Language Learning & Technology
ISSN 1094-3501 BY-NC-ND

June 2021, Volume 25, Issue 2
pp. 178–191

ARTICLE

The effects of captions on L2 learners' comprehension of vlogs

Dukhayel Aldukhayel, *Qassim University*

Abstract

This study investigated the effects of captions on the listening comprehension of vlogs. A total of 96 EFL learners watched three vlogs under one of three conditions: L2 captions, L1 captions, and no captions. Each group included low-, mid-, and high-level proficiency learners. The vlogs differed in the pictorial support of the audio, with Vlog 1 being highly supported, Vlog 2 being partially supported, and Vlog 3 being slightly supported by pictorial images. After each vlog, the participants took a multiple-choice test measuring their comprehension of details. Afterwards, participants completed a questionnaire about their perception of captions. The findings suggest that the availability of captions may not necessarily lead to better listening comprehension because students, particularly lower proficiency learners, were unable to simultaneously process the multiple modalities (images, audio, and captions) due to their limited capacities of working memory and cognitive load. High-proficiency learners achieved better comprehension than low- and mid-proficiency learners and achieved their best comprehension with L2 captions. A significant increase in comprehension of vlogs caused by high pictorial support was detected, with the inverse relationship also being true. Analysis of the questionnaire indicated that participants consider L2 captions useful. For both L2 and L1 captions, students think that their listening comprehension would decrease without captions. When considering vlogs for L2 listening, language proficiency and pictorial support are better indicators of levels of comprehension. Captions might be beneficial when learners' proficiency level is high. When visual images are highly supportive for the audio, better comprehension of vlogs is likely.

Keywords: *L2 Listening, Captions, YouTube, Vlog*

Language(s) Learned in This Study: *English*

APA Citation: Aldukhayel, D. (2021). The effects of captions on L2 learners' comprehension of vlogs. Language Learning & Technology, 25(2), 178–191. http://hdl.handle.net/10125/73439

Introduction

Language learning demands a suitable and sufficient input, whether it is written or spoken. The input should be authentic and comprehensible. However, providing sufficient aural input to meet these conditions might be a challenge in the EFL context (Rodgers & Webb, 2017; Vanderplank, 2016). Utilizing authentic videos, such as episodes of well-known TV programs, in L2 classes is helpful because the visual images in videos provide authenticity to real-life listening that is lacking in aurally-based texts (Yeldham, 2018). In addition to the traditionally utilized video genres in L2 classes, such as television programs, news clips, movies, and the like, vlogs intended for an English-speaking audience may also be a potential source of L2 aural input (Watkins & Wilkins, 2011). Vlogging is a new genre that has proliferated in recent years due to the Internet and technological advancements, Wood (2019) confirms. A *vlog* is short for "video blog," while *blog* is already a blended word (from "web log"). Snelson (2015) points out that vlogs have become more interesting and have shifted from home-based settings to mobile settings, in which vloggers discuss a wide variety of topics with the presence of more cameras in a wide array of settings and contexts. Snelson adds that vloggers began to tap into wider audiences' interest areas to document and share a contextualized view of their engagement with hobbies, daily experiences, and encounters with other people.

Considering their distinguished contextual and interactional parameters, vlogs have been identified as a recognizable "new genre" by L2 researchers (e.g., Frobenius, 2011, 2014; Werner, 2012; Wood, 2019). The most comprehensive—and possibly only—analysis of vlogs as a recognizable genre was performed by Wood (2019). Applying the speaking model by the linguistic anthropologist Hymes (1974) within a genre analysis, Wood qualitatively analyzed the contextual and interactional parameters of a number of representative vlogs in order to discover what differentiates vlogs from other video genres. She found that vlogs are distinguished in their settings, participants, goal of vlogging, act sequence of vlogging, key (the overall mode or tone), vlogging instrumentalities, and norms and genres of the vlog. It is outside the scope of this study to detail each of these features; however, Wood concludes, "it is evident that 21st century individuals who spend much of their life in the online sphere can and do recognize these digital genres" (The Vlog as a New Form of Interaction section, para. 2).

However, whether vlogs serve as comprehensible input for learners of all proficiency levels is a concern. One feature of YouTube that might make vlogs comprehensible for less proficient learners is closed captions and subtitles. Captions refer to transcriptions of the dialogue in the language of the medium, whereas subtitles refer to transcriptions or translations of the dialogue in the viewer's first language (Markham et al., 2001; Vanderplank, 2016). However, since the two terms are often used interchangeably, we will refer to both as *captions* in this paper. The aim of this study is to explore the potential effectiveness of captions on listening comprehension of English language vlogs across three different conditions: L2 captions (English), L1 captions (Arabic), and no captions.

