International Journal of TESOL Studies (2026)
Vol. 8 (3) 204-229 https://doi.org/10.58304/ijts.260615

Article

The Effect of Textual Enhancement on Incidental Learning of Single Words and Multi-word Items From L2 Captioned Viewing

Maryam Ehsani
University of Tehran, Iran

Received: 16 January 2026 / Received in revised form: 29 May 2026 / Accepted: 14 June 2026 / Available online: 24 June 2026

Abstract
There has been limited research examining the impact of textual enhancement on incidental learning of multi-word items (MWIs) and single words from captioned viewing, particularly at the level of form recall. This study aims to address this gap. Sixty-two EFL learners encountered 16 MWIs and 18 single words in one of the following conditions: (a) MWI-enhanced captioned viewing (MWI-enhanced CV) (b) single-word enhanced captioned viewing (SW-enhanced CV), (c) uncaptioned viewing, (d) unenhanced captioned viewing (Unenhanced CV). Learning was gauged using a form-recall test with re-arranged items as a pretest, immediate posttest, and delayed posttest. Regarding MWI learning, the MWI-enhanced captioned viewing group significantly outperformed the other three groups on both immediate and delayed posttests. Regarding single-word learning, however, no significant differences were found among the four groups. Learners' prior vocabulary knowledge had a significantly positive correlation merely with the delayed gains in the single-word enhanced captioned viewing and the uncaptioned viewing conditions. Altogether, the findings demonstrate that instructional support embedded in captions through textual enhancement plays a critical role in incidental vocabulary learning. Theoretically, the results extend the Noticing Hypothesis by suggesting that perceptual salience can selectively support MWI learning without an apparent increase in cognitive load, as inferred from the vocabulary learning gains.

Keywords
Captioned viewing, incidental vocabulary learning, multi-word item, single word, textual enhancement

1 Introduction

Vocabulary is a critical element of language proficiency (Webb & Nation, 2017). For instance, Hu and Nation (2000) found that knowledge of 98% of the running words in a text is necessary for unassisted reading comprehension. Regarding single words, Nation (2006) analyzed different text types from both written and spoken English using the British National Corpus and demonstrated that 3,000-4,000 word families are necessary to achieve 95% text coverage and 6,000-9,000 word families are needed to

Email: Maryam.Ehsani1996@ut.ac.ir; howanagan@gmail.com

reach 98% text coverage (Nation, 2006). Multi-Word Items (MWIs, henceforth) include a vast array of expressions involving more than a single word, such as collocations and idioms (Boers, 2020). Previous corpus analyses have shown that MWIs, also called formulaic sequences or multiword expressions (e.g., Teng, 2025a), constitute a great percentage of both text and speech (e.g., Erman & Warren, 2000). Explicit teaching of all the single words and MWIs that one needs to know to communicate effectively in an L2 seems to be a challenging task. For that reason, both teachers and students appear to recognize the need to identify different methods to accelerate the vocabulary learning process (Alshumrani, 2024). Using different sources of input to encourage incidental vocabulary development is one of these methods.

Numerous sources of input have been studied with regard to their potential for incidental vocabulary learning. Empirical evidence from prior research suggests that vocabulary can be learned incidentally through reading (Ma & Reynolds, 2023; Pellicer-Sánchez et al., 2022; Zhou & Day, 2020), listening (Jin & Webb, 2020), reading-while-listening (Brown et al., 2008), viewing without captions (Peters & Webb, 2018), and viewing with captions (Teng, 2025b, 2026). Furthermore, Webb et al.'s (2023) meta-analysis demonstrated that incidental vocabulary learning occurs across different meaning-focused input modalities, including reading, listening, viewing, and reading-while-listening. However, lexical gains achieved through exposure to input-only appear to be relatively modest (Webb et al., 2023). This has led some researchers to develop a number of techniques to improve the word learning process. Two such techniques are providing captions for videos (see Kurokawa et al., 2025, for a meta-analytic review, and Wi & Boers, 2025, for a synthesis study) and textually enhancing target items in the input using underlining or similar techniques (Hsieh, 2020).

Although recent meta-analytic evidence has demonstrated that captioned viewing promotes incidental single-word learning (Kurokawa et al., 2025), the picture appears to be less than clear for MWI learning. To illustrate, while some studies revealed that captioned videos led to more MWI learning than uncaptioned videos (Majuddin et al., 2021; Teng, 2019b), there are studies that found no significant disparity between the two conditions (Dang et al., 2022a). Moreover, there seems to be a dearth of research on how textual enhancement in captions influences the acquisition of MWIs (Kurokawa et al., 2025). Typographic or textual enhancement (TE, henceforth) is a technique designed to help learners attend to specific linguistic items in the input, using methods such as underlining and/or bolding of those items (Lee, 2021). To the best of my knowledge, to date, only two studies have compared the impacts of typographically enhanced captions with normal captions on recalling the forms of MWIs (Majuddin et al., 2021; Puimège et al., 2023). These two studies differed in their research designs, which may partly explain their divergent findings (see Section 2.2 for details). The results of Majuddin et al.'s (2021) study showed that although both unenhanced captioned viewing and textually-enhanced captioned viewing positively influenced the learning of MWIs, no significant variation existed between them. On the other hand, Puimège et al. (2023) found that the presence of TE led to significantly higher MWI form-recall scores, when total reading times were not controlled for.

The present study aimed to extend the above-mentioned line of research by addressing the following research objectives. Firstly, not only the effects of captioning and TE on acquiring MWIs were explored, but also their effects on learning single words were investigated. Secondly, in contrast to the study conducted by Puimège et al. (2023) which used no delayed posttests, the current study included a delayed posttest. Delayed posttests are important because immediate posttest scores may reflect short-term recall of recently encountered lexical items rather than long-term retention. Thirdly, the relationship between students' prior vocabulary knowledge and learning gains from each treatment condition was also examined. Finally, a largely unexplored question concerns the potential trade-off effects of textual enhancement in captioned viewing. Specifically, does enhancing MWIs divert attention away from unenhanced single words, and conversely, does enhancing single words reduce the likelihood of noticing unenhanced MWIs? Given that attentional resources during viewing are limited (Mayer, 2021), and that the Noticing Hypothesis (Schmidt, 1990, 1994, 2001) posits that only noticed items can be acquired, it is theoretically possible that drawing learners' attention toward one category of lexical items through

TE may reduce the noticing and learning of unenhanced items of another category. This question is particularly significant given that previous research on incidental vocabulary learning through captioned viewing has examined TE effects on either single words or MWIs, without considering whether enhancing one type of item comes at a cost to learning the other. To my knowledge, no prior study has directly examined these potential trade-off effects, and the present study represents the first empirical investigation into this issue.

2 Review of the Literature

2.1 Vocabulary learning from captioned viewing

Incidental vocabulary learning takes place when lexical items are gained as a result of learner engagement in meaning-focused activities (Webb, 2020). Studies of incidental vocabulary development from uncaptioned viewing reveal that learning gains are rather small (Dang et al., 2022b; Peters & Webb, 2018; Puimège & Peters, 2019, 2020; Rodgers & Webb, 2020). For example, as for single words, participants in Peters and Webb's (2018) study learned approximately four words out of 64 target words following the viewing of a one-hour television program. As for MWIs, a study conducted by Puimège and Peters (2019) demonstrated that watching a 30-minute British reality TV program resulted in absolute gains of 2.20, 3.95, 2.4 out of 20 MWIs on measures of form recognition, form recall, and meaning recall, respectively. These limited gains are congruent with the "Noticing Hypothesis" (Schmidt, 1990, 1994, 2001), which postulates that forms must be noticed for intake to occur. During meaning-focused audiovisual viewing, learners' primary goal is typically to comprehend the content, which may limit the attentional resources available for processing unfamiliar lexical items. Moreover, according to Mayer's (2021) "Cognitive Theory of Multimedia Learning", the simultaneous processing of visual imagery and audio may leave little capacity for attending to novel words.

Sutton and Webb's (2026) recent meta-analysis of 56 experiments (N = 1,954) demonstrated that viewing uncaptioned audiovisual input had a medium, significant effect on single-word learning; however, its effect on learning MWIs was small and non-significant. They attributed the small effect of uncaptioned viewing on MWI learning to the relatively small sample sizes of the studies included in their meta-analysis that examined this issue.

Although incidental vocabulary learning from exposure to meaning-focused input has been suggested to be utilized (Nation, 2022; Webb, 2020), researchers have sought ways to accelerate the rate of incidental vocabulary development from different input modalities. Two of these methods that have been widely researched during the last decade include captioning videos and textually enhancing unknown vocabulary. Captions are the written text displayed on screen, providing a same-language transcription of the audio (Vanderplank, 2016a, 2016b). The majority of previous studies show that captioned viewing leads to a higher rate of single-word learning than uncaptioned viewing (Neuman & Koskinen, 1992; Peters, 2019; Teng, 2019a, 2022a; Teng & Cui, 2024, 2025). For example, Teng and Cui (2024), in a between-subjects study involving 129 young Chinese ESL learners, compared the incidental learning of single words under three captioning conditions: full captions, keyword captions, and no captions. Vocabulary learning was measured via meaning recall tests administered before, immediately after, and two weeks after the treatment. The results of this study showed that viewing with full captions was the most effective condition for single-word learning. However, there are studies that show a different pattern (Dang et al., 2023; Hao et al., 2022; Sydorenko, 2010; Wang, 2019). For example, Dang et al. (2023) compared the learning of 50 single words through uncaptioned and captioned viewing of an academic lecture. They found that while both conditions facilitated meaning recall, only uncaptioned viewing led to significant learning gains. This finding implies that, under certain conditions, captions may not always facilitate vocabulary development.

