Applied Research on English Language

V. 15 N. 3 2026
pp: 63-84
http://are.ui.ac.ir

DOI: 10.22108/are.2026.147433.2659
Document Type: Research Article

The Effects of Captioning and Textual Enhancement on L2 Viewing Comprehension: Insights from Iranian EFL Learners

Maryam Ehsani [1], Hossein Karami [1], Tahmineh Khalili [2*]

[1] *Department of English Language and Literature, Faculty of Foreign Languages and Literatures, University of Tehran, Tehran, Iran*

[2] *Department of Teaching English as a Foreign Language (TEFL), Faculty of Humanities, University of Hormozgan, Bandar Abbas, Iran*

Received: 2025/11/13 **Accepted: 2026/07/11**

Abstract: Studies on the effects of caption condition and the textual enhancement of lexical items in captions on L2 viewing comprehension have shown mixed results. This study investigated whether caption condition and textual enhancement influence Iranian EFL learners' viewing comprehension. To this end, the following viewing conditions were examined: 1) captioned viewing with 16 textually-enhanced multiword items, 2) captioned viewing with 18 textually-enhanced single words, 3) uncaptioned viewing, and 4) unenhanced captioned viewing. Sixty-two Iranian EFL learners from four intact classes were randomly allocated to one of the viewing conditions. Viewing comprehension was assessed using a 10-item comprehension test. The results showed that the textual enhancement of multiword items and single words had a positive impact on their comprehension scores. Another noteworthy outcome was that unenhanced captioned viewing and uncaptioned viewing did not lead to significantly different comprehension scores. Additionally, questionnaire data indicated that participants were accustomed to watching English videos and that the majority of them watched English videos with L2 captions. The findings suggest that textual enhancement can be incorporated into captioned materials without compromising comprehension among high-intermediate to advanced EFL learners. The findings have implications for technological developments in L2 education, including future research, instructional design, and teacher education, particularly in EFL contexts.

Keywords: Caption, L2 Learning, Listening, Textual Enhancement, Viewing Comprehension.

* **Corresponding Author.**

Authors' Email Address:

[1] **Maryam Ehsani** (Maryam.Ehsani1996@ut.ac.ir), [2] **Hossein Karami** (hkarami@ut.ac.ir), [3] **Tahmineh Khalili** (tkhalili@hormozgan.ac.ir)

2322-5343 © University of Isfahan
This is an open access article under the CC BY-NC-ND/4.0/ License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

The meaning of "comprehension" seems to differ among individuals (Lee et al., 2025). What researchers seem to agree on about the meaning of comprehension might be a range of abilities and processes involved in the comprehension process (Kendeou et al., 2007). Various types of comprehension share basic processes, including interpreting content, utilizing existing knowledge to help this interpretation, and eventually generating a coherent mental representation of the processed content (Kendeou et al., 2007). Recently, the majority of input comprehension studies have focused on reading (Choi & Zhang, 2021; Tong et al., 2023; Zhang & Zhang, 2022), listening (Du & Man, 2022; Vafaee & Suzuki, 2020; Zhang & Zhang, 2022) and reading-while-listening (Hui, 2024; Pellicer-Sánchez et al., 2020; Serrano & Pellicer-Sánchez, 2022), but this trend has started to change during the last decade given the popularity of audio-visual materials. This change seems to be timely because research indicates that nowadays L2 learners tend to receive more audio-visual input than other input types, such as reading or listening materials (Muñoz et al., 2023). To illustrate, in a study undertaken by Ahrabi Fakhr et al. (2021), the questionnaire results showed that Iranian EFL students majoring in English watched English videos more frequently than they read English books or listened to English materials. Thus, it can also be summarized that technology could promote L2 comprehension dramatically.

Viewing comprehension may pose a challenge for L2 learners (Schroeders et al., 2010). Various techniques have been devised to facilitate comprehension, with captioning being one of the most widely used ones (Teng, 2025). Some research evidence suggests that captions can enhance viewing comprehension (Sidorov et al., 2020). For example, Pujadas and Muñoz (2024) found that captioned viewing led to superior comprehension scores compared to uncaptioned viewing. In a study with a between-subjects design, Teng (2025) found that viewing captioned videos resulted in improvements in comprehension scores.

Few studies can be found on the effect of textual enhancement (TE) of vocabulary items in captions on viewing comprehension. TE has been defined as manipulating a text to make certain linguistic targets more perceptually salient (Lee, 2021). In vocabulary learning studies, this technique has been employed to draw L2 learners' attention to target lexical items and promote their acquisition. However, the impact of implementing this technique on students' viewing experience and viewing comprehension is yet unresolved. In Hsieh's (2020) study, the results showed that uncaptioned viewing, fully-captioned viewing, and fully captioned viewing plus textually enhanced target single words did not differ significantly from each other in terms of comprehension scores. However, the fully-captioned viewing

condition resulted in non-significantly better scores than the other conditions. The writer pointed out that the presence of textually enhanced words in the captioning line might have directed learners' attention to vocabulary rather than video content.