Literature Review

Captioning studies have differed largely with regard to their design of treatment conditions, but there are two general types of studies. There are those studies that (a) compared full captions with keyword captions and/or no captions (e.g., Guillory, 1998; Montero Perez et al., 2014a, 2014b; Rodgers & Webb, 2017; Teng, 2019; Winke et al., 2010, 2013) and those that (2) compared the L2 captions with the L1 captions and/or no captions (e.g., Markham & Peter, 2003; Markham et al., 2001; Hayati & Mohmedi, 2011). Aside from differences in the treatment conditions, captioning studies also differed in participants' L1 backgrounds, proficiency levels, age, type and duration of watching materials, measurement instruments and questions, and so forth. Due to these differences in research design, the benefit of captions for facilitating listening comprehension of videos was not always supported, and the studies have also presented contradicting results, sometimes for no obvious reasons. These differences will be examined in detail in the following sections.

One of the early studies that showed the benefit of full and keyword captioning came from Guillory (1998), who compared the influence of full captions, keyword captions, and no captions on French language learners' listening comprehension of two educational videos. Keyword captions are words identified as being important to the video. To measure comprehension, participants (n = 202) in all treatment groups completed two short-answer comprehension tests, each consisting of seven questions on the recall of details and inferencing from the information presented in the videos. The results showed that the full captions group (M = 10.1) and the keyword captions group (M = 9.2) outperformed the no captions group (M = 7.3), but there was no significant difference between the two caption groups.

The benefit of full and keyword captioning was, in some cases, totally unobserved, such as in Montero et al. (2014a). They used three short authentic clips to explore the effects of captions on French language learners' (n = 133) listening comprehension under one of four conditions: full captions, keyword captions, full captions with highlighted keywords, and no captions. Three comprehension tests were used and included 41 items: 19 short open-ended questions, 14 true-false items, and eight combination items. The results revealed that participants in all conditions achieved similar scores on the comprehension tests. Montero Perez et al. explained that the results could be attributed to the type of information targeted by the comprehension questions. They also indicated that the nature of the clips, which presented concrete factual information, did not allow for asking inferencing questions, which are considered more challenging.

In other cases, the benefit of full and keyword captioning depended on the type of questions, such as in the case of global comprehension versus comprehension of details. For example, see Montero et al. (2014b), who investigated the effect of full captions and keyword captions on the listening comprehension of three short French videos by French language learners (*n* = 226). Listening comprehension was measured by a comprehension test, which consisted of questions on global understanding and on details. Findings revealed that the full caption group outperformed both the keyword and the no caption group on the global comprehension questions. However, no difference was found between the keyword caption and the no caption group. Results of the detail comprehension questions revealed no differences among the three conditions.

Other studies showed that the benefit of full and keyword captioning was affected by both the type of questions and level of proficiency such as Teng (2019). Teng's participants (*n* = 182), who were primary school ESL learners, watched two short story videos in one of three conditions: full captions, keyword captions, and no captions. Each group included learners with higher and lower proficiency levels. Two comprehension tests were developed for each video: a written recall protocol instrument measuring global comprehension and a 20-item multiple-choice test measuring detail comprehension. Regarding global comprehension, findings revealed that the treatment with full captions produced significantly higher scores than the version with keyword captions, and the treatment with keyword captions produced significantly higher scores than the treatment without captions, regardless of proficiency level. With regards to comprehension of details, the same results only occurred for the high-proficiency learners. For low-proficiency learners, differences were significant only between the keyword treatment and the no caption treatment. Learners with high proficiency levels scored significantly higher than those with low proficiency levels, irrespective of the captioning condition.

The benefit of captions might decrease or disappear entirely in the later stages of long experiments, as Rodgers and Webb (2017) found. The authors focused on L2 television programs, rather than the short videos often used in previous studies. They investigated the listening comprehension of 372 Japanese university students who watched ten 42-minute episodes of an American television program in two conditions: L2 captions and no captions. The measures used were true/false items, multiple-choice items, and sequencing items, which were designed to measure detail comprehension, inferencing ability, and global comprehension. Results indicated that the learners who watched captioned videos had significantly higher comprehension scores for only three of the ten episodes. By the tenth episode, the learners in the no captions group benefited from their accumulated knowledge to reach a level where the potential support from the captions did not produce a substantial difference in comprehension between the two groups. The authors concluded that captions are likely to significantly aid comprehension, particularly when the episode is comparatively difficult.

Other studies have compared the benefits of L1 and L2 captions. Some of those studies revealed the advantage of the L1 captions over L2 captions, such as Markham et al. (2001) and Markham and Peter (2003), who compared the effects of captions on Spanish language learners' listening comprehension of a documentary in one of three treatment conditions: L2 (Spanish) captions, L1 (English) captions, or no captions. In the 2001 study, participants' (*n* = 169) comprehension was measured with a written summary and 10 multiple-choice items. In the 2003 study, a 20-item multiple-choice test was used to measure learners' (*n* = 213) comprehension. The two studies' findings were the same. They revealed that learners in the L1 caption groups outperformed those in the L2 captions group and dramatically outperformed those in the no caption groups on all tests. In contrast, other studies found an advantage for the L2 captions over L1 captions, such as Hayati and Mohmedi (2011), who examined the influence of L2 (English) captions, L1 (Persian) captions, and no captions on the comprehension of six parts of a movie. EFL Iranian learners' (*n* = 90) comprehension was measured via six sets of multiple-choice tests. However, their results revealed that the L2 caption group performed significantly higher than the L1 caption group, which in turn performed significantly higher than the no caption group on the listening tests. The authors speculated that L1 captions involve translation, requiring learners to switch from one language to another, which takes time, so learners may lag behind and lose track.