Furthermore, Wang (2019) examined the impacts of different caption types on English-major students' meaning-recall learning and found that in two out of the four classes that were studied, there were no significant disparities between the L2-English captioned and uncaptioned viewing modes. Additionally, in a study with intermediate and advanced EFL learners, Hao et al. (2022) found that L2-English captioned and uncaptioned TED talks yielded different vocabulary learning patterns depending on proficiency level. The outcomes of the meaning recognition test for intermediate learners indicated no statistically significant variation among the two modes whereas the findings for advanced learners showed a significant disparity among them; to be more specific, the no caption group significantly outperformed the English caption group. This result suggests that proficiency level possibly moderates how captions affect vocabulary acquisition. This finding aligns with Kurokawa et al.'s (2025) meta-analysis, according to which captioned videos were most beneficial for intermediate-level learners (g = 0.81) compared to beginning to low-intermediate learners (g = 0.28) and high-intermediate to advanced learners (g = 0.29). Altogether, the above-mentioned findings call for further research on the topic to clarify the overall pattern and discover the potential moderating factors.

Given the importance of MWIs for fluent and proficient language use (Durrant & Schmitt, 2010; Tavakoli & Uchihara, 2020), identifying effective ways to facilitate their acquisition can be considered important within the realm of L2 vocabulary research. This is particularly relevant because although MWIs constitute a substantial proportion of language (Erman & Warren, 2000), they are often difficult for L2 learners to acquire (Nguyen & Webb, 2017). Moreover, due to the prevalence of MWIs in language, explicit instruction alone is unlikely to provide sufficient coverage within the constraints of classroom time (Puimège & Peters, 2020). As audiovisual materials provide learners with readily accessible opportunities for incidental language exposure, recent years have witnessed several studies aimed at comparing the effects of captioned and uncaptioned viewing on incidental MWI learning at the level of form (Dang et al., 2022a; Majuddin et al., 2021; Teng, 2019b). In a between-subjects study with Chinese primary school ESL learners, Teng (2019b) reported that full captions resulted in highest incidental collocation learning gains, followed by keyword captions and no captions. In contrast, in a quasi-experimental investigation on EAP students' collocation form-recognition acquisition, Dang et al. (2022b) found that both captioned and uncaptioned viewing led to form-recognition development, with no significant dissimilarities between them. Rather unexpectedly, learners' pre-existing vocabulary knowledge did not influence the learning outcomes significantly. The inconsistency between Teng (2019b) and Dang et al. (2022b) may reflect methodological differences, including learner age (children vs adults), instructional context (general English at school level vs academic English at university), and video genre (storytelling video vs. academic lecture). Consequently, it remains unclear whether captioning benefits MWI learning generally or only under specific learning conditions and the inconsistent findings underscore the need for further research to elucidate the issue.

2.2 Textual enhancement in captioned viewing

Textual enhancement (TE) has been defined as manipulating a text to make certain linguistic elements more perceptually salient (Lee, 2021). One of the theories that supports the idea that textual enhancement of target items may boost their learning is "Noticing Hypothesis" (Schmidt 1990, 1994, 2001), according to which for language learning to occur, the individual needs to consciously notice the linguistic items. Regarding the effect of TE on incidental single-word learning, previous research has mainly focused on reading, with the overall results indicating the effectiveness of TE for single-word learning at the levels of form recognition (Sánchez Gutiérrez et al., 2019; Vu & Peters, 2022b) and form and meaning recall (Vu & Peters, 2022b). Yet extending TE to audiovisual input introduces additional cognitive demands: learners must coordinate the processing of images, audio, captions, and enhancements simultaneously. According to the "Cognitive Theory of Multimedia Learning" (Mayer, 2021), because of human's limited capacity of processing channels, TE may hinder learning by adding extraneous cognitive load.

Only few studies have explored the effect of textually enhanced captions on single-word learning (Cintrón-Valentín & García-Amaya, 2021; Hsieh, 2020; Montero Perez et al., 2014). Cintrón-Valentín and García-Amaya (2021) recruited 369 L2 Spanish learners to investigate the impact of viewing textually enhanced captioned videos on L2 vocabulary and grammar development. Three captioning conditions were examined: captioned video with TE on target vocabulary, captioned video with TE on target grammar, and no captioning. Findings of immediate posttests revealed that captioning was effective for vocabulary learning; however, the analyses of the delayed posttest showed that long-term effects were not as stable for low-frequency items.

Hsieh (2020) compared the effects of the following viewing modes on low-intermediate EFL learners' incidental vocabulary learning (form and meaning recognition, and meaning recall): 1. uncaptioned, 2. fully-captioned, 3. fully-captioned with textually-enhanced target-words. Results of form-recognition posttest scores revealed no significant difference between modes 2 and 3. Fully-captioned viewing led to significantly greater meaning recognition and recall scores than the other two modes, but no significant differences were detected between modes 1 and 3. Fully-captioned viewing with enhanced target words caused form-recognition development at the expense of meaning recognition and recall. The uncaptioned viewing was not effective for vocabulary acquisition.

Montero Perez et al. (2014) studied the impacts of uncaptioned viewing (control), fully- captioned viewing, and textually-enhanced fully-captioned viewing on incidental learning of single words and MWIs. The results demonstrated that all captioned viewing conditions led to equally well scores on the form recognition test and that they resulted in significantly greater scores compared to the control condition. The textually-enhanced fully-captioned viewing condition resulted in significantly higher scores than the control condition on meaning recognition. However, captioning did not affect meaning recall. Another notable finding was that vocabulary size significantly correlated with the learning gains. In a similar vein, Montero Perez et al. (2018) observed that the captioned viewing group significantly outperformed the uncaptioned viewing group in terms of form recognition of 18 target items, including 6 MWIs and 12 single words.

Similar to single-word learning, the effect of TE on MWI learning has been examined mostly in the area of reading (Boers et al., 2017; Choi, 2017; Jung et al., 2025; Puimege et al., 2024; Toomer & Elgort, 2019; Vu & Peters, 2022a) and reading-while-listening (Jung & Lee, 2024; Vu & Peters, 2023). The overall results of the research undertaken in this area show that TE can possibly increase attention to target MWIs and improve their learning. Recognizing the scarcity of research on the efficacy of captioned viewing plus TE on incidental learning of MWIs, Majuddin et al. (2021) implemented a study involving Malaysian L2 English learners. Intact classes were randomly allocated to different caption conditions: no captions, unenhanced captions, and enhanced captions (via bolding and underlining). The results of the immediate form-recall posttest showed that both caption types yielded higher form recall scores than the no captions condition. However, no significant variation was detected between the scores of the two caption conditions. Another noteworthy outcome was that the participants' VST score significantly predicted form-recall scores, indicating the presence of a "Matthew effect" (e.g., Stanovich, 2009).

Adopting a counterbalanced within-subjects design, Puimège et al. (2023) investigated the impact of TE on 26 EFL learners' attention to and learning of MWIs from viewing a 30-minute BBC documentary. Two versions of the same video were formed, each divided into an unenhanced and an enhanced section. Items enhanced in the first version were unenhanced in the second, and vice versa. Results revealed that TE had a significantly positive influence on MWI form-recall scores only when reading times were excluded from the analysis. As Puimège et al. (2023) highlighted, this implies that TE can facilitate the incidental acquisition of MWIs under certain conditions. According to the findings, they argued that TE may benefit learning if only learners read captions fluently and allocate their attention efficiently.

Since EFL learners tend to receive larger amounts of L2 input through viewing television programs and films than through reading (Lindgren & Muñoz, 2013; Peters, 2018), it seems necessary to

investigate the effectiveness of this mode of input (either with or without captions and TE) for learning single words and MWIs. This was the primary focus of the present investigation. Moreover, because empirical findings on the relationship between prior vocabulary knowledge and both single-word and MWI learning from viewing are mixed, the secondary objective of this research was to address this issue.

2.3 Research questions

The present investigation addresses the following research questions:

1. To what degree do learners acquire and retain form-recall knowledge of the target multi-word items (MWIs) and single words after each of the four treatment conditions (MWI-enhanced CV, SW-enhanced CV, uncaptioned viewing, and unenhanced CV)?
2. How do the four treatment conditions differ in their effects on learners' form-recall gains for (a) target MWIs and (b) target single words on the immediate and delayed posttests?
3. Is there any relationship between prior vocabulary knowledge and form-recall gains under each treatment condition?

3 Method

3.1 Research design

This study tried to explore the effect of TE of single words and MWIs in captioned videos on their immediate learning and one-week retention at the level of form recall. The learning conditions included the following, randomly distributed among four intact classes: 1. captioned viewing with textually enhanced target MWIs (MWI-enhanced CV), 2. captioned viewing with textually enhanced target single words (SW-enhanced CV), 3. uncaptioned viewing, 4. captioned viewing with no TE (Unenhanced CV). The screenshots of the four viewing conditions are displayed in Figure 1.

