. 2009a. “Vocabulary Demands of Television Programs.” *Language Learning* 59 (2): 335–66. https://doi.org/10.1111/j.1467-9922.2009.00509.x.

Yoshii, Makoto. 2006. “L1 and L2 Glosses: Their Effects on Incidental Vocabulary Learning.” *Language Learning & Technology* 10 (3): 85–101.