Another line of research that informed this study concerns orthographic differences between learners' native language and the target language. Two studies highlight these differences: Winke et al. (2010, 2013). These two studies investigated the behavior of reading captions by English-speaking learners of Arabic, Chinese, Russian, and Spanish who watched videos with L2 captions. The studies found that the Arabic- and Chinese-language learners spent significantly less time reading captions in the video than the Spanish and Russian learners. The authors concluded that multiple input modalities (images, audio, and captions) may be understood differently depending, in part, on orthographical differences between the L1 and L2 captions. Learners of a language whose orthography is closer to that of the target language may be better able to use the written modality as an initial source of information. Thus, the researchers also concluded that when there is a great distance between the L1 and L2 orthographies, learners may have to rely on listening more because the written symbols have not yet been mastered. Their conclusions were based on the dual-processing theory of working memory (Baddeley, 2007; Mayer & Moreno, 1998), which suggests that people have limited cognitive resources for processing information, as well as on the cognitive load theory, which suggests that when listeners' attention is split between two visual input modalities (captions and visual images) to infer meaning, their limited visual working memory capacity will be overburdened (Chandler & Sweller, 1991).

Overall, several gaps can be identified in captioning research, particularly in the studies mentioned above. First, the listening comprehension advantage of captions has been examined solely when viewing traditional video genres, such as television programs, news clips, movies, documentaries, or story videos. However, the question of whether the influence of captions can be transferred to vlogs, as a recognized online video genre, remains unanswered, although vlogs are the most popular video genre on YouTube (Werner, 2012). While vlogs "have attracted researchers from social, communicative and linguistic perspectives" (Wood, 2019, Studies About the Vlogging Phenomena section, para. 1), there remains scant scholarly research on vlogs in the field of second language acquisition. Vanderplank (2016) points out the accessibility of online audiovisual materials on YouTube and other Internet sources, stressing that they "may well have an impact on the practices of teachers, class-based learners and informal learners, and a future study would be needed to see this effect" (p. 230).

The second concern in captioning research is that it lacks investigation of the influence of pictorial support on listening comprehension, particularly some studies (e.g., Durbahn et al., 2020) found that the presence of imagery plays a role in comprehension of videos. The level of audio/video correlation and how much the audio input is supported and contextualized by the visual images in the video (Garza, 1991), and its relationship with captions and language proficiency, have not been investigated in the above-reviewed studies. Indeed, Rodgers and Webb (2017) speculated that other factors might affect listening comprehension, including audio/video correlation, and commented that it could be the impetus for future research on exactly how and to what extent the comprehension of television is influenced by these factors and how the presence of captions affects this relationship.

The third concern is that, to the best of our knowledge, only two studies have investigated the differential effects of captions across languages, namely Winke et al. (2010, 2013). Bearing in mind the differences between Arabic and English orthographies, the extent to which Arabic-speaking learners of English can read L2 captions remains unanswered. Hence, the present study seeks to address these research gaps by examining the effects of three captioning types—L2 captions, L1 captions, and no captions—on the listening comprehension of vlogs by Arab EFL learners. To enhance knowledge in this area, two variables (i.e., learners' English proficiency and pictorial support) were also examined. To this end, the following research questions were posed:

1. Do L2 captions or L1 captions result in better listening comprehension of vlogs for Arab EFL learners?
2. To what extent do English proficiency levels affect listening comprehension of vlogs, and does the captioning condition lead to better performance for any proficiency levels?
3. To what extent does pictorial support affect the listening comprehension of vlogs?

Method

Participants

The participants were 96 Arabic-speaking university male students (M_{age}= 17.38 years, *SD* = .64). They were enrolled in a mandatory English program that focused on basic skills in reading, writing, speaking, and listening. Participants who had studied English for a minimum of seven years were enrolled in three different classrooms. Each classroom was assigned to one of three groups: L2 Captions Group (L2CG; *n* = 35), L1 Captions Group (L1CG; *n* = 32), and No Captions Group (NCG; *n* = 29). Given the importance of vocabulary size for video understanding (Webb & Rodgers, 2009), we included a vocabulary size test rather than a general proficiency test (Montero Perez et al., 2014b). As an indicator of English proficiency, the participants were administered the Vocabulary Levels Test (VLT; Schmitt et al., 2001) at the 2,000-, 3,000-, and 5,000-word levels. The total possible score for all three sections of the test is 90 marks (30 marks for each section). Accordingly, a score between 0–30 was considered a low proficiency level, 30–60 was considered a medium proficiency level, and 60–90 was considered a high proficiency level. According to the VLT results, each group included low (*M* = 26.86, *SD* = 1.3), mid (*M* = 45.85, *SD* = 1.72), and high (*M* = 64.28, *SD* = 2.05) proficiency students (See Table 1). ANOVA (*F*(2, 93) = .248, *p* = .78) indicated no significant difference among the combined scores of these tests for the L2CG (*M* = 43.34, *SD* = 13.47), L1CG (*M* = 45.21, *SD* = 14.02), and NCG (*M* = 43.27, *SD* = 14.03).