Figure 1

Screenshots of the Four Viewing Conditions

Condition 1: MWI-enhanced CV

Condition 2: SW-enhanced CV

Condition 3: Uncaptioned viewing

Condition 4: Unenhanced CV

3.2 Participants

Initially, 79 Iranian English-major university students from four intact classes participated in the experiment, but data of some students were removed from the main analysis, resulting in 62 participants. The removal included those who were absent in one of the testing sessions or those who failed to meet the requirement of knowing the most frequent 3,000 English word families according to the results of Webb et al.'s (2017) Updated Vocabulary Levels Test (UVLT). This criterion was adopted because knowledge of the 3,000 most frequent word families was necessary to achieve adequate viewing comprehension of the documentary (Durbahn et al., 2020; Webb and Rodgers, 2009). Participants' age ranged from 18 to 23 (mean age = 20.46, SD = 1.30) and the sample comprised 18 male and 44 female students. Vocabulary knowledge was also measured employing Nation and Beglar's (2007) Vocabulary Size Test (VST). The participants' mean VST score was 76.11 out of 140 (SD = 14.37), and their score range indicated they knew between 4,000-11,000 most frequent English word families receptively. Normality was assessed via skewness and kurtosis values, and Levene's test confirmed homogeneity of variance across the four groups. ANOVA results revealed no statistically significant differences in UVLT and VST scores among the four classes ($p > .05$ for both UVLT and VST). The two vocabulary tests administered in this study served different purposes. The UVLT was used to assess participants' vocabulary mastery across frequency bands and to determine whether their vocabulary knowledge was sufficient for meeting the lexical demands of the documentary. In contrast, the VST was used to estimate participants' overall receptive vocabulary size. It was included to address RQ3, which examined the relationship between learners' prior vocabulary knowledge and their incidental vocabulary learning gains. Ten additional English-major university students who met the vocabulary requirement of the study were recruited in the pilot study.

3.3 Material

Following a pilot study and considering input from the students' English teachers, the first 32 minutes of the first episode of The World's Sneakiest Animals, a BBC documentary series, was used as the video

material. The pilot group watched the documentary with captions and reported that the content was interesting to them. The episode titled 'Staying alive' was about visual trickery. The duration of the video was 32 minutes and 15 seconds and its script consisted of 2,923 words. Analyzing the script employing RANGE (Nation & Heatley, 2002) showed that 91.5% of the words belonged to the 3,000 most frequent English word families. It should be noted that recent research on audiovisual comprehension suggests that lexical coverage levels below 95% may still allow adequate understanding of audiovisual materials (Durbahn et al., 2020, 2024). Following Durbahn et al. (2020), this project deemed 90% coverage as appropriate. According to the participants' scores on the UVLT, VST, and comprehension test, as well as the results of the pilot study, the participants were believed to be able to follow the video with relative ease.

3.4 Target items

The video was carefully reviewed to identify MWIs and single words presumed to be unfamiliar to the participants. After generating a list of possible target MWIs and single words, I sought input from three of the learners' English teachers concerning the learners' probable knowledge of the chosen items. Fifty-four items, including 25 MWIs and 29 single words, were piloted with 10 students. A total of 16 MWIs and 18 single words which were shown to be unknown by 80% or more of the pilot study participants on the form-recall test were chosen as the final items. The 80% cut-off point follows the convention adopted in prior incidental vocabulary learning studies (Dang et al., 2022b; Peters & Webb, 2018; Puimège et al., 2023). Target MWIs included 10 collocations, 3 phrasal verbs, and 3 idioms. The single words consisted of 6 nouns, 5 verbs, 6 adjectives, and 1 adverb. All but three single words appeared only once in the documentary. The MWIs had a minimum MI score (a widely accepted cut-off point for association strength) of 3.0 (Dang et al., 2022b; Majuddin et al., 2021; Puimège et al., 2023) and yielded at least 100 hits in COCA (Davies, 2008) and all the single words fell beyond the most frequent 3,000 English words. For ecological validity, neither videos nor target items were manipulated. See Appendix A for target MWIs and Appendix B for target single words.

3.5 Instruments

A paper-and-pencil test was employed to assess the participants' recall of the target MWIs and single-word forms. The first section of this test measured knowledge of MWIs and the second section knowledge of single words. Participants were required to write the English form of each target item, prompted by its corresponding Persian definition (or English description, for only one item). An exception was the word "mush", for which an English description was provided instead of a Persian definition because no suitable single-word Persian equivalent was available. Following prior research (e.g., Puimège & Peters, 2020), the initial letter and the number of letters of each word was provided to prevent participants from writing other potential words, as can be seen in the example below:

محافظت کردن = To c - - - - - p - - - - - - - - - (To confer protection)

The same test was administered as a pretest, an immediate posttest, and a delayed posttest, with a one-week interval between each administration. To reduce testing effects, the order of items was randomized in each testing phase. Aside from the target items, in alignment with prior studies (Dang et al., 2022b; Webb & Chang, 2022), several single words (e.g., prey) and MWIs (e.g. stay alive) known by 80% or more of pilot study participants were incorporated in the test to maintain participant motivation and to minimize deliberate memorization of the target items. However, responses to these distractor items were excluded from the primary analysis. Altogether, the test contained 16 target MWIs, 18 target single words, and 20 distractors (including 9 MWIs and 11 single words). Right after the treatment, participants completed a 10-item true/false test assessing their comprehension of the BBC documentary. This test did not include any of the target items.

3.6 Procedure

Data were collected over a four-week period. In the first week, participants were informed about the general research (without revealing its specific aims) before providing their consent to participate. In the second week, participants completed the UVLT, the VST, and the pretest. One week later, they viewed the video according to their assigned experimental conditions, using the multimedia system (computer, projector, and high-quality speakers) of their classrooms. Prior to exposure to the learning materials, they were instructed that they would subsequently answer comprehension questions. Immediately after the treatment, they completed the comprehension test, followed by the immediate posttest. In the final week, participants took the unannounced delayed posttest. They were directed to focus their attention on understanding the video's content and to refrain from using mobile phones, consulting dictionaries, or taking notes during the experiment. The researcher, who taught all four classes, monitored compliance with the experimental instructions. After the delayed posttest, participants were debriefed regarding the study's real purposes.

3.7 Scoring

Responses were scored using a binary method (1 = correct, 0 = incorrect). A lenient scoring system was employed, in which minor errors received a full mark (e.g., using a plural instead of a singular form or vice versa; wrongly spelled words that did not result in a different word or change in meaning). All pretests and posttests were initially scored by the researcher, followed by a random secondary scoring of 50% of the tests by a Persian-speaking TEFL PhD holder. Inter-rater reliability analysis, as measured by Cohen's kappa, revealed almost perfect agreement (.98, .96, and .98 for the pretest, immediate posttest, and delayed posttest, respectively). Any discrepancies were resolved through discussion.

3.8 Data analysis

Data analysis was carried out in SPSS (version 26.0). Normality was confirmed by checking the skewness and kurtosis values. Since the data were normally distributed, mixed within-between-subjects ANOVAs were used to answer RQ1. Several ANCOVAs with learners' pretest scores as a covariate were performed to answer RQ2. In line with Norouzian and Plonsky's (2018) recommendations, effect sizes as well as the corresponding confidence intervals are reported. Following Richardson's (2011) suggestion for interpreting partial eta squared, partial eta squared values of .01, .06, and .14 are considered small, medium, and large effects, respectively. To answer RQ3, two Pearson product-moment correlations were performed. According to Rodgers and Webb (2020), absolute gains may not accurately represent vocabulary learning because participants with higher pretest scores have less room for improvement than those with lower scores. Therefore, following previous studies (Ahrabi Fakhr et al., 2021; Fievez et al., 2020; Peters & Webb, 2018; Rodgers & Webb, 2020), relative gains were used to perform the correlations. To compute relative gains, the following formula was applied:

[absolute gains/ (number of target items - number of known words)] × 100.

4 Results

Participants' comprehension test scores showed that, irrespective of their viewing conditions, all four groups comprehended the content of the video well. The average comprehension score was 7.98 (standard deviation= 1.32).

RQ1: To what degree do learners acquire and retain form-recall knowledge of the target multi-word items (MWIs) and single words after each of the four treatment conditions (MWI-enhanced CV, SW-enhanced CV, uncaptioned viewing, and unenhanced CV)?

A one-way ANOVA was conducted to compare the four groups' pretest scores. Unenhanced CV had the best performance on the pretest, followed by Uncaptioned viewing, MWI-enhanced CV, and SW-enhanced CV, respectively (Unenhanced CV > Uncaptioned viewing > MWI-enhanced CV > SW-enhanced CV). Levene's test for equality of variances indicated homogeneity among the four groups' pretest scores ($p > .05$). The ANOVA result showed that the pretest scores did not differ from each other significantly, F (3, 58) = 1.831, $p = .152$, $p > .05$, partial $\eta^2 = 0.086$, [90% CI: 0.00, 0.17] (medium effect).

Two other one-way ANOVAs were executed to compare the four groups in terms of their score on the MWI section (including 16 target items) and the single words section (including 18 target items) of the pretest. Regarding the MWI section, the ANOVA revealed no statistically significant disparities among the four groups' scores, F (3, 58) = 2.067, $p = .115$, $p > .05$. Similarly, with respect to the single words section, no significant difference was detected, F (3, 58) = 1.166, $p = .331$, $p > .05$.