Majuddin et al. (2021) conducted a study with Malaysian L2 English learners randomly allocated to one of the experimental conditions. These conditions differed in their caption condition (no captions, unenhanced captions, captions + TE on target multiword items). The results showed that learners in the unenhanced captioned viewing condition significantly outperformed those in the uncaptioned condition. However, rather surprisingly, enhanced captioned viewing did not result in significantly higher comprehension scores compared to the uncaptioned condition, suggesting a negative effect of TE on video comprehension.

Despite many advances in TE and captioning, very few studies have focused on this aspect of EFL education in Iranian settings (Roohani et al., 2013). Existing findings are inconsistent, making it difficult for teachers and materials developers to determine whether textual enhancement should be incorporated into captioned materials for Iranian EFL learners. So, the question remains as to whether directing Iranian EFL learners' attention toward enhanced lexical items facilitates or interferes with content comprehension. Considering the abundance of audiovisual input received by many EFL learners and the changing conditions of EFL learning that increase opportunities for language exposure, the paucity of research in this area needs to be addressed. Applying this approach in research can also pave the way for practical applications in EFL education.

Literature Review

Theoretical Background

Two theories that are closely related to the present investigation are the Noticing Hypothesis (Schmidt, 1990) and the Cognitive Theory of Multimedia Learning (Mayer, 2021). The Noticing Hypothesis posits that conscious attention to linguistic forms is a necessary condition for language learning. According to this hypothesis, language input cannot become intake unless learners consciously notice specific linguistic features. Textual Enhancement (TE) was developed as an instructional technique to increase the perceptual salience of target linguistic forms and thereby promote noticing. By making target items more visually prominent, TE is expected to increase the likelihood that learners notice and subsequently acquire them.

According to the Cognitive Theory of Multimedia Learning (Mayer, 2021), meaningful learning occurs when learners actively select relevant verbal and visual information, organize

it into coherent mental representations, and integrate these representations with prior knowledge. Consequently, presenting words together with corresponding images generally promotes deeper learning than presenting words alone. However, this theory also assumes that working memory has a limited processing capacity. Learners can process only a finite amount of information in the verbal and visual channels at any given time (Baddeley, 2010). When instructional materials exceed this limited capacity, cognitive overload may occur, reducing learning effectiveness. This assumption has important implications for captioned video viewing. While captions may facilitate comprehension and vocabulary learning by providing additional verbal input, excessive on-screen text may increase extraneous cognitive load and distract learners' attention from the video content.

Captions and Viewing Comprehension

Captions refer to on-screen text in the same language (Sydorenko, 2010; Vanderplank, 2016a; 2016b). They were initially used to enable deaf and hard-of-hearing individuals to understand videos, but were later exploited to benefit those with normal hearing who nevertheless had difficulties with L2 listening comprehension (Vanderplank, 2016b; 2016a). Recently, there has been a proliferation of studies investigating the potential benefits of captions for language learning purposes. Although many studies have aimed to explore the efficacy of captions for vocabulary learning, other language areas have not been neglected. In fact, research suggests that captioning videos can boost learning of different language components, including grammar (e.g., Pattemore, 2022), pronunciation (e.g., Mohsen & Mahdi, 2021), and vocabulary (e.g., Fievez et al., 2023; Teng, 2022a; 2022b).

In Winke et al.'s (2010) study, second- and fourth-year learners of Arabic, Chinese, Spanish, and Russian watched three short L2 videos twice, once with captioning and once without. For each target language, one group saw captions during the first viewing of the video and another saw captions during the second viewing. Learners of Spanish included two extra groups, one which viewed the videos twice without captions and one which viewed them twice with captions. After the second viewing, a comprehension test was given. Results demonstrated that captioned videos were more beneficial than uncaptioned videos for overall video comprehension.

Pujadas and Muñoz (2024) examined the effects of L2 proficiency (A1 to C2, measured by the OPT) and vocabulary knowledge on the comprehension of captioned and uncaptioned English TV series. The results revealed that captioned viewing led to superior comprehension

scores compared to uncaptioned viewing. Additionally, both L2 proficiency and vocabulary knowledge were positively correlated with comprehension scores.

Apart from studies that have probed the effect of various on-screen texts on viewing comprehension, a few studies have compared comprehension of captioned videos and comprehension of other modes of input. To exemplify, in a study with 80 pre-intermediate Vietnamese EFL learners, Vu et al. (2023) found that reading-while-listening resulted in better comprehension scores than captioned TV viewing.

Textual Enhancement and Content Comprehension

Typographic or Textual Enhancement (TE) 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). TE has been investigated mostly in the area of grammar learning through reading and more recently, through captioned viewing (Della Putta, 2019; Jahan & Kormos, 2015; LaBrozzi, 2016; Lee, 2021; Lee & Jung, 2024; Lee & Révész, 2018; Lee, 2007; Meguro, 2019; Pattemore, 2022; Rassaei, 2015; Simard, 2009). The findings of this line of research, however, indicate mixed patterns. For instance, the findings of Lee's (2007) research showed that TE facilitated the learning of target grammatical forms but had unfavorable effects on reading comprehension. Contrary to Lee's (2007) findings, in Meguro's (2019) study, TE of target grammatical constructions did not hinder L2 English learners' reading comprehension. LaBrozzi (2016) compared the effects of different types of TE on recognition of L2 grammatical forms and reading comprehension. Different types of TE were explored, including underlining, italicizing, bolding, using capital letters, increasing font size, and changing the font. The results demonstrated that reading comprehension was unaffected by the type of TE used.