Table 1

Participants' Proficiency Level by Group

Proficiency level	L2CG	L1CG	NCG	Total
Low	11	9	10	30
Mid	17	15	13	45
High	7	8	6	21

Materials

Three YouTube vlogs of different pictorial support were selected for the three experimental groups. Vlog 1 (6'54'') was *Living on $1 For 24 Hours in NYC! (Day #1)* by LivingBobby (2018); Vlog 2 (6'52'') was *How to spend 1 day in Lucerne | Switzerland Travel Vlog* by One Girl One Suitcase (2015); Vlog 3 (5'13'') was *Visit Russia - The DON'Ts of Visiting Russia* by Wolters World (2017; the original form of the vlog titles was preserved). As research shows that 90% lexical coverage (known words in a text) is likely to be enough for listening/viewing comprehension of informal texts (Durbahn et al., 2020; Noreillie et al., 2018; Van Zeeland & Schmitt, 2012), all vlogs reached 90% coverage at the 1,000 word-level, as measured by Cobb's Lextutor's VocabProfile. This assured that the input would not be difficult for the participants of all levels of proficiency to understand. Three experienced English language teachers were consulted in the selection of the vlogs. The teachers used the following linguistic and paralinguistic criteria when selecting the vlogs: situational appropriateness, grammatical and lexical complexity, and inherent interest value to university-level students (Garza, 1991).

Because captions are automatically generated by YouTube, the teachers—who were native Arabic speakers—were asked to check the accuracy of the L2 and L1 captions of the three vlogs and decide how much the captions matched the dialogue in the vlogs. All the raters decided that both the L2 and L1 captions matched 95% or more of the vlogs' dialogue.

Teachers were also asked to rate the vlogs in terms of the pictorial support of the vlog (i.e., how much the audio input is supported and contextualized by the visual images in the vlog). The method established for determining pictorial support was based on how much of the vlog duration included visual elements/images that instantly/directly support, contextualize, or match the audio. By consulting with the teachers, it was determined that pictorial support occurring during 66% or more of the vlog time is considered high, between 33% and 65% of the vlog time is considered partial, and pictorial support during 32% or less of the vlog time is considered slight. The teachers reached a consensus on the pictorial support of the vlogs ultimately chosen for the study. They regarded the pictorial images as highly supportive of Vlog 1, partially supportive of Vlog 2, and slightly supportive of Vlog 3. They also reached a consensus on the difficulty level and appropriateness of the vlogs that were chosen for the study.

Listening Comprehension Tests and Scoring

In the present study, listening comprehension is defined as understanding the meaning of spoken words while watching videos. A multiple-choice test was developed for each vlog. Each test included 10 questions designed to measure participants' listening comprehension of details in the vlogs. Tests were administered immediately after vlog watching, and participants were asked to choose the best answer from four options for each question. The questions were carefully developed so they could not be answered directly without a participant understanding the aural content of the vlogs. The multiple-choice test format seemed to be particularly appropriate in this study, as the low- and mid-level students clearly lacked highly developed English language writing ability at this point in their study of the target language (Markham & Peter, 2003).

Following the research procedures of other captioning studies (e.g., Markham & Peter, 2003; Rodgers & Webb, 2017), the study's tests were translated into the participants' L1 (Arabic), taking into consideration Buck's (2001) suggestion that "if the test-takers are all from the same L1 background, giving the questions in the first language works very well and is probably a good way of ensuring that listening test scores are not contaminated with other skills" (p. 143). Translation was done by the author in consultation with the three teachers. Cross-referencing between the tests' materials and the Arabic captions of all three vlogs was performed to assure that the tests would not be prejudicial in favor of the L1CG.

To determine the suitability of each vlog, a pilot study was conducted with 30 students from the same program with similar language-learning backgrounds and similar English proficiency levels. An item analysis was performed for each of the pilot-tested multiple-choice items, and problematic items were rewritten, altered, or deleted. Prior to the item-analysis procedure, each comprehension test had 15 multiple-choice items. Following the item-analysis procedure, the comprehension tests were distilled to 10 items. Pilot participants expressed that the text and item difficulty levels were appropriate. The learners also selected the correct multiple-choice answer at least 70% of the time at the lower end of the test performance range after viewing the vlog. Also, pilot students were asked to rate the vlogs in terms of pictorial support, and their ratings were similar to the teachers'.

The scoring system for the tests allotted one point for each correct answer, with a possible total score of 10. A meeting was held with the three teachers and the author in which all agreed that 10 items could adequately reflect the vlogs' content. The tests were scored by the author and two teachers.

Questionnaire

Participants completed a mini questionnaire consisting of a five-point Likert scale, including statements on the usefulness of L2 and L1 captions (e.g., ''My comprehension improved because of the English/Arabic captions'') and questions about their caption needs (e.g., ''My comprehension would be better with the Arabic/English captions''). The L2CG and the L1CG responded to three items, while the NCG responded to four items. The questionnaire data was used to clarify the findings of the quantitative research.