To answer RQ1, a Repeated Measures ANOVA was run. The design incorporated Time (pretest, immediate posttest, delayed posttest) as a within-subjects factor and Group (treatment condition) as a between-subjects factor. Mauchly's test confirmed the sphericity assumption, x2(2) = .906, $p = .61$. However, results from Box's M test suggested that the homogeneity of covariance matrices assumption was not met ($p = .000$, $p < .05$). Therefore, rather than Wilks' lambda, Pillai's trace was utilized as the F-statistic to interpret main effects (Cramer & Howitt, 2004, as cited in Teng, 2022a). Multivariate test results demonstrated a significant Time × Group interaction, F (3, 58) = 3.35, $p = .004$, $p < .05$, partial $\eta^2 = 0.148$ [90% CI: 0.01, 0.25] (large effect).

Twelve paired-samples t-tests were run to further analyze the interaction effect (with a Bonferroni-adjusted alpha of .004). The analysis revealed that Groups 1 and 2 showed significant improvement from pretest to immediate posttest ($p = .000$, $p = .001$, respectively) while Groups 3 and 4 exhibited no significant gains ($p = .025$, $p = .01$, respectively). No significant variations were observed between the immediate and delayed posttests, nor between the pretests and delayed posttests, in any of the four groups.

After that, two separate mixed within-between-subjects ANOVAs were performed on MWI (Table 1) and single words (Table 2) sections of the tests. Regarding the MWI section, because the Box's M was statistically significant ($p = .000$, $p < .05$), the Pillai's Trace in the Multivariate Test was consulted. Because the statistic was significant for Time × Group interaction ($p = .000$, $p < .05$), follow-up paired-samples t-tests were conducted. Again, the alpha was adjusted to .004 (.05 ÷ 12). The results showed that a significant difference occurred only between the pretest and immediate posttest scores in Group 1 ($p = .000$).

Another mixed within-between-subjects ANOVA was run, this time on the single-word section scores of the four groups across the three testing points. Levene's test showed that the variances were homogeneous for the pretest, immediate posttest, and delayed posttest scores on the single-word section. Following a significant Mauchly's test result, $\chi^2(2) = 0.793$, $p = .001$, $p < .05$, the Greenhouse-Geisser correction ($\varepsilon = 0.829$) was employed to address the violation of sphericity. Given that the Box's M was statistically significant ($p = .000$, $p < .05$), the Pillai's Trace in the Multivariate Tests was consulted. The Pillai's Trace showed no significant Time × Group Interaction, F (3, 58) = 1.60, $p = .151$, $p > .05$, partial $\eta^2 = 0.077$ [90% CI: 0.00, 0.16] (medium effect). The Pillai's Trace row in the multivariate tests revealed a significant main effect of "Time", Pillai's Trace = 0.319, F (2, 57) = 13.326, $p = .000$, $p < .05$, partial $\eta^2 = 0.319$[90% CI: 0.14, 0.44] (large effect). The large significant effect size of time revealed that all groups performed the best on the immediate posttest, followed by the delayed posttest. Compared with the other two testing points, all groups achieved the lowest score on the single words section of the pretest. According to the Tests of Between Subjects Effects, the independent variable "Group" was not significant, F (3, 58) = 0.163, $p = .921$, $p > .05$, partial $\eta^2 = 0.008$, indicating that the four groups did not differ significantly in their knowledge of single-word items.

Table 1

Descriptive Statistics for the Performance of the Four Groups on the MWI Section of the Test across the Three Testing Points

	Group	Mean	Std. Deviation	N
PreMWI	Group 1	0.437	0.963	16
	Group 2	0.250	0.447	16
	Group 3	1.071	1.774	14
	Group 4	1.000	0.966	16
	Total	0.677	1.141	62
ImMWI	Group 1	4.500	2.943	16
	Group 2	0.750	0.856	16
	Group 3	2.000	2.417	14
	Group 4	1.562	1.209	16
	Total	2.209	2.443	62
DelMWI	Group 1	2.312	2.891	16
	Group 2	0.437	0.512	16
	Group 3	1.714	2.367	14
	Group 4	1.250	1.238	16
	Total	1.4194	2.044	62

Note. Max. = 16. PreMWI = pretest score on the MWI section; ImMWI = immediate posttest score on the MWI section; DelMWI = delayed posttest score on the MWI section.

Table 2

Descriptive Statistics for the Performance of the Four Groups on the Single-word Section of the Test across the Three Testing Points

	Group	Mean	Std. Deviation	N
PreSingle	Group 1	0.937	1.181	16
	Group 2	0.500	0.730	16
	Group 3	0.571	0.851	14
	Group 4	1.187	1.682	16
	Total	0.806	1.185	62
ImSingle	Group 1	1.375	1.543	16
	Group 2	1.750	1.527	16
	Group 3	1.500	1.605	14
	Group 4	1.437	1.750	16
	Total	1.516	1.575	62
DelSingle	Group 1	1.375	1.627	16
	Group 2	0.875	1.310	16
	Group 3	1.000	1.109	14
	Group 4	1.250	1.653	16
	Total	1.129	1.431	62

Note. Max. = 18. PreSingle = pretest score on the single-word section; ImSingle = immediate posttest score on the single-word section; DelSingle = delayed posttest score on the single-word section.

RQ2: How do the four treatment conditions differ in their effects on learners' form-recall gains for (a) target MWIs and (b) target single words on the immediate and delayed posttests?

To answer RQ2, two one-way ANCOVAs were conducted separately, analyzing the immediate and delayed posttest scores, respectively, with Group as the between-subjects factor. Pretest scores and total UVLT scores were initially included in the model as covariates, but the UVLT scores were later removed due to their significant correlation with pretest scores, Pearson's $r = 0.293$, p (sig. 2-tailed) = .021, $p < .05$.

Before running the ANCOVA, the assumption of homogeneity of regression slopes was tested by examining the Group × Pretest interaction for each dependent variable. Regarding the immediate posttest scores, the Group × Pretest interaction was non- significant, $p = .5$, $p > .05$. Similarly, for the delayed posttest scores, the Group × Pretest interaction did not reach significance, $p = .136$, $p > .05$. Therefore, as can be seen, the homogeneity of regression slopes assumption was satisfied for each dependent variable. The ANCOVA results indicated that after controlling for the pretest scores, there was a significant disparity between the four treatment groups on immediate posttest scores, F (3, 57) = 7.585, $p = .000$, $p < .05$, partial $\eta^2 = 0.285$ [90% CI: 0.10, 0.40] (large effect).

A Bonferroni post-hoc analysis was run on the immediate posttest scores to detect the differences between the four groups. As shown in Table 3, Group 1 (MWI-enhanced CV) significantly outscored the other three groups ($p < .05$ in all cases). The variations between other groups did not reach statistical significance.

Table 3

Bonferroni Post-hoc Analysis on the Immediate Posttest Scores

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Group 1	Group 2	2.608*	.840	.018	.314	4.903
	Group 3	2.704*	.863	.016	.345	5.062
	Group 4	3.871*	.844	.000	1.563	6.180
Group 2	Group 1	-2.608*	.840	.018	-4.903	-.314
	Group 3	.095	.876	1.000	-2.298	2.489
	Group 4	1.263	.870	.911	-1.114	3.640
Group 3	Group 1	-2.704*	.863	.016	-5.062	-.345
	Group 2	-.095	.876	1.000	-2.489	2.298
	Group 4	1.168	.867	1.000	-1.201	3.537
Group 4	Group 1	-3.871*	.844	.000	-6.180	-1.563
	Group 2	-1.263	.870	.911	-3.640	1.114
	Group 3	-1.168	.867	1.000	-3.537	1.201

Note. *$p < .05$

Additionally, the ANCOVA revealed that after controlling for the pretest scores, a significant variation existed between the four treatment groups on the delayed posttest scores, F (3, 57) = 5.075, p = .003, p < .05, partial η^2 = 0.211[90% CI: 0.04, 0.32] (large effect). A Bonferroni post-hoc analysis was performed to identify the specific between-group differences. As can be seen in Table 4, it was discovered that the only significant difference was between the scores of Group 1 (MWI-enhanced CV) and Group 4 (unenhanced CV), with Group 1 significantly outperforming Group 4, p = .002, p < .05.

Table 4

Bonferroni Post-hoc Analysis on the Delayed Posttest Scores

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Group 1	Group 2	1.515	.602	.088	-.130	3.160
	Group 3	1.342	.618	.205	-.349	3.032
	Group 4	2.305*	.605	.002	.651	3.960
Group 2	Group 1	-1.515	.602	.088	-3.160	.130
	Group 3	-.173	.628	1.000	-1.889	1.542
	Group 4	.790	.623	1.000	-.913	2.494
Group 3	Group 1	-1.342	.618	.205	-3.032	.349
	Group 2	.173	.628	1.000	-1.542	1.889
	Group 4	.964	.621	.758	-.735	2.662
Group 4	Group 1	-2.305*	.605	.002	-3.960	-.651
	Group 2	-.790	.623	1.000	-2.494	.913
	Group 3	-.964	.621	.758	-2.662	.735

Note. *p < .05

After checking for ANCOVA assumptions, including the assumption of homogeneity of regression slopes, and ensuring that they were met, four other ANCOVAs were run, two on the MWI-section and two on the single-word section of the immediate and delayed posttests. The ANCOVA showed that after controlling for the MWI pretest scores, there was a significant disparity between the MWI immediate posttest scores of the four groups, F (3, 57) = 15.41, p = .000, p < .05, partial η^2 = 0.448 [90% CI: 0.26, 0.54] (large effect). A Bonferroni post-hoc analysis was run on the MWI immediate posttest scores to discover the differences between the four groups. The results (Table 5) revealed that Group 1 (MWI-enhanced CV) significantly outscored the other three groups (p < .001 in all cases). The variations between other groups did not reach statistical significance.