The effect of TE on vocabulary items on content comprehension in different input modes has also been investigated. As for reading, Puimege et al. (2024) asked L2 learners of English to read 10 English texts adapted from a popular science book and TED-ED scripts. Post-intervention results indicated that participants primarily focused on content comprehension, and TE did not cause learners to pay conscious attention to the form of L2 collocations.

Taking the contradictory findings of previous studies into consideration, the present study aims to examine the impact of captions and TE of multiword items and single words on EFL learners' viewing comprehension. Additionally, learners' attitudes and viewing experiences are scrutinized. From a theoretical perspective, whether TE facilitates or hinders

learning during captioned viewing may depend on the balance between increased noticing (Schmidt, 1990) and increased processing demands (Mayer, 2021). If the enhancement successfully directs learners' attention without overwhelming working memory, vocabulary learning and comprehension may improve simultaneously. Conversely, if the additional visual salience imposes excessive cognitive load, learners may struggle to process both the message and the highlighted lexical items. Comparing different caption conditions, therefore, provides an opportunity to examine how lexical salience and cognitive load interact during L2 viewing. Although previous studies have explored these issues, their findings remain inconsistent, and evidence from Iranian EFL contexts is scarce. Accordingly, empirical investigation is required to determine whether textual enhancement of single words and multiword items in captions facilitates learning without compromising viewing comprehension. The present study, therefore, addresses the following two research questions:

1. Does caption condition affect Iranian EFL learners' viewing comprehension significantly?
2. What do learners report regarding their viewing experience and attention to content and vocabulary under each captioning condition?

Methods

This study tried to explore the effect of captions and TE of multiword items (MWIs) and single words in captions on L2 learners' viewing comprehension. The viewing conditions included the following, randomly distributed among four intact classes: 1) captioned viewing with textually enhanced MWIs (MWI-enhanced captioned viewing), 2) captioned viewing with textually enhanced single words (single-word enhanced captioned viewing), 3) uncaptioned viewing, and 4) captioned viewing with no TE (unenhanced captioned viewing).

Participants

Sixty-two Iranian undergraduates majoring in English as a Foreign Language (TEFL) from four intact classes participated in this study. All participants knew the most frequent 3,000 English words according to the results of Webb et al.'s (2017) Updated Vocabulary Levels Test (UVLT). The UVLT was administered to ensure that participants possessed sufficient lexical knowledge to understand the documentary. Participants' ages ranged from 18 to 23 (mean age = 20.46, SD = 1.30), and the sample included 18 male and 44 female students. Table 1 displays participants' background information.

Table 1. Descriptive Statistics for Background Information of Participants

		Age					Gender	
		N	Minimum	Maximum	Mean	Standard Deviation	Male	Female
							N	N
Group	Group 1	16	18.00	21.00	19.19	.98	7	9
	Group 2	16	19.00	22.00	20.19	.83	5	11
	Group 3	14	20.00	23.00	20.57	.94	3	11
	Group 4	16	21.00	23.00	21.94	.57	3	13

Materials and Instruments

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 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 documentary script using RANGE (Nation & Heatley, 2002) revealed that 91/5% of the words came from the 3,000 most frequent English word families. In a study on the relationship between vocabulary and viewing comprehension, Durbahn et al. (2020) found that a lexical coverage of about 90% is enough for adequate viewing comprehension without external help. Additionally, van Zeeland and Schmitt (2013) found no significant difference in comprehension of spoken input between 90 % and 95% text coverage. Following Durbahn et al. (2020) and van Zeeland and Schmitt (2013), this project deemed 90% coverage as appropriate. According to the participants' VLT scores and the results of the pilot study, the participants of the main study were believed to be able to follow the video with relative ease.

As mentioned earlier in the literature review section, some evidence from prior research indicates that vocabulary knowledge and L2 proficiency are correlated with L2 input comprehension (Pujadas & Muñoz, 2020, 2024). Taking this into account, and to ensure inter-group homogeneity, we decided to measure not only participants' vocabulary knowledge but also their L2 proficiency. Participants' vocabulary knowledge was assessed using Webb et al.'s (2017) Updated Vocabulary Levels Test (UVLT) and Nation and Beglar's (2007) Vocabulary Size Test (VST). The UVLT scores showed that all participants knew the most frequent 3,000 English words. 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. Another test that was given to the participants was the

Quick Oxford Placement Test (OPT) (Allan, 2004). The result of this test demonstrated that students' general English proficiency ranged from 30 to 54 (mean= 40.62, SD= 6.91). These results indicated that the participants' English proficiency level ranged from pre-intermediate or B1 (according to the Common European Framework of Reference) to advanced or C1. Three one-way analyses of variance (ANOVA) were run, showing no significant differences between the groups' UVLT, VST, and OPT scores ($p. > 0.05$ in all cases).