Procedures

In the laboratory, participants were told to watch the vlogs twice, adhering to the tradition of playing a text at least twice (Field, 2008). Each student worked individually at a computer with a headset. The L2CG and L1CG were instructed on how to view captions, while the NCG was told to watch without any captions. Although less proficient learners may have needed to watch a vlog several times or slow down the speech rate to enhance comprehension, they were not allowed to change the speed or replay the vlog, which ensured that the researcher could examine learners' proficiency levels and their relationship with the treatment effects. The author and two teachers monitored the participants to ensure that they followed the procedure. Once they finished watching a vlog, participants completed a paper-and-pencil multiple-choice comprehension test. The time for each test, as determined through the pilot study, was 15 minutes, and the experiment took approximately one hour to complete. At the end of the experiment, participants filled out the questionnaire.

Results

Research Question 1

To determine whether there were any statistically significant differences among the three groups (L2CG, L1CG, and NCG) on the comprehension scores of the vlogs, a multivariate analysis of variance (MANOVA) was conducted due to there being three dependent measures (Vlog 1, Vlog 2, and Vlog 3) and one independent measure with three levels. The MANOVA demonstrated that captions did not produce significantly different outcomes, Wilks' $\Lambda = 0.98$, $F(6, 182) = .32$, $p = .93$, multivariate η2 = .010. As can be seen in Table 2, the three groups have similar means for each vlog.

Table 2

Descriptive Statistics of Comprehension Scores by Treatment Condition

	L2CG's mean (*SD*)	**L1CG's mean (*SD*)**	**NCG's mean (*SD*)**
Vlog 1	7.91 (2.07)	7.68 (1.20)	7.86 (2.06)
Vlog 2	6.11 (2.16)	5.43 (2.50)	5.72 (2.32)
Vlog 3	5.34 (1.97)	5.09 (1.89)	5.00 (1.81)

Note. Possible score for each vlog = 10 points

Research Question 2

Another MANOVA was conducted to determine whether there were any statistically significant differences among the three proficiency levels on the comprehension scores of the vlogs. The MANOVA demonstrated that the different proficiency levels produced significantly different outcomes (Wilks' $\Lambda = 0.27$, $F(6, 182) = 28.32$, $p < .001$, $\eta^2 = .48$). The effect size as measured by *d* was 0.69, which is considered a medium effect size, according to Plonsky and Oswald (2014). Follow-up univariate analyses indicated that students of each proficiency level significantly differed in their scores across the three vlogs ($p < .001$). In order to determine which proficiency level was superior to the other, a post hoc test was run. The results of the Tukey HSD test revealed that differences among the groups were significant. As shown in Table 3, high-proficiency students ($n = 21$) had significantly higher scores than mid-proficiency students ($n = 45$), followed by low-proficiency students ($n = 30$). The level of significance was set at 0.05.

Table 3

Descriptive Statistics of Comprehension Scores by Proficiency Level

	Low proficiency learners' mean (*SD*)	**Mid proficiency learners' mean (*SD*)**	**High proficiency learners' mean (*SD*)**
Vlog 1	5.83 (1.88)	8.40 (1.40)	9.42 (2.02)
Vlog 2	3.70 (1.51)	6.00 (1.91)	8.23 (1.18)
Vlog 3	3.50 (1.33)	5.46 (1.58)	6.85 (1.20)

Note. Possible score for each vlog = 10 points

To determine whether the captioning condition leads to better performance in comprehension for any of the three proficiency levels, three ANOVAs were performed. In terms of low- and mid-proficiency learners, the tests showed that the three treatment conditions produced no significant differences in comprehension ($F(2, 27) = 1.70$, $p = .203$ and $F(2, 42) = 3.207$, $p = .051$, respectively). However, a significant difference was detected for high-proficiency learners ($F(2, 18) = 18.036$, $p < .001$). Post-hoc testing indicates that the treatment with L2 captions produced significantly higher scores than L1 captions and no captions ($p < 0.05$) but shows no significant difference between L1 captions and no captions ($p > 0.05$). As can be seen in Table 4, high-proficiency learners in the L2CG achieved the best comprehension.

Table 4

Interaction Between Treatment and Proficiency Level (Means and Standard Deviations on the Total Scores of All Vlogs)

	Low proficiency learners' mean (*SD*)	**Mid proficiency learners' mean (*SD*)**	**High proficiency learners' mean (*SD*)**
L2 captions	12.27 (2.93)	20.76 (1.03)	27.14 (1.09)
L1 captions	12.44 (1.58)	18.60 (2.64)	24.00 (1.70)
No captions	14.40 (3.62)	20.15 (3.41)	22.16 (1.72)