The ANCOVA results showed that after controlling for the MWI pretest scores, there was a significant disparity between the MWI delayed posttest scores of the four groups, F (3, 57) = 6.616, p = .001, p < .05, partial η^2 = 0.258 [90% CI: 0.08, 0.37] (large effect). A subsequent Bonferroni post-hoc analysis was run to discover which groups differed significantly. The results (Table 6) revealed that Group 1 (MWI-enhanced CV) significantly outscored the other three groups (p < .05 in all cases). The variations between other groups did not reach statistical significance.

Table 5

Bonferroni Post-hoc Analysis on the MWI Section of the Immediate Posttest

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Group 1	Group 2	3.566*	.611	.000	1.895	5.237
	Group 3	3.122*	.645	.000	1.360	4.884
	Group 4	3.489*	.621	.000	1.792	5.186
Group 2	Group 1	-3.566*	.611	.000	-5.237	-1.895
	Group 3	-.444	.653	1.000	-2.230	1.342
	Group 4	-.077	.629	1.000	-1.796	1.643
Group 3	Group 1	-3.122*	.645	.000	-4.884	-1.360
	Group 2	.444	.653	1.000	-1.342	2.230
	Group 4	.367	.632	1.000	-1.359	2.094
Group 4	Group 1	-3.489*	.621	.000	-5.186	-1.792
	Group 2	.077	.629	1.000	-1.643	1.796
	Group 3	-.367	.632	1.000	-2.094	1.359

Note. *$p < .05$

Table 6

Bonferroni Post-hoc Analysis on the MWI Section of the Delayed Posttest

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Group 1	Group 2	1.621*	.453	.004	.382	2.860
	Group 3	1.457*	.478	.021	.150	2.764
	Group 4	1.824*	.460	.001	.566	3.083
Group 2	Group 1	-1.621*	.453	.004	-2.860	-.382
	Group 3	-.164	.485	1.000	-1.489	1.160
	Group 4	.203	.466	1.000	-1.072	1.478
Group 3	Group 1	-1.457*	.478	.021	-2.764	-.150
	Group 2	.164	.485	1.000	-1.160	1.489
	Group 4	.368	.468	1.000	-.913	1.648
Group 4	Group 1	-1.824*	.460	.001	-3.083	-.566
	Group 2	-.203	.466	1.000	-1.478	1.072
	Group 3	-.368	.468	1.000	-1.648	.913

Note. *$p < .05$

The ANCOVA results revealed that after controlling for the single-word pretest scores, there was no significant variation between the single-word immediate posttest scores of the four groups, F (3, 57) =

2.275, $p = .09$, $p > .05$. Similarly, it was revealed that after controlling for the single-word pretest scores, the four groups did not differ significantly in their single-word delayed posttest scores, $F(3, 57) = 1.009$, $p = .396$, $p > .05$.

RQ3: Is there any relationship between prior vocabulary knowledge and form-recall gains under each treatment condition?

Two Pearson correlations were calculated, one between the VST scores and immediate relative gains and the other one between VST scores and delayed relative gains. The combined results of all the four groups indicated that VST scores were not significantly correlated with either immediate or delayed relative gains ($p > .05$ in both cases). Four correlations were run between VST scores and immediate relative gains per group and four were performed between VST scores and delayed relative gains per group. Analysis of correlations per group showed that in Group 2 ($r = .576$, $p = .019$, $p < .05$) and Group 3 ($r = 0.746$, $p = .002$, $p < .05$) there were significant positive correlations between VST scores and delayed relative gains. These results suggest that the more vocabulary the participants in Group 2 (SW-enhanced CV) and Group 3 (uncaptioned viewing) knew, the greater their delayed vocabulary learning gains were.

5 Discussion

This study examined the efficacy of four different captioning conditions in bringing about initial and medium-term vocabulary learning from viewing a 30-minute documentary. It also explored the relationship between pre-existing vocabulary knowledge and vocabulary learning from each condition. First and foremost, the higher form-recall posttest scores compared to pretest scores corroborates previous research (Puimège et al., 2023; Puimège & Peters, 2019, 2020) and sheds more light on the efficacy of viewing audio-visual materials for vocabulary form-recall development. Separate analysis of the participants' performance on MWI- (including 16 items) and single-word sections (including 18 items) of the tests depicted the same pattern. These findings lend support to previous studies which reported that viewing uncaptioned and captioned videos is beneficial for incidental learning of MWIs (Dang et al., 2022; Majuddin et al., 2021; Puimège & Peters, 2019, 2020; Vu et al., 2023) and single words (Ahrabi Fakhr et al., 2021; Montero Perez, 2020; Puimège & Peters, 2019; Pujadas & Muñoz, 2019; Teng, 2023a; Wang & Pellicer-Sánchez, 2022) at the level of form.

In answer to RQ1, the results revealed that in Group 1 (MWI-enhanced CV) and Group 2 (SW-enhanced CV), the immediate posttest scores were significantly greater than the pretest scores, but no significant differences were found between the pretest and immediate posttest scores in Group 3 (uncaptioned viewing) and Group 4 (unenhanced CV). Although failing to reach statistical significance, the pretest-immediate posttest improvement of the uncaptioned viewing group is tentatively in line with Puimège and Peters (2019), in which the outcomes of the form-recall posttest showed that both MWIs and single words were acquired even from one exposure to that item through uncaptioned viewing. The modest, non-significant improvement in the uncaptioned viewing group is also consistent with broader meta-analytic evidence indicating that uncaptioned viewing can promote incidental vocabulary learning, although the learning gains are often limited. Sutton and Webb (2026) reported a medium overall effect of uncaptioned viewing on L2 learning ($g = 1.01$) across 56 experiments, thereby confirming that uncaptioned viewing possesses genuine learning potential. Similarly, Webb et al. (2023), in their meta-analysis of incidental vocabulary learning from meaning-focused input, found that vocabulary learning occurred across all input modes (i.e., reading, listening, reading-while-listening, and viewing). Nevertheless, viewing produced the smallest proportional gains, with approximately 7% of target words learned on immediate posttests and 5% on delayed posttests. Webb et al. (2023) suggested that this pattern may be attributable to learners' greater familiarity with written and spoken input in instructional settings as well as the frequent use of authentic, L1-oriented audiovisual materials, which

are often linguistically more demanding than materials specifically designed for L2 learners. This may partly explain why the gains observed in the uncaptioned viewing condition did not reach statistical significance. The video material in the present study was an authentic documentary and it was not specifically designed for L2 learners.

Moreover, the increase in the immediate posttest scores of the unenhanced CV group provides tentative evidence in support of Montero Perez et al. (2014) and Montero Perez et al. (2018), which found that captioned viewing facilitates the incidental acquisition of MWIs and single words at the form-recognition level. However, because the pretest-immediate posttest variations failed to reach statistical significance in the current study, it needs to be reiterated that these interpretations are only tentative and further evidence is required to confirm them. The non-significant improvement observed in the unenhanced CV group can be interpreted in light of several moderating variables identified in previous research. To exemplify, Kurokawa et al. (2025) found a medium overall effect of captioning on incidental vocabulary learning (g = 0.56), but reported that the effect was substantially reduced in studies employing pretest-posttest designs (g = 0.29 –0.35) compared with studies without a pretest (g = 1.01). Since the present study adopted a pretest-posttest design, this methodological feature may have diminished the magnitude of vocabulary gains in the unenhanced CV group. In addition, Kurokawa et al. (2025) found that captioning effects were considerably larger when videos were designed for L2 learners (g = 1.04) than when authentic L1-targeted videos were used (g = 0.46). Consequently, the L1-targeted documentary used in the current study may have contributed to the relatively modest gains observed in the unenhanced CV condition.

Separate analysis of the MWI-section of the tests demonstrated that the immediate posttest scores were significantly higher than the pretest scores only in Group 1 (MWI-enhanced CV), indicating that TE of MWIs helped learners not only notice those items but also gain initial knowledge of their form. This outcome is consistent with Puimège et al.'s (2023) study, in which the target MWIs were more likely to be known in the immediate posttest than in the pretest. The disparities between other scores did not reach statistical significance. The non-significant improvement of the uncaptioned viewing group contradicts Puimège and Peters (2020) and Dang et al. (2022a), who found significant pretest-posttest improvement in students' form-recall and form-recognition scores, respectively. As mentioned earlier, according to Kurokawa et al. (2025) and Webb et al. (2023), pretest design and the use of L1-targeted videos both reduce incidental vocabulary gains from viewing (either with or without captions), which may account for the more modest outcomes in the present study relative to those reported by Puimège and Peters (2020) and Dang et al. (2022a).