Participants completed a 10-item true/false test designed to assess their comprehension of the BBC documentary. This test did not include any of the textually-enhanced items (which were all unknown to the participants). The questions tapped into skills such as discerning the main idea, identifying supporting details, and drawing inferences from the context. The internal consistency of the comprehension test was examined using the Kuder-Richardson formula 20 (KR-20), which is the appropriate reliability estimate for dichotomously scored items. The resulting coefficient (KR-20= .75) indicated acceptable reliability for the test. A post-viewing questionnaire, adapted from Puimège and Peters (2020), Peters and Webb (2018), and Puimège and Peters (2019), was given to the participants as well. The questionnaire was translated into Persian (participants' native language) by the first researcher. Content validity was confirmed using judgments of two experts in Applied Linguistics. Internal consistency of the questionnaire in the current study was acceptable (Cronbach's $\alpha = .78$). Both the comprehension test and the questionnaire were scored by the first researcher.

Procedure

The data were gathered within three weeks. In the initial week, participants were briefed on the research (without revealing its specific aims) before they gave consent to take part in the study. They were given the OPT test during this week to measure their English language proficiency. In the second week, they completed the UVLT and the VST. One week later, participants watched the video under their assigned conditions, using the multimedia system (computer, projector, and speakers) in their classrooms. Group 1 watched the documentary with captions that included 16 textually-enhanced unknown MWIs. Group 2 watched the documentary with captions that involved 18 textually-enhanced unknown single words. Textual enhancement was in the form of bolding and underlining. Group 3 watched the documentary without captions, and Group 4 watched it with unenhanced captions. Before exposure to the learning materials, participants were told they would later answer comprehension questions. Immediately after the treatment, they took the comprehension test

(reliability = .75), followed by the post-viewing questionnaire. After completion of the questionnaire, participants were debriefed about the real purposes of the study.

Data Analysis

Responses were scored using a binary method, according to which correct responses received 1 and incorrect responses received 0. Normality was confirmed by checking the skewness and kurtosis values. All statistical analyses were implemented using SPSS software (version 26.0). Given the normal distribution of the data, a one-way ANOVA was used to answer RQ1. RQ2 was answered through descriptive analyses of the questionnaire data.

Results and Discussion

Results

Comprehension Test

Participants' comprehension test scores showed that, irrespective of their viewing condition, all four groups comprehended the content of the video well. The average comprehension score was 7.95 (standard deviation = 1.36). Table 2 summarizes the results of the comprehension test for each group. Levene's test for homogeneity of variances (Table 3) revealed that the variance in scores was the same for the four groups ($p = .615$, $p > 0.05$).

Table 2. Descriptive Statistics for the Comprehension Test per Group

	N	Mean	Std. Deviation
Group 1	16	8.6875	1.07819
Group 2	16	8.3750	1.20416
Group 3	14	7.9286	1.49174
Group 4	16	6.8125	.91059
Total	62	7.9516	1.36018

Note: The maximum possible score was 10.

Table 3. Results of the Levene's Test for Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Comp.	Based on Mean	.604	3	58	.615
	Based on Median	.603	3	58	.616
	Based on Median and with adjusted df	.603	3	51.565	.616
	Based on trimmed mean	.566	3	58	.640

ANOVA results (Table 4) showed that there was a significant difference between participants' comprehension scores, F (3, 58) = 7.752, p = .000. Post-hoc comparisons using the Tukey HSD test (Table 5) indicated that the mean score for Group 4 (M = 6.81, SD = .91) was significantly lower than that for Group 1 (M = 8.68, SD = 1.07) and Group 2 (M = 8.37, SD = 1.2). Other mean scores were not significantly different from each other. According to Cohen's (1988) eta squared guidelines (as cited in Pallant, 2016, p. 212), 0.01 is considered a small effect, 0.06 a medium effect, and 0.14 a large effect. Here, the effect size, calculated using eta squared, was 0.286, which indicates that the difference in comprehension test mean scores between the groups was large.

Table 4. ANOVA Results

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	32.301	3	10.767	7.752	.000
Within Groups	80.554	58	1.389		
Total	112.855	61			

Table 5. Multiple Comparisons for Comprehension Scores

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Group 1	Group 2	.31250	.41666	.876	-.7896	1.4146
	Group 3	.75893	.43129	.303	-.3819	1.8997
	Group 4	1.87500*	.41666	.000	.7729	2.9771
Group 2	Group 1	-.31250	.41666	.876	-1.4146	.7896
	Group 3	.44643	.43129	.730	-.6944	1.5872
	Group 4	1.56250*	.41666	.002	.4604	2.6646
Group 3	Group 1	-.75893	.43129	.303	-1.8997	.3819
	Group 2	-.44643	.43129	.730	-1.5872	.6944
	Group 4	1.11607	.43129	.057	-.0247	2.2569
Group 4	Group 1	-1.87500*	.41666	.000	-2.9771	-.7729
	Group 2	-1.56250*	.41666	.002	-2.6646	-.4604
	Group 3	-1.11607	.43129	.057	-2.2569	.0247

*. The mean difference is significant at the 0.05 level.

Questionnaire

Viewing habits: The results of the first section of the questionnaire showed that 98.3% of the respondents (out of 59) watch English videos (including films, series, TV programs, YouTube, etc.). Of 57 respondents, 34 (59.6%) reported that they watch English videos every day, 20 (35.1%) claimed that they watch English videos once a week, and 3 (5.3%) stated that they watch English videos once a month. Of 58 respondents, 12 (20.7%) reported that they watch English videos without any captions or subtitles, 14 (24.1%) reported that they use Persian subtitles while watching English videos, 24 (41.4%) stated that they watch English videos with English captions, 7 (12.1%) claimed that they watch English videos sometimes with Persian subtitles and sometimes with English captions, and 1 person (1.7%) reported that she watches English videos sometimes with and sometimes without English captions.