Note. Possible score for three vlogs: 3 × 10 = 30

Research Question 3

A repeated measures ANOVA with a Greenhouse-Geisser correction was conducted to assess whether there were differences in students' comprehension caused by different pictorial support. As stated earlier, the audio input was highly supported in Vlog 1, partially supported in Vlog 2, and slightly supported in Vlog 3 by pictorial images. Results indicated that there was a significant effect of pictorial support ($F(1.86, 176.80) = 81.03$, $p < .001$, $R^2 = .46$, $\text{eta}^2 = .46$). Post hoc tests using the Bonferroni correction revealed that comprehension in Vlog 3 significantly declined ($M = 5.16$, $SD = 1.88$) compared to that of Vlog 2 ($M = 5.77$, $SD = 2.33$), which also significantly declined compared to that of Vlog 1 ($M = 7.82$, $SD = 2.02$). Examination of these means suggests that pictorial support really does have an effect on comprehension. Specifically, our results suggest that comprehension scores decrease as pictorial support decreases. Inversely, as pictorial support increases, so does comprehension. This result was consistent across both the L2CG, L1CG, and NCG and across low-, mid-, and high-proficiency learners, implying that when pictorial support is higher, comprehension gets significantly better, regardless of the captioning condition and proficiency level (see Table 2 and Table 3).

Questionnaire

The descriptive statistics for each questionnaire item are listed in Table 5. Results of the usefulness of captions statements revealed that the L2CG considered captions to be more useful than the L1CG group did. The L2CG found the L2 captions relatively useful for vlog comprehension, with average scores higher than 3.80 on the 5-point scale. However, both the L2CG and the L1CG agreed that their comprehension would be negatively affected without captions. Although the NCG did not agree that their comprehension was negatively affected because of the absence of captions, they agreed that their comprehension would be better with the L2 captions.

Table 5

Participants' Perceptions of L2 and L1 Captions' Usefulness and Need

	L2CG (*n* = 35)		L1CG (*n* = 32)		NCG (*n* = 29)	
	M	*SD*	*M*	*SD*	*M*	*SD*
My comprehension improved because of the English/Arabic captions.	3.86	.81	3.12	.63		
My comprehension would be negatively affected without the English/Arabic captions.	3.37	1.50	3.96	.84		
My comprehension would be better with the Arabic/English captions.	2.94	1.22	3.10	.73		
My comprehension was negatively affected without the English captions.					2.89	1.34
My comprehension was negatively affected without the Arabic captions.					2.19	1.39
My comprehension would be better with the English captions.					3.96	1.05
My comprehension would be better with the Arabic captions.					3.26	1.60

Note. 1= I completely disagree, 5 = I completely agree

Discussion

In contrast to some previously conducted studies, where L2 and L1 captions invariably led to increased listening comprehension (e.g., Guillory, 1998; Hayati & Mohmedi, 2011; Markham & Peter, 2003; Markham et al., 2001; Rodgers & Webb, 2017; Teng, 2019; Winke et al., 2010), this study found no differences in listening comprehension among the treatments for the L2CG, L1CG, and NCG across the three vlogs viewed. Our findings are in line with Montero Perez et al. (2014a) and Montero Perez et al.

(2014b), who found no influence of captions on the listening comprehension of details. The mean comprehension scores for all three vlogs were similar for the three captioning treatments, whether captions were present or absent, which contradicts other studies that demonstrated some or full support for full captions, keyword captions, L2 captions, and L1 captions on the comprehension of details.

The first possible reason for the similar test scores among the three treatment conditions might be that students, particularly low-proficiency ones, did not read the captions because they had to focus on the video (i.e., visual images) and attempt to match the audio to the visual images in order to comprehend the aural input. As suggested by Winke et al. (2010, 2013), because of the limited capacities of working memory and cognitive load, students, particularly lower proficiency ones, were not able to simultaneously process multiple modalities (images, audio, and captions). We speculate that L2 captions more significantly affected students' abilities because of the orthographical differences. Another reason for the non-significant results can be that the study's tests only measured detail comprehension rather than global comprehension, the listening ability where significant differences were observed in other studies (e.g., Montero Perez et al., 2014b; Teng, 2019).

A question might then be asked, "Why didn't the L1 captions help the L1CG in this study whose L1 is Arabic?" Someone might also argue that when L1 captions are provided, there should be 100% comprehension, or when someone watches a movie with L1 captions, they should understand the movie as much as someone who watches it in their L1. The answer is that watching an L2 movie with L1 captions does not necessarily guarantee the viewer will comprehend the movie as L1 speakers do, because L1 captions could split learners' attention and prevent them from concentrating on the audio input due to, as stated earlier, limited working memory and cognitive load. That was pointed out by Hayati and Mohmedi (2011), whose participants commented that the Persian (L1) captions diverted their attention and prevented them from concentrating on the audio input. Similarly, this study's questionnaire data showed that the L1CG found the L1 captions only slightly useful, and when L1 captions were made available for them, their comprehension was not better than the other groups. Just because the study's results indicate no difference in comprehension among the treatment groups, it does not necessarily mean that L1 or L2 captions are not useful; it simply means that these students were possibly unable to make use of them.

Although the effectiveness of captions was not reflected in different comprehension scores among the treatment groups, the questionnaire results indicated that participants in the L2CG considered captions to be useful for understanding the vlogs and the L1CG indicated that their comprehension would be negatively affected without captions. While the NCG indicated that their comprehension was not negatively affected by the absence of captions, they unambiguously agreed that their comprehension would be better with the L2 captions.