Separate analysis of the single-word section of the tests showed a large significant effect size for time, revealing that altogether, the groups performed the best on the immediate posttest, followed by the delayed posttest, and they achieved the lowest score on the pretest. The lower scores of the participants in the delayed posttest compared to the immediate posttest align with previous findings (Cintrón-Valentín & García-Amaya, 2021; Teng, 2023a, 2024b). For example, Cintrón-Valentín and García-Amaya (2021) found that in the full captions + vocabulary TE condition, a considerable reduction occurred in learners' form-recall of the target single words between the immediate and the delayed posttest. As argued by the researchers, it is possible that learners' lower scores at the delayed posttest was because of not having an opportunity to revisit the target vocabulary (Cintrón-Valentín & García-Amaya, 2021), as was the case in the current investigation.

In-depth analysis of the outcome of the single-word section of the uncaptioned viewing group reveals that the absolute gain from the pretest to the immediate posttest was almost one word (1.5 – 0.57 = 0.93), which represents about 5.2 % of the target single words (n = 18). This finding is congruent with the outcomes of the meta-analysis undertaken by Webb et al. (2023), which showed that participants acquired between 5% and 7% of the target vocabulary from uncaptioned viewing, based on immediate posttest scores.

The non-significant single-word learning in the unenhanced CV group contradicts most of the previous studies investigating development of form-recognition (Ahrabi Fakhr et al., 2021; Teng, 2022b, 2023b, 2024; Wang & Pellicer-Sánchez, 2022) and form-recall (Pujadas & Muñoz, 2019) through captioned viewing. For example, this outcome differs from that of Teng (2023b), who reported that for form recognition through video exposure, viewers were 1.35 times more likely than non-viewers to score one level higher on the immediate posttest. Furthermore, this finding is in contrast with Pujadas and Muñoz (2019), who found that captioned viewing brought about a significant difference between form-recall pretest and posttest scores.

In answer to RQ2, a post-hoc analysis on the immediate posttest scores revealed that Group 1 (MWI-enhanced CV) significantly outperformed the other three groups. This finding suggests that TE of MWIs is particularly effective for vocabulary improvement through captioned viewing. The absence of a significant disparity between the uncaptioned viewing and the unenhanced CV conditions aligns with Dang et al. (2022a), who also found comparable form-recognition gains for collocations across both modes. However, this finding differs from Montero Perez et al. (2014) and Montero Perez et al. (2018), in which captioned viewing led to significantly higher immediate form-recognition posttest scores than uncaptioned viewing. The significant difference between Group 1 and Group 4 contradicts the findings of Montero Perez et al. (2014), where no significant variations were found between the form-recognition gains of the unenhanced and enhanced captioned viewing conditions. One factor that may have contributed to this discrepancy is that while both Montero Perez et al. (2014) and Montero Perez et al. (2018) included MWIs among their target items, single words constituted the majority of their target set (6 MWIs vs. 12 single words in Montero Perez et al., 2018, for example), and it is possible that the absence of enhancement effect was driven primarily by those items. In the present study, a more proportionate number of MWIs (k = 16) and single words (k = 18) were used. Nonetheless, this explanation is tentative, and future research is needed to clarify it.

In addition, the post-hoc analysis on the delayed posttest scores showed that the only significant difference occurred between the scores of Group 1 (MWI-enhanced CV) and Group 4 (unenhanced CV), with Group 1 significantly outperforming Group 4. This finding partially corroborates Majuddin et al. (2021), showing that participants who scored 0 on the pretest were more likely to achieve a full score in the delayed posttest when using enhanced captions compared to unenhanced captions.

Separate ANCOVA analyses of the MWI sections of the posttests revealed that Group 1 significantly outperformed the other groups on both the immediate and delayed posttests, indicating superior initial learning and medium-term retention. This finding contradicts Majuddin et al. (2021), who found no significant disparities in immediate form-recall scores between groups with unenhanced and MWI-enhanced captions. However, it is consistent with Puimège et al. (2023), where TE was a significant predictor of pretest-immediate posttest form-recall gains, with a higher likelihood of MWI-learning in the enhanced condition than its unenhanced counterpart when total reading time was not taken into account. It is useful to compare the findings not only with prior video-based studies, but also with research on reading. Accordingly, the present finding aligns with earlier reading studies showing that, compared to unenhanced items, textually enhanced MWIs capture more learner attention and subsequently promote MWI learning at the level of form-recognition (Boers et al., 2017; Jung et al., 2025; Toomer & Elgort, 2019) and form-recall (Choi, 2017; Jung et al., 2025; Toomer & Elgort, 2019; Vu & Peters, 2022a, 2023). For example, in Toomer and Elgort's (2019) research on incidental collocation acquisition through reading, enhanced reading (bolding of collocations) brought about greater form-recall and form-recognition gains compared to unenhanced reading.

The significant pretest-immediate posttest variation between the MWI-enhanced captioned viewing and uncaptioned viewing indicates that participants in the MWI-enhanced group did not appear to become overwhelmed by the multiplicity of materials presented to them. In other words, although

they were exposed to moving pictures, normal words, and textually-enhanced words while listening to the speaker, they managed to focus on textually-enhanced MWIs and retain them in memory. In fact, the real-time nature of viewing did not significantly interfere with learning the forms of the textually-enhanced MWIs.

Lack of a significant difference between MWI learning in the uncaptioned and captioned viewing groups contradicts Teng's (2019b) and Majuddin et al.'s (2021) results, according to which captioned viewing led to significantly higher immediate posttest scores than uncaptioned viewing in terms of form-recognition and form-recall. However, it is consistent with Dang et al. (2022a), who found no significant disparity between form-recognition collocation gains resulting from these two viewing conditions. One possible explanation for the contradictory findings between the present study and that of Teng (2019b) and Majuddin et al. (2021) may be differences in L2 proficiency level. According to Kurokawa et al. (2025), captioned videos are most beneficial for intermediate-level learners (g = 0.81) compared to beginning to low-intermediate learners (g = 0.28) and high-intermediate to advanced learners (g = 0.29). The participants in the current study were high-intermediate to advanced learners and the discrepancy in the findings may be attributable to this difference in the participants' L2 proficiency level. Another reason can be pretest administration, which likely reduced the captioning advantage by heightening attention to target forms (Kurokawa et al., 2025).

Separate ANCOVA analyses on the single-word sections of the posttests showed that no significant difference existed between the performance of the four groups. The lack of significant variations between the four groups concerning their knowledge of single-word items contrasts with Cintrón-Valentín and García-Amaya (2021), in which the results of both the immediate and two-week delayed form-recall posttests on the L2-Spanish target single words showed that all captioning groups (target vocabulary and grammatical features highlighted using TE) outperformed the uncaptioned control group. Additionally, lack of a significant disparity between the captioned viewing conditions (unenhanced and SW-enhanced) and uncaptioned viewing is in contrast with Hsieh's (2020) study, in which both captioned conditions led to significantly better immediate form-recognition scores than the uncaptioned condition. Moreover, lack of a significant difference between the unenhanced CV and the uncaptioned viewing differs from the studies carried out by Teng (2022a, 2023b) and Wang and Pellicer-Sánchez (2022), where the former condition led to significantly more vocabulary gains at the level of form than the latter.

However, the lack of a significant variation between the learning gains of the unenhanced CV condition and the SW-enhanced CV condition is consistent with Hsieh (2020), in which there were no significant differences between these two conditions in terms of the immediate form-recognition scores. This finding suggests that TE may not be as effective for single words as it is for MWIs. One reason may explain this result. Some of the target MWIs used in the present study consisted of components that were possibly familiar to the participants. For example, in the MWI "play tricks", participants were most probably already familiar with the forms "play" and "tricks" and this may have boosted later recall of the whole MWI; but this was not the case for single words which were among the less frequent English words (e.g. seamlessly). Moreover, one may argue that because the first letter of each component was provided for MWIs, recalling the familiar components can be easier than when the first letter of a totally unfamiliar form is provided. Overall, this interpretation should be taken with a grain of salt because different factors may impact learning of MWIs and single words differently and comparing MWIs with single- word items can be misleading. In fact, since the MWIs and single words are not completely comparable in terms of corpus frequency, difficulty, part of speech, concreteness, and number of letters/syllables, the claim that TE is more effective for MWIs than for single words should be interpreted with caution. This represents a limitation that future research should take into consideration.

In answer to RQ3, it was revealed that in Group 2 (SW-enhanced CV) and Group 3 (uncaptioned viewing) there were significant positive correlations between VST scores and delayed relative gains. These results indicate that the more vocabulary the participants in Group 2 and Group 3 knew, the

greater their delayed lexical gains were. The finding that there was a positive correlation between form-recall delayed gains and prior vocabulary knowledge is consistent with Puimège and Peters (2019b, 2020). In fact, it substantiates the results of prior studies demonstrating that learners with stronger vocabulary knowledge are more likely to acquire new lexical items through viewing than those with more limited vocabulary knowledge (Majuddin et al., 2021; Puimège & Peters, 2019, 2020). This finding supports the "Matthew effect," according to which less knowledgeable students learn less and more knowledgeable students learn more (Stanovich, 2009). Other relative gains, however, did not have a significant correlation with VST scores. This indicates that in the present study, a Matthew effect existed only in the captioned viewing with enhanced single words and uncaptioned viewing conditions, but not in the other conditions. Lack of a significant correlation between prior vocabulary knowledge and incidental vocabulary learning at the level of form in the unenhanced CV condition contradicts several of the previous studies (Ahrabi Fakhr et al., 2021; Montero Perez et al., 2014; Teng, 2024; Vu et al., 2023). However, it supports Dang et al.'s (2022a) study, in which pre-existing vocabulary knowledge did not have any significant relationship with collocation learning (form-recognition) from captioned viewing.