Affective response and content comprehension: According to the first item of the second section of the questionnaire, 46 participants (74.2%) found the topic of the video interesting (agree or strongly agree), 12 (19.4%) were uncertain, and a few did not find the topic interesting. Regarding the length of the video, 31 participants (50%) reported that it was appropriate, 16 (25.8%) were uncertain, and others believed that the length was not appropriate (it was long). As for the understandability of the video, 38 participants (61.3 %) stated that its content was easily understandable, 13 (21%) participants were uncertain, and 10 (16.2%) did not find the content easily understandable. Forty-six participants (74.2%) reported that they mainly focused on the content of the video; 8 participants (12.9%) were uncertain, and 7 (11.3%) either disagreed or strongly disagreed with the claim.

Vocabulary awareness: Thirty-two participants (51.6%) reported that they paid attention to the vocabulary items that were used in the video, 19 (30.6%) were uncertain, and 10 (16.1%) stated that they did not pay attention to the vocabulary items that were used in the video.

Attitudes towards viewing L2 videos and language learning: Answers to the sixth question of the second section of the questionnaire showed that 54 participants (88.5%) either agreed or strongly agreed that watching videos is a good way to improve one's English proficiency. Answers to the other three questions of this section were also illuminating. It was shown that 56 students (91.8%) either agreed or strongly agreed that watching English videos is a good way to improve L2 learners' listening skills. Fifty-two participants (85.3%) either agreed or strongly agreed that watching videos is a good way to improve one's English vocabulary. Participants appeared less certain about the effectiveness of English videos for

improving L2 learners' English grammar; 32 of them (53.4%) selected either agree or strongly agree, and 14 students (23.3%) rated this item 3, indicating their lack of certainty.

Open-ended questions: The other section of the questionnaire tapped into learners' comprehension of the documentary. The answers to this section indicated that participants had no trouble understanding the gist of the content. Specifically, the questionnaire item "what have you learned in terms of content?" elicited references to central topics in the documentary. For example, participants mentioned "I have learned about the camouflage of animals," "how animals stay alive in nature, including strategies and solutions," "secrets behind how animals manage to stay alive," "survival strategies in different animals," "camouflage of sea animals," and "I have learned that animals can be sneaky, too". Some participants mentioned the details they remembered from the documentary, for example:

- "octopuses are skillful in camouflage,"
- "two snakes that looked similar while one of them was venomous and the other was not,"
- "The part about snakes and the effect of zebra stripes on its predators was very interesting to me."
- "I learned new information about the tricks used by squirrels, and I learned that there are other animals which are better at camouflage than the chameleon"
- "I have learned that the camouflage in animals is dependent on a combination of their bodily features and color."

Several participants mentioned that they enjoyed the pictures and the visual effects of the documentary. One participant reported that because the presenter used casual language to speak, he was encouraged to listen to the content more carefully.

The questionnaire item "What have you learned in terms of language?" elicited illuminating answers. Two participants mentioned that although they could not remember the words and expressions from the video, they had picked up some items subconsciously. A few participants generally stated that they learned new content and new words and expressions without mentioning any examples. However, many participants named expressions and words they had learned from the documentary.

In the first group, in which the students watched the video with captions that were highlighted for target MWIs, one participant mentioned that although she comprehended the gist of the message, she did not learn the meaning of words and expressions because we were not allowed to look them up. The same participant stated that because they were tested on the words and expressions before, she was familiar with words and expressions and understood

them in context, but was not able to remember their meanings. This answer is enlightening as the participant indicated her awareness about the importance of dictionary look-up and the effect of pretesting on later treatment. A participant from the third group, who watched the video without captions, said, "If I could, I would like to re-watch the video several times to take notes of important words and expressions and expand my vocabulary". Another participant from the fourth group, who watched the video with normal captions, mentioned, "I remember venomous, predator, and gecko because they were repeated several times throughout the video". This response is interesting because it not only suggests that the frequency of occurrence may boost vocabulary learning from captioned viewing but also shows that some learners pay attention to repetitions consciously.

As for the questionnaire item about whether or not participants wanted to re-watch the video, several participants mentioned that now that they understood the content of the video, they wanted to re-watch the video to learn new words and expressions, while others claimed that they wanted to re-watch it because they found the content interesting and wanted to pay attention to details as they re-watched the video. Still, others mentioned that because they understood all the content, they preferred not to re-watch the video. Several participants mentioned that they did not like to re-watch the video because if one understands the topics introduced in a video, re-watching that video would be a boring activity. One participant claimed that "it is not necessary to re-watch what I have understood, but if I intend to learn its vocabulary, re-watching would be a good idea". These answers are valuable as they suggest that there may be individual differences between students regarding how they view a repeated activity.