Language proficiency was found to be more of a key factor than the presence of captions in the comprehension of vlogs. Similar to Teng's (2019) findings, high-proficiency learners gained higher listening comprehension scores than low-proficiency learners, regardless of the captioning condition. Moreover, higher proficiency learners of any treatment condition outperformed lower proficiency learners of any treatment condition. For instance, high-proficiency learners' listening comprehension without captions was better than that of mid- and low-proficiency learners with L2 or L1 caption support. Teng's suggestion that learners' proficiency levels play a key role in dictating whether they could be engaged with captioned videos and Yeldham's (2018) observation that higher proficiency learners engage more fully with the multiple cues provided by the speakers, captions, and visuals support our speculation because the effectiveness of captions was only found to be significant for high-proficiency learners. In fact, L2 captions produced significantly higher scores for high-proficiency learners rather than L1 captions because, as Hayati and Mohmedi (2011) point out, L1 captions involve translation, requiring learners to switch from one language to another, which takes time, so learners may lag behind and lose track.

Finally, with regard to the effect of pictorial support on listening comprehension (i.e., whether higher pictorial support of the audio in a vlog positively affects performance on listening comprehension), it has been emphasized that factors specific to viewing videos could affect listening comprehension, including the relationship of images to audio (Gruba, 2004; Wagner, 2002). Similar to Durbahn et al. (2020) who found that learners rely more on the use of visual cues and less on the textual ones, our results demonstrated the benefit of pictorial support; namely, better listening comprehension was observed when pictorial support was high, as in Vlog 1, regardless of the presence of captions and proficiency level. Unlike Vlog 2 and Vlog 3, Vlog 1 was filmed and edited more professionally, and the visual images corresponded very closely to the content of the audio, which likely led to the participants' highest comprehension scores. It could be argued that other factors, such as accent, pronunciation, and speaker speed (Buck, 2001) might have negatively affected listening comprehension scores for Vlog 2 and Vlog 3. Yet, while this might be true for Vlog 3, what needs to be emphasized is that the difficulty of those factors were not different between Vlog 2 and Vlog 1. These results do not support Markham and Peter's (2003) claim that the influence of pictorial support for listening comprehension could be ruled out. This difference in results could perhaps be explained due to Markham and Peter using only one video for their study, when it was necessary to have at least two videos with different pictorial support in order to draw a conclusion on its effect.

Conclusion

The purpose of this study was to explore what kind of caption treatment is more effective in developing listening comprehension of vlogs in the context of EFL students: L2CG, L1CG, or NCG. All three groups were exposed to three different treatments. L2CG watched the vlogs with L2 captions and an English soundtrack. L1CG viewed the same vlogs with L1 captions and English audio, while NCG watched the same vlogs without captions but with the same soundtrack. The results of the present study suggest that the presence of captions may not lead to better listening comprehension. This is possibly caused by the limited capacity of working memory and limited cognitive load. Listening comprehension was invariably influenced by proficiency level. Moreover, viewers with and without captions were able to make significant gains to their listening comprehension when the pictorial support was high.

Limitations and Suggestions for Future Research

As this is the first study to investigate the effectiveness of captions on the listening comprehension of vlogs, further research is clearly needed on captioned vlogs, particularly due to the limitations identified. First, keyword captioning was not explored because this study compared captions of the target and native languages to no captions. A comparison of keyword captions to full captions and no captions has been considered in previous studies on videos. Thus, future studies might be conducted to examine whether full captions, keyword captions, and no captions produce differences in the listening comprehension of vlogs. Second, the assessment of comprehension was measured only with multiple-choice tests. The drawback here lies in the need to use other assessment techniques that measure various aspects of listening comprehension (i.e., global comprehension and inferencing ability). Third, three vlogs as the study material is a very small number, representing only a small portion of the vlog genre. Including more vlogs that differ in terms of topics, lengths of duration, accents, speaker speed, pictorial support, and more should increase our understanding of the relationship between vlogs and captions. Fourth, while females might have different reactions to videos and images, only male students were recruited as participants in this study. Hence, whether and how female students' listening comprehension of vlogs differs is worthy of investigation. Finally, the sample population was Arabic-speaking EFL learners. Therefore, we cannot generalize the findings to EFL learners of other L1s. This study could also be replicated with learners who have other L1s. In the same vein, non-English vlogs should be considered in future research to examine whether and how learners deal with captions of other foreign languages.

Pedagogical Implications

Despite the limitations, the study offers some pedagogical implications. Like other video genres, vlogs lend themselves to ESL/EFL curricula since various listening activities can be created for vlogs. When using vlogs for listening comprehension activities, both L1 and L2 captions might be beneficial when learners' language proficiency level is high, with L2 captions being more useful. High-proficiency learners are more capable of simultaneously processing the multiple modalities (visual images, audio, and captions) in the vlogs. The degree of pictorial support for the audio is also crucial when considering vlogs for teaching L2 listening. Namely, when visual images are highly supportive of the audio, better comprehension of vlogs is likely. The more supportive the visual images are, the better the comprehension of the vlogs.

References

Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford University Press.

Buck, G. (2001). *Assessing listening*. Cambridge University Press.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293–332.