The findings of this study offer at least two theoretical implications. First, the results suggest that the effects of perceptual salience on learning are moderated by the type of lexical item, with textual enhancement facilitating the acquisition of MWIs, but not single words. This indicates that the Noticing Hypothesis may operate differently depending on the type of target item. Second, the fact that enhanced captions supported learning of MWIs despite the substantial processing demands of audiovisual input suggests that textual enhancement may help learners allocate attention to MWIs more effectively during audiovisual input.

6 Limitations and Future Directions

The current study is subject to several limitations. First, it only evaluated vocabulary learning at the level of form recall. To gain a more comprehensive picture of how textual enhancement in captions affects incidental vocabulary learning, future research can use several tests to measure vocabulary development at different sensitivity levels (form recognition, meaning recall/recognition). Second, a transfer-appropriate processing concern arises from the present design. According to the transfer-appropriate processing framework (Morris et al., 1977), learning performance is influenced by the congruency between the conditions of learning and the conditions of testing. Since learners in the captioned groups had access to written target forms during treatment while the uncaptioned group did not, the written posttest format may have systematically advantaged the captioned groups due to learning-testing congruency, making it difficult to determine the extent to which group differences reflected vocabulary learning rather than test-format effects. Third, the initial letters of the target items were supplied as clues to prevent learners from writing alternative items. This differs from L2 users' production of vocabulary items in real-life situations. Fourth, the study was conducted at a single site with English-major university students, limiting generalisability. Vitta et al. (2022) highlighted the importance of multisite designs in L2 vocabulary research. Fifth, English-major students represent a specialised learner profile with potentially different characteristics than non-major learners. Future multisite research with more diverse learner profiles is therefore warranted. Sixth, the sample size per condition was relatively small. However, it should be mentioned that previous research on the effect of mode of input on incidental vocabulary learning has used the same number of participants, at least in one of their experimental conditions (Feng & Webb, 2020; Majuddin et al., 2021). Seventh, word- and learner-related variables, such as frequency, working memory, and visual support (see Teng & Cui, 2025; Kaderoğlu, 2026), were not investigated. Finally, to obtain a more fine-grained picture of the effect of the treatment conditions, future studies are recommended to include a test-only control group.

7 Conclusion

This study primarily aimed to investigate the effect of TE and captioning on incidental learning of MWIs and single words from viewing. The findings reveal that TE can increase the likelihood of noticing and encoding specific lexical items, but its effectiveness varies by lexical unit type (MWI vs single-word) and learner characteristics (prior vocabulary knowledge). Despite the limitations, this research provides empirical support for the favorable impact of textually enhanced captions on MWI learning through viewing. This study extends the earlier body of research on the efficacy of TE for incidental MWI learning through different modes of input, including reading (Choi, 2017; Vu & Peters, 2022a, 2023) and captioned viewing (Puimège et al., 2023).

Taken together, the results of this study highlight that TE can guide learner attention in ways that support vocabulary learning during video viewing. For MWIs in particular, TE appears to lower processing barriers associated with parsing multiword sequences in fast, transient input. Pedagogically, the study suggests that language teachers and material designers should consider selectively enhancing lexical items in captioned videos when the learning objective involves vocabulary acquisition, especially for MWIs. However, TE may need to be used judiciously for single words, as enhancing unfamiliar and infrequent forms might increase cognitive load without guaranteeing retention.

Appendix A

Target MWIs

Target MWI	Type	Frequency of occurrence	MI	Target MWI	Type	Frequency of occurrence
Optical illusion	Collocation (Adj+ N)	3	8/86	Make (sth) out	Phrasal verb	1
Beg the question	Collocation (V+N)	1	4/28	Conjure up	Phrasal verb	1
Deadly serpent	Collocation (Adj+ N)	1	4/23	Pull off	Phrasal verb	1
Play tricks	Collocation (V+N)	1	3/57	Stand the test of time	Idiom	1
Age-old question	Collocation (Adj+ N)	1	5/08	Stick out like a sore thumb	Idiom	1
Sheer beauty	Collocation (Adj+ N)	1	4/92	Vanish into thin air	Idiom	1
Invisibility cloak	Collocation (N+ N)	1	11/77			
Pose a threat	Collocation (V+N)	1	7/89			
Confer protection	Collocation (V+N)	1	4/88			
Broad daylight	Collocation (Adj+ N)	1	8/21			

Appendix B

Target single words

Target word	Part of Speech	Frequency of occurrence	Target word	Part of Speech	Frequency of occurrence
Limb	Noun	1	Deploy	Verb	1
Mush	Noun	1	Conniving	Adjective	1
Fringe	Noun	1	Sinister	Adjective	1
Con	Noun	2	Phony	Adjective	1
Giveaway	Noun	1	Cunning	Adjective	2
Ploy	Noun	1	Barefaced	Adjective	1
Outwit	Verb	2	Pygmy	Adjective	1
Gallop	Verb	1	Seamlessly	Adverb	1
Deter	Verb	1			
Trot	Verb	1			

References

Ahrabi Fakhr, M., Borzabadi Farahani, D., & Khomeijani Farahani, A. A. (2021). Incidental vocabulary learning and retention from audiovisual input and factors affecting them. *English Teaching & Learning*, *45*(2), 167–188. https://doi.org/10.1007/s42321-020-00066-y

Alshumrani, H. A. (2024). Unveiling vocabulary teaching and learning beliefs of teachers and learners in an EFL context. *Asian-Pacific Journal of Second and Foreign Language Education*, *9*(1), 1–19. https://doi.org/10.1186/s40862-023-00242-0

Boers, F. (2020). Factors affecting the learning of multiword items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 143–157). Routledge. https://doi.org/10.4324/9780429291586

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

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136–163. http://hdl.handle.net/10125/66816

Choi, S. (2017). 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. 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

Dang, T. N. Y., Lu, C., & Webb, S. (2022a). Incidental learning of collocations in an academic lecture through different input modes. *Language Learning*, *72*(3), 728–764. https://doi.org/10.1111/lang.12499

Dang, T. N. Y., Lu, C., & Webb, S. (2022b). Incidental learning of single words and collocations through viewing an academic lecture. *Studies in Second Language Acquisition*, *44*(3), 708–736. https://doi.org/10.1017/S0272263121000474

Dang, T. N. Y., Lu, C., & Webb, S. (2023). Open access academic lectures as sources for incidental vocabulary learning: Examining the role of input mode, frequency, type of vocabulary, and elaboration. *Applied Linguistics*, *44*(4), 747–770. https://doi.org/10.1093/applin/amac044

Davies, M. (2008). *Corpus of contemporary American English (COCA)*. https://corpus.byu.edu/coca/

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, *88*, 102166. https://doi.org/10.1016/j.system.2019.102166

Durbahn, M., Rodgers, M., Macis, M., & Peters, E. (2024). Lexical coverage in L1 and L2 viewing comprehension. *Studies in Second Language Acquisition*, *46*(4), 1045–1068. https://doi.org/10.1017/S0272263124000391

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163–188. https://doi.org/10.1177/0267658309349431

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text - Interdisciplinary Journal for the Study of Discourse*, *20*(1), 29-62. https://doi.org/10.1515/text.1.2000.20.1.29

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, *42*(3), 499–523. https://doi.org/10.1017/S0272263119000494

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2020). Vocabulary learning through viewing captioned or subtitled videos and the role of learner- and word-related factors. *CALICO Journal*, *37*(3), 233–253. https://doi.org/10.1558/cj.39370

Hao, T., Sheng, H., Ardasheva, Y., & Wang, Z. (2022). Effects of dual subtitles on Chinese students' English listening comprehension and vocabulary learning. *The Asia-Pacific Education Researcher*, *31*(5), 529–540. https://doi.org/10.1007/s40299-021-00601-w

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, *33*(5–6), 567–589. https://doi.org/10.1080/09588221.2019.1577898

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430. https://doi.org/10.64152/10125/66973

Jin, Z., & Webb, S. (2020). Incidental vocabulary learning through listening to teacher talk. *The Modern Language Journal, 104*(3), 550–566. https://doi.org/10.1111/modl.12661

Jung, J., & Lee, M. (2024). Incidental collocational learning from reading-while-listening and the impact of synchronized textual enhancement. *International Review of Applied Linguistics in Language Teaching*, *62*(2), 1935-1958. https://doi.org/10.1515/iral-2023-0029

Jung, J., Stainer, M. J., & Tran, M. H. (2025). The impact of textual enhancement and frequency manipulation on incidental learning of collocations from reading. *Language Teaching Research*, *29*(6), 2679-2708. https://doi.org/10.1177/13621688221129994

Kaderoğlu, K. (2026). Incidental vocabulary learning from TV: Subtitles and imagery. *International Journal of TESOL Studies*, *8*(3), 67-81. https://doi.org/10.58304/ijts.250907

Kurokawa, S., Hein, A. M., & Uchihara, T. (2025). Incidental vocabulary acquisition through captioned viewing: A meta-analysis. *Language Learning, 75*(4), 939–987. https://doi.org/10.1111/lang.12697

Lee, B. J. (2021). The effects of proficiency and textual enhancement technique on noticing. *System*, *96*, 102407. https://doi.org/10.1016/j.system.2020.102407

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*(1), 105–129. https://doi.org/10.1080/14790718.2012.679275

Ma, X., & Reynolds, B. L. (2023). The effect of reading purpose on incidental acquisition and retention of vocabulary from reading. *Asian Journal of English Language Teaching, 32*(1), 63-104. https://doi.org/10.65961/AJELT-2023-1-004

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

Mayer, R. E. (2021). Evidence-based principles for how to design effective instructional videos. *Journal of Applied Research in Memory and Cognition*, *10*(2), 229–240. https://doi.org/10.1016/j.jarmac.2021.03.007

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, *42*(4), 749–773. https://doi.org/10.1017/S0272263119000706

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*(1), 118–141.

