Examining adolescent EFL learners' TV viewing comprehension

through captions and subtitles

Authors:

Geòrgia Pujadas

University of Barcelona

ORCiD: 0000-0002-0290-1158

georgia.pujadas.jorba@gmail.com

Carmen Muñoz

University of Barcelona

ORCiD: 0000-0002-7001-4155

munoz@ub.edu

Abstract

This study explores the differential effects of captions and subtitles on extensive TV viewing comprehension by adolescent beginner FL learners, and how their comprehension is affected by factors related to the learner, pre-teaching of target vocabulary, the lexical coverage of the episodes and the testing instruments. Four classes of secondary school students took part in an 8-month intervention viewing 24 episodes of a TV series, two classes with captions, and two with subtitles. One class in each language condition received explicit instruction on target vocabulary. Comprehension was assessed through multiple-choice and true-false items, which included a combination of textually explicit and inferential items. Results showed a significant advantage of subtitles over captions for content comprehension, and prior vocabulary knowledge emerged as a significant predictor – particularly in the captions condition. Comprehension scores were also mediated by test-related factors, with true-false items receiving overall more correct responses while textually explicit and inferential items scores differed according to language of the on-screen text. Lexical coverage also emerged as a significant predictor of comprehension.

Keywords: extensive TV viewing, comprehension, captions, subtitles, adolescents

Increasing the amount of exposure to comprehensible input in the target language is beneficial for second language (L2) acquisition (Ellis, 2013), even more so for developing listening competence – which is an often overlooked skill in the language classroom (Nation & Newton, 2009; Vandergrift, 2007). Furthermore, improving the understanding of oral discourse is one of the most difficult challenges FL learners with limited L2 input encounter, especially when they find themselves in an environment - such as the case of Spain - where they are not regularly exposed to the target language.

One way to increase the amount of L2 exposure is through extensive viewing (e.g. Webb & Rodgers, 2009a), which can provide authentic input in an environment that has limited L2 presence (Webb, 2015). TV programmes in particular have been shown to be an effective source of comprehensible input and of natural, contextualized spoken dialogue (Vulchanova, Aurstad, Kvitnes & Eshuis, 2015), having the additional semantic support provided by the images (Rodgers, 2013). Also, compared to other sources of comprehensible input such as reading, TV can provide a large amount of input in a short time, and it is already consumed in large quantities across the EU, with 81% of the population watching it daily (European Commission, 2017). This figure goes up to 88% in Spain, a traditionally dubbing country, where the foreign language soundtrack of films and TV programmes is replaced by a native language soundtrack. Therefore, most of this input is in Spanish (Almeida & Costa, 2014) rather than in the original version (OV), generally English. Compared to other European countries – in which learners are frequently exposed to English through television, movies or newspapers (e.g. Vulchanova et al., 2015) – in Spain most of the exposure to English is limited to formal instructional settings (Muñoz, in press), where there is not enough time to provide learners with as much exposure to the L2 as needed. If learners were to watch TV in

the L2 for enjoyment it could be a valuable source of meaning-focused input (Webb, 2015). Additionally, language learners are highly motivated to watch visual media for language learning (e.g. Vanderplank, 2019).

However, some TV programmes might be too difficult for learners whose linguistic skills are not advanced enough. The addition of on-screen text in the L2 (henceforth captions) or the L1 (henceforth subtitles) can make this input comprehensible (e.g. Danan, 2004). The use of either L1 or L2 text support has been a matter of debate in the past decades, and while research has shown that captions seem to be more adequate for aspects such as vocabulary or pronunciation, subtitles seem to be more effective for content comprehension (e.g. Bianchi & Ciabattoni, 2008; Markham, Peter & McCarthy, 2001). The present study will try to shed light on the matter of the respective benefits of captions and subtitles for comprehension and the possible interactions with other factors related to the learner, the audio-visual materials and the instruments used to measure comprehension.

Audio-visual Input and Comprehension

Although generally associated with entertainment, TV programmes and films can be an effective way to increase foreign language learners' exposure to authentic, naturalistic input. This media complies with Nation's (2007) five conditions for suitable input (Rodgers, 2013): it is processed in large quantities; it is familiar to the language learners; it provides contextual cues (i.e. through image and dialogue); it is comprehensible (Rodgers & Webb, 2011); and it is engaging (Webb, 2010). TV series – as well as other audio-visual materials – have several features that can contribute to the facilitation of information processing, the main ones being the presence of imagery, the possibility to accumulate background knowledge and the recurrence of vocabulary.

The first and most obvious attribute of audio-visual materials is imagery, which, as a semiotic resource, is a powerful mode of meaning-making (see The Douglas Fir Group, 2016). In the context of television viewing, comprehension requires that viewers construct meaning from available sources of information, either linguistic (e.g. phonological, prosodic, lexical) or non-linguistic (e.g. contextual knowledge) (Buck, 2001). Viewers use this knowledge to create a mental representation of input through two types of processing: top-down processing (using context and prior knowledge to make inferences) and bottom-up processing (constructing meaning from the smaller components of words up to discourse-level features) (Vandergrift, 2007). Comprehension occurs when sufficient information is acquired through both types of processing (Buck, 2001). Research has shown that low-proficiency listeners rely more on top-down processing, because of their limited word segmentation skills and proficiency. Imagery – which can work as a "compensatory mechanism" (Vandergrift, 2007: 193) – provides beginner-level learners with contextual knowledge which allows them to shift attention away from top-down processing and focus their attention on details of the story, which in turn can positively affect comprehension (Rodgers, 2013