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, *88*, 102–166. https://doi.org/10.1016/j.system.2019.102166

Field, J. (2008). *Listening in the language classroom.* Cambridge University Press.

Frobenius, M. (2011). Beginning a monologue: The opening sequence of video blogs. *Journal of Pragmatics*, *43*(3), 814–827. https://doi.org/10.1016/j.pragma.2010.09.018

Frobenius, M. (2014). *The pragmatics of monologue: Interaction in video blogs*. [Unpublished doctoral dissertation]. Saarland University.

Garza, T. (1991). Evaluating the use of captioned video material in advanced foreign language learning. *Foreign Language Annals*, *24*(3), 239–258. https://doi.org/10.1111/j.1944-9720.1991.tb00469.x

Gruba, P. (2004). Understanding digitized second language videotext. *Computer Assisted Language Learning*, *17*(1), 51–82. https://doi.org/10.1076/call.17.1.51.29710

Guillory, H. G. (1998). The effects of keyword captions to authentic French video on learner comprehension. *CALICO Journal*, *15*, 89–108. https://doi.org/10.1558/cj.v15i1-3.89-108

Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology*, *42*(1), 181–192. https://doi.org/10.1111/j.1467-8535.2009.01004.x

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. University of Pennsylvania Press.

LivingBobby. (2018, May 18). *Living on $1 for 24 hours in NYC! (Day #1)* [Video]. YouTube. https://youtu.be/L7FoY03_9FY

Markham, P., & Peter, L. A. (2003). The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *Journal of Educational Technology Systems*, *31*(3), 331–341. https://doi.org/10.2190/bhuh-420b-fe23-ala0

Markham, P., Peter, L. A., & McCarthy, T. J. (2001). The effects of native language vs. target language captions on foreign language students’ DVD video comprehension. *Foreign Language Annals*, *34*(5), 439–445. https://doi.org/10.1111/j.1944-9720.2001.tb02083.x

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*, 312–320.

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*(01), 21–43. https://doi.org/10.1017/s0958344013000256

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. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44357/1/18_01_Monteroperezetal.pdf

Noreillie, A. S., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages: An approximate replication study of Stæhr (2009). *ITL - International Journal of Applied Linguistics*, *169*(1), 212–231. https://doi.org/10.1075/itl.00013.nor

One Girl One Suitcase. (2015, August 15). *How to spend 1 day in Lucerne | Switzerland travel vlog* [Video]. YouTube. https://youtu.be/TJpS5WoPbig

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

Rodgers, M. P. H., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English-language television programs. *CALICO Journal*, *34*(1), 20–38. https://doi.org/10.1558/cj.29522

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Snelson, C. (2015). Vlogging about school on YouTube: An exploratory study. *New Media and Society*, *17*(3), 321–339. https://doi.org/10.1177/1461444813504271

Teng, F. (2019). Maximizing the potential of captions for primary school ESL students' comprehension of English-language videos. *Computer Assisted Language Learning*, *32*(7), 665–691. https://doi.org/10.1080/09588221.2018.1532912

Van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457–479.

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching*. Palgrave Macmillan.

Wagner, E. (2002). Video listening tests: A pilot study. *Columbia University Working Papers in TESOL & Applied Linguistics*, *2*(1), 1–39. https://tesol-dev.journals.cdrs.columbia.edu/wp-content/uploads/sites/12/2015/05/4.-Wagner-2002.pdf

Watkins, J., & Wilkins, M. (2011). Using YouTube in the EFL classroom. *Language Education in Asia*, *2*(1), 113–119. https://doi.org/10.5746/leia/11/v2/i1/a09/Watkins_wilkins

Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407–427. https://doi.org/10.1093/applin/amp010

Werner, E. A. (2012). *Rants, reactions, and other rhetorics: Genres of the YouTube vlog* [Unpublished doctoral dissertation]. University of North Carolina at Chapel Hill.

Winke, P., Gass, S., & Syodorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 65–86. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44203/1/14_01_Winkegasssydorenko.pdf

Winke, P., Gass, S., & Syodorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, *97*(1), 254–275. https://doi.org/10.1111/j.1540-4781.2013.01432.x

Wolters World. (2017, July 16). *Visit Russia - The don'ts of visiting Russia* [Video]. YouTube. https://youtu.be/rzuPtcmsOKA

Wood, M. (2019, August 3). What makes a vlog a vlog? *Diggit Magazine*. https://www.diggitmagazine.com/academic-papers/what-makes-vlog-vlog

Yeldham, M. (2018). Viewing L2 captioned video: What’s in it for the listener? *Computer Assisted Language Learning*, *31*(4), 367–389. https://doi.org/10.1080/09588221.2017.1406956

About the Author

Dukhayel Aldukhayel is an Associate Professor in the Department of English Language and Translation, College of Arabic Language and Social Studies, at Qassim University in Saudi Arabia. He holds an MA degree in TESL/TEFL from Colorado State University and a PhD in Applied Linguistics from the University of Memphis. His research interests include L2 vocabulary, L2 listening, and CALL.

E-mail: dmdkhiel@qu.edu.sa