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, *31*(1–2), 1–26. https://doi.org/10.1080/09588221.2017.1375960

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*, 9–13.

Nation, I. S. P., & Heatley, A. (2002). *Range: A program for the analysis of vocabulary in texts.* http://www.vuw.ac.nz/lals/staff/paulnation/nation.aspx

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, *27*(1), 94-106. https://doi.org/10.2307/747835

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, *21*(3), 298–320. https://doi.org/10.1177/1362168816639619

Norouzian, R., & Plonsky, L. (2018). Eta- and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research*, *34*(2), 257–271. https://doi.org/10.1177/0267658316684904

Pellicer-Sánchez, A., Siyanova-Chanturia, A., & Parente, F. (2022). The effect of frequency of exposure on the processing and learning of collocations: A comparison of first and second language readers' eye movements. *Applied Psycholinguistics*, *43*(3), 727–756. https://doi.org/10.1017/S014271642200011X

Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, *169*(1), 142–168. https://doi.org/10.1075/itl.00010.pet

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. https://doi.org/10.1002/tesq.531

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing l2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Puimège, E., Montero Perez, M., & Peters, E. (2023). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study. *Second Language Research*, *39*(2), 471–492. https://doi.org/10.1177/02676583211049741

Puimège, E., Montero Perez, M., & Peters, E. (2024). The effects of typographic enhancement on L2 collocation processing and learning from reading: An eye-tracking study. *Applied Linguistics*, *45*(1), 88–110. https://doi.org/10.1093/applin/amad003

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, *47*(4), 424–438. https://doi.org/10.1080/09571736.2019.1638630

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing l2 television and factors that affect learning. *Studies in Second Language Acquisition*, *42*(3), 525–549. https://doi.org/10.1017/S027226311900055X

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, *47*(4), 479–496. https://doi.org/10.1080/09571736.2019.1616806

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Rodgers, M. P. H., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *ITL - International Journal of Applied Linguistics*, *171*(2), 191–220. https://doi.org/10.1075/itl.18034.rod

Sánchez Gutiérrez, C. H., Serrano, M. P., & García, P. R. (2019). The effects of word frequency and typographical enhancement on incidental vocabulary learning in reading. *Journal of Spanish Language Teaching*, *6*(1), 14–31. https://doi.org/10.1080/23247797.2019.1590000

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, *11*, 11–26.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (1st ed., pp. 3–32). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.003

Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education*, *189*(1–2), 23–55. https://doi.org/10.1177/0022057409189001-204

Sutton, D., & Webb, S. (2026). The effects of audiovisual input on second language learning A meta-analysis. *Studies in Second Language Acquisition*, 1–25. Published online. https://doi.org/10.1017/S0272263126101612

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73.

Tavakoli, P., & Uchihara, T. (2020). To what extent are multiword sequences associated with oral fluency? *Language Learning*, *70*(2), 506–547. https://doi.org/10.1111/lang.12384

Teng, M. F. (2019a). Incidental vocabulary learning for primary school students: The effects of L2 caption type and word exposure frequency. *The Australian Educational Researcher*, *46*(1), 113–136. https://doi.org/10.1007/s13384-018-0279-6

Teng, M. F. (2019b). The effects of video caption types and advance organizers on incidental L2 collocation learning. *Computers & Education*, *142*, 103655. https://doi.org/10.1016/j.compedu.2019.103655

Teng, M. F. (2022a). Incidental L2 vocabulary learning from viewing captioned videos: Effects of learner-related factors. *System*, *105*, 102736. https://doi.org/10.1016/j.system.2022.102736

Teng, M. F. (2022b). Vocabulary learning through videos: Captions, advance-organizer strategy, and their combination. *Computer Assisted Language Learning*, *35*(3), 518–550. https://doi.org/10.1080/09588221.2020.1720253

Teng, M. F. (2023a). Effectiveness of captioned videos for incidental vocabulary learning and retention: The role of working memory. *Computer Assisted Language Learning*, 1–29. https://doi.org/10.1080/09588221.2023.2173613

Teng, M. F. (2023b). Incidental vocabulary learning from captioned videos: Learners' prior vocabulary knowledge and working memory. *Journal of Computer Assisted Learning*, *39*(2), 517–531. https://doi.org/10.1111/jcal.12756

Teng, M. F. (2024). Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System*, *124*, 103381. https://doi.org/10.1016/j.system.2024.103381

Teng, M. F. (2025a). Incidental learning of multiword expressions from bilingual captioned videos for language minority students. *The JALT CALL Journal, 21*(1), 1201. https://doi.org/10.29140/jaltcall.v21n1.1201

Teng, M. F. (2025b). Modality of input and factors affecting incidental vocabulary learning: reading, listening, and viewing with captions. *Applied Linguistics Review, 16*(4), 1607-1635. https://doi.org/10.1515/applirev-2024-0021

Teng, M. F. (2026). Incidental vocabulary learning from captioned video genres: Proficiency, working memory, and aptitude. *Computer Assisted Language Learning, 39*(3), 644-686. https://doi.org/10.1080/09588221.2024.2421517

Teng, M. F., & Cui, Y. (2024). Comparing incidental learning of single words and collocations from different captioning conditions: The role of vocabulary knowledge and working memory. *Journal of Computer Assisted Learning, 40*(3), 973–989. https://doi.org/10.1111/jcal.12910

Teng, M. F., & Cui, Y. (2025). Incidental vocabulary learning from captioned viewing: A forest plot of word- and learner-related factors. *Language Teaching Research*, 1-32. Published online. https://doi.org/10.1177/13621688251372987

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

Vanderplank, R. (2016a). 'Effects of' and 'effects with' captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, *49*(2), 235–250. https://doi.org/10.1017/S0261444813000207

Vanderplank, R. (2016b). The state of the art II: Selected research on other issues in watching captioned TV, films and video. In R. Vanderplank, *Captioned Media in Foreign Language Learning and Teaching* (pp. 105–148). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-50045-8_5

Vitta, J. P., Nicklin, C., & McLean, S. (2022). Effect size–driven sample-size planning, randomization, and multisite use in L2 instructed vocabulary acquisition experimental samples. *Studies in Second Language Acquisition*, *44*(5), 1424-1448. https://doi.org/10.1017/S0272263121000541

Vu, D. V., & Peters, E. (2022a). Incidental learning of collocations from meaningful input: A longitudinal study into three reading modes and factors that affect learning. *Studies in Second Language Acquisition*, *44*(3), 685–707. https://doi.org/10.1017/S0272263121000462

Vu, D. V., & Peters, E. (2022b). Learning vocabulary from reading-only, reading-while-listening, and reading with textual input enhancement: Insights from Vietnamese EFL learners. *RELC Journal*, *53*(1), 85–100. https://doi.org/10.1177/0033688220911485

Vu, D. V., & Peters, E. (2023). A longitudinal study on the effect of mode of reading on incidental collocation learning and predictors of learning gains. *TESOL Quarterly*, *57*(1), 5–32. https://doi.org/10.1002/tesq.3111

Wang, A., & Pellicer-Sánchez, A. (2022). Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study. *Language Learning*, *72*(3), 765–805. https://doi.org/10.1111/lang.12495

Wang, Y. (2019). Effects of l1/l2 captioned TV programs on students' vocabulary learning and comprehension. *CALICO Journal*, *36*(3), 204–224. https://doi.org/10.1558/cj.36268

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge. https://doi.org/10.4324/9780429291586

Webb, S., & Chang, A. C.-S. (2022). How does mode of input affect the incidental learning of collocations? *Studies in Second Language Acquisition*, *44*(1), 35–56. https://doi.org/10.1017/S0272263120000297

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning, 59*(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, *168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web

Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, *56*(2), 161–180. https://doi.org/10.1017/S0261444822000507

Wi, I., & Boers, F. (2025). A synthesis of research on L2 vocabulary learning through audiovisual input and on-screen text. *Digital Applied Linguistics*, *3*, 102774. https://doi.org/10.29140/dal.v3.102774

Zhou, J., & Day, R. R. (2020). The incidental learning of L2 Chinese vocabulary through reading. *Reading in a Foreign Language, 32*(2), 169-193. https://doi.org/10.64152/10125/67379

Maryam Ehsani holds a PhD in Applied Linguistics from the University of Tehran, Iran. She currently teaches undergraduate courses in TEFL in the ELT Department, University of Hormozgan, Iran. Her primary research interest is second language vocabulary acquisition, use, and assessment, with a particular focus on incidental vocabulary learning. Her other areas of interest include L2 viewing comprehension, technology-based language learning, and critical discourse analysis.