Discussion

The primary purpose of the current study was to examine the effect of captioning and TE of multiword items and single words on Iranian EFL learners' viewing comprehension. The secondary purpose of this study was to gain additional insights into the students' viewing experience. The descriptive statistics demonstrated that the performance of the four groups on the comprehension test was as follows: Group 1 > Group 2 > Group 3 > Group 4.

Regarding the first research question, the ANOVA results revealed a significant difference between the comprehension scores of the four groups. Post-hoc analyses revealed that Group 1 and Group 2 significantly outperformed Group 4 on the comprehension test. This finding suggests that in this study, TE of captions had a positive impact on students' viewing comprehension. The significant impact of caption condition on viewing

comprehension scores differs from Montero Perez et al. (2014a) and Hsieh (2020) but is similar to Majuddin et al. (2021). It appears that in the current study, the presence of TE motivated learners to pay closer attention to the video content.

Interestingly, rather than distracting from the overall message, TE of lexical forms appears to have supported comprehension. From the perspective of Mayer's (2021) Cognitive Theory of Multimedia Learning, this outcome implies that the additional visual salience introduced by TE did not exceed learners' available cognitive capacity. Rather than imposing extraneous cognitive load, which would be expected to impair learning by overwhelming the verbal and visual processing channels, TE appears to have guided learners toward meaningful selection and organization of relevant linguistic and content information. This suggests that when TE is applied judiciously, it can assist learners in integrating verbal input with the visual and auditory channels of the video without causing cognitive overload.

As for TE of single words, the significantly greater viewing comprehension of Group 2 compared to Group 4 contradicts Hsieh's (2020) study, in which no significant difference was observed between the viewing comprehension scores of the single-word enhanced captioned viewing and unenhanced captioned viewing conditions. Interestingly, even the descriptive statistics of that study and the current study do not align, because in that study, unenhanced fully captioned viewing led to higher scores than both uncaptioned and single-word enhanced captioned viewing conditions.

A theoretical account of why single-word TE aided comprehension in the present study, but not in Hsieh's (2020), may relate to the interaction between noticing and learner proficiency. For higher-proficiency learners, as in the present sample, the act of noticing a highlighted word may require relatively less cognitive effort, freeing up working memory resources for simultaneous meaning construction. This is consistent with Mayer's (2021) theory, which posits that meaningful learning depends on the learner's ability to actively select and integrate relevant information without exceeding processing capacity. For lower-proficiency learners, conversely, the same TE might generate extraneous cognitive load by pulling attention away from the overall message at a point when available cognitive resources are already heavily engaged in basic decoding.

With respect to MWIs, the significantly higher viewing comprehension scores of Group 1 compared to Group 4 are in contrast with Majuddin et al. (2021), who found no significant disparity between the comprehension scores of the MWI-enhanced captioned viewing and the uncaptioned viewing conditions. However, this result receives partial support from one of the findings of Majuddin et al. (2021). According to that finding, when watching the video twice,

the MWI-enhanced captioned viewing group non-significantly outscored the unenhanced captioned group in terms of viewing comprehension score.

A plausible theoretical explanation for why TE aided comprehension, rather than distracting from it, may lie in the Noticing Hypothesis (Schmidt, 1990) and its interaction with learners' proficiency level. For high proficiency learners (as in the current study), the additional processing load imposed by TE may be within cognitive capacity, allowing them to notice highlighted items without sacrificing comprehension of the overall message. The present results, therefore, suggest that TE may be beneficial specifically when learner proficiency is sufficient to accommodate the dual demands of comprehension and form-focused attention.

The lack of a significant difference between the comprehension score of the uncaptioned viewing group compared to the unenhanced captioned viewing group aligns with Montero Perez et al. (2014b) and Hsieh (2020) but differs from Majuddin et al. (2021), who observed that learners in the unenhanced captioned viewing condition significantly outperformed those in the uncaptioned condition. Moreover, this finding is in contrast with Montero Perez et al. (2013), who reported that captioned viewing had a significantly large effect size on listening comprehension.

Although not reaching statistical significance, the higher mean comprehension score of the uncaptioned viewing group compared to the unenhanced captioned viewing group is similar to Sydorenko (2010), whose findings indicated that uncaptioned viewing developed listening comprehension. However, this finding conflicts with Winke et al. (2010) and Pujadas and Muñoz (2024), who found that captioned viewing led to superior comprehension scores compared to uncaptioned viewing. One plausible explanation for this finding can be the high proficiency level of the participants. As stated earlier, Pujadas and Muñoz (2024) discovered that learner reliance on captions can vary according to their proficiency level. That is, lower-level learners may rely on captions for listening comprehension more than higher-level learners. Additionally, some research findings indicate that captions may cause frustration or hinder comprehension in higher-level learners (Leveridge & Yang, 2013).

The lack of a significant difference between the comprehension score of Group 3 and that of groups 1 and 2 is in line with Majuddin et al. (2021) and Hsieh (2020), respectively. Finally, our questionnaire data indicate that a great proportion of participants regularly watched English-language videos either with or without captions/subtitles, a finding that confirms the popularity of L2 audio-visual materials among Iranian English-major EFL learners. The finding that most participants used L2 captions while watching English videos (41.4%) is consistent

with Ahrabi Fakhr et al. (2021), who found that Iranian EFL students watched English-captioned videos for almost three hours each week. However, in that study, students reported watching uncaptioned English videos more than L1-subtitled videos. This pattern differs from the viewing habits of the participants in the current study, where more participants reported using Persian subtitles (24.1%) compared to no captions/subtitles (20.7%).

Another interesting insight from the questionnaire is that while research suggests repeated viewing can enhance comprehension (Majuddin et al., 2021), not all individuals enjoy this repetition; therefore, personal preferences should be taken into consideration. Instead of watching the same video repeatedly, students could be encouraged to watch TV series or YouTube vlogs with similar topics. This way, relevant content is repeated without causing learners unnecessary boredom and frustration.

Conclusions and Implications

Given the inconsistent findings of previous studies regarding the effect of TE and caption condition on L2 viewing comprehension, the present study aimed to clarify the effect of four captioning conditions on Iranian EFL learners' viewing comprehension. It contributes to a growing body of research on L2 viewing comprehension by demonstrating that TE of lexical items in captions does not necessarily compromise, and can in fact support, viewing comprehension, at least among high-proficiency EFL learners. This finding reframes a concern that has persisted in the literature (e.g., Hsieh, 2020; Majuddin et al., 2021), namely that drawing attention to lexical items may come at the cost of meaning comprehension. The results suggest this trade-off is not inevitable and that proficiency may act as a moderating variable. Another notable finding was that unenhanced captioned and uncaptioned viewing did not lead to significantly different comprehension outcomes, a finding that challenges the assumption that captions always aid viewing comprehension. Finally, the questionnaire results indicated that participants were used to watching English videos and most of them (41.4%) watched English videos with L2 captions.

The findings of the present investigation have some pedagogical implications. First, the presence of TE in captioned videos to draw learners' attention to certain lexical items does not necessarily lead to a reduction in content comprehension. Conversely, it may positively impact their comprehension if students are of high language proficiency. Therefore, provided that learners' L2 proficiency is high, teachers and materials developers can textually enhance target lexical items in the captions without worrying about a trade-off between attention to textually-enhanced items and video comprehension. Second, the lack of a significant

difference in the mean comprehension scores of the unenhanced captioned viewing group and the uncaptioned group indicates that L2 teachers can encourage high-proficiency learners to turn the captions off while watching English videos, so that the learners become more independent and more confident viewers.

Like any study, this investigation had several limitations. One limitation was the fact that, because of using intact classes, the study had a quasi-experimental, rather than true experimental, design. In addition, although the sample size (n = 62 across four groups) was moderate, it was not particularly large for between-group comparisons, which may have limited the statistical power to detect smaller effects. Another limitation was that, due to time restrictions, it was not extended longitudinally. Furthermore, participants consisted of Iranian university students from a single university. To be able to generalize the results, future studies involving learners from other academic levels and cultural backgrounds are required because the comprehension patterns for other learner groups might be different. Finally, eye-tracking data are required to provide us with helpful insights into the way learners process textually-enhanced captions.

Acknowledgements

We extend our sincere gratitude to the students and colleagues who provided invaluable assistance during the data collection phase.

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

Allan, D. (2004). *Oxford Placement Test 1.* Oxford University Press.

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), 136–140. https://doi.org/10.1016/j.cub.2009.12.014

Choi, Y., & Zhang, D. (2021). The relative role of vocabulary and grammatical knowledge in L2 reading comprehension: A systematic review of literature. *International Review of Applied Linguistics in Language Teaching*, *59*(1), 1–30. https://doi.org/10.1515/iral-2017-0033

Della Putta, P. (2019). Promoting learning and unlearning through textual enhancement in a closely related L1-L2 relationship: The results of a bidirectional study with Spanish-

speaking students of Italian and Italian-speaking students of Spanish. *European Journal of Applied Linguistics*, *7*(1), 1–30. https://doi.org/10.1515/eujal-2018-0005

Du, G., & Man, D. (2022). Person factors and strategic processing in L2 listening comprehension: Examining the role of vocabulary size, metacognitive knowledge, self-efficacy, and strategy use. *System*, *107*, 102801. https://doi.org/10.1016/j.system.2022.102801

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

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2023). 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, 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

Hui, B. (2024). Scaffolding comprehension with reading while listening and the role of reading speed and text complexity. *The Modern Language Journal*, *108*(1), 183–200. https://doi.org/10.1111/modl.12905

Jahan, A., & Kormos, J. (2015). The impact of textual enhancement on EFL learners' grammatical awareness of future plans and intentions. *International Journal of Applied Linguistics*, *25*(1), 46–66. https://doi.org/10.1111/ijal.12049

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. (2007). Preschool and early elementary comprehension: Skill development and strategy interventions. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 27-45). Lawrence Erlbaum Associates.

LaBrozzi, R. M. (2016). The effects of textual enhancement type on L2 form recognition and reading comprehension in Spanish. *Language Teaching Research*, *20*(1), 75–91. https://doi.org/10.1177/1362168814561903

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

Lee, C. E., Godwin, H. J., Blythe, H. I., & Drieghe, D. (2025). Individual differences in skilled reading and the word frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 51*(8), 1281–1302. https://doi.org/10.1037/xlm0001428

Lee, M., & Jung, J. (2024). Effects of textual enhancement and task manipulation on L2 learners' attentional processes and grammatical knowledge development: A mixed methods study. *Language Teaching Research*, *28*(4), 1552–1571. https://doi.org/10.1177/13621688211034640

Lee, M., & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: An eye-tracking study. *The Modern Language Journal*, *102*(3), 557–577. https://doi.org/10.1111/modl.12503

Lee, S. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, *57*(1), 87–118. https://doi.org/10.1111/j.1467-9922.2007.00400.x

Leveridge, A. N., & Yang, J. C. (2013). Testing learner reliance on caption supports in second language listening comprehension multimedia environments. *ReCALL*, *25*(2), 199–214. https://doi.org/10.1017/S0958344013000074

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

Meguro, Y. (2019). Textual enhancement, grammar learning, reading comprehension, and tag questions. *Language Teaching Research*, *23*(1), 58–77. https://doi.org/10.1177/1362168817714277

Mohsen, M. A., & Mahdi, H. S. (2021). Partial versus full captioning mode to improve L2 vocabulary acquisition in a mobile-assisted language learning setting: Words pronunciation domain. *Journal of Computing in Higher Education*, *33*(2), 524–543. https://doi.org/10.1007/s12528-021-09276-0

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

Montero Perez, M., Peters, E., & Desmet, P. (2014b). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL*, *26*(1), 21–43. https://doi.org/10.1017/S0958344013000256

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*(3), 720–739. https://doi.org/10.1016/j.system.2013.07.013

Muñoz, C., Pujadas, G., & Pattemore, A. (2023). Audio-visual input for learning L2 vocabulary and grammatical constructions. *Second Language Research*, *39*(1), 13-37. https://doi.org/10.1177/02676583211015797

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

Pallant, J. (2016). *SPSS survival manual* (6th ed.). Open University Press.

Pattemore, A. (2022). *Learning grammatical constructions from audio-visual input* [Phd thesis]. https://doi.org/10.13140/RG.2.2.36540.72324

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition*, *42*(3), 577–598. https://doi.org/10.1017/S0272263120000091

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

Puimege, 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. (2020). Examining adolescent EFL learners' TV viewing comprehension through captions and subtitles. *Studies in Second Language Acquisition*, *42*(3), 551–575. https://doi.org/10.1017/S0272263120000042

Pujadas, G., & Muñoz, C. (2024). When to switch captions off? Exploring the effects of L2 proficiency and vocabulary knowledge on comprehension of captioned and

uncaptioned TV. *Studies in Second Language Learning and Teaching*. https://doi.org/10.14746/ssllt.38036

Rassaei, E. (2015). Effects of textual enhancement and input enrichment on L2 development. *TESOL Journal*, *6*(2), 281–301. https://doi.org/10.1002/tesj.149

Roohani, A., Rahimi Domakani, M., & Alikhani, F. (2013). The effects of captioning texts and caption ordering on L2 listening comprehension and vocabulary learning. *Applied Research on English Language*, *2*(2), 51-64. https://doi.org/10.22108/are.2013.15470

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

Schroeders, U., Wilhelm, O., & Bucholtz, N. (2010). Reading, listening, and viewing comprehension in English as a foreign language: One or more constructs?. *Intelligence*, *38*(6), 562-573. https://doi.org/10.1016/j.intell.2010.09.003

Serrano, R., & Pellicer-Sánchez, A. (2022). Young L2 learners' online processing of information in a graded reader during reading-only and reading-while-listening conditions: A study of eye-movements. *Applied Linguistics Review*, *13*(1), 49–70. https://doi.org/10.1515/applirev-2018-0102

Sidorov, O., Hu, R., Rohrbach, M., Singh, A. (2020). TextCaps: A Dataset for Image Captioning with Reading Comprehension. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), *Computer Vision – ECCV 2020* (Lecture Notes in Computer Science, Vol. 12347, pp. 742–758). Springer. https://doi.org/10.1007/978-3-030-58536-5_44

Simard, D. (2009). Differential effects of textual enhancement formats on intake. *System*, *37*(1), 124–135. https://doi.org/10.1016/j.system.2008.06.005 Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73.

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. (2025). Incidental vocabulary learning from captioned video genres: vocabulary knowledge, comprehension, repetition, and working memory. *Computer Assisted Language Learning*, *38*(5-6), 1301-1340. https://doi.org/10.1080/09588221.2023.2275158

Tong, Y., Hasim, Z., & Abdul Halim, H. (2023). The relationship between L2 vocabulary knowledge and reading proficiency: The moderating effects of vocabulary fluency. *Humanities and Social Sciences Communications*, *10*(1), 555. https://doi.org/10.1057/s41599-023-02058-2

Vafaee, P., & Suzuki, Y. (2020). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition*, *42*(2), 383–410. https://doi.org/10.1017/S0272263119000676

van Zeeland, H. & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457–479. https://doi.org/10.1093/applin/ams074

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

Vu, D. V., Noreillie, A.-S., & Peters, E. (2023). Incidental collocation learning from reading-while-listening and captioned TV viewing and predictors of learning gains. *Language Teaching Research*, 136216882211510. https://doi.org/10.1177/13621688221151048

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

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 65–86.

Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, *26*(4), 696–725. https://doi.org/10.1177/1362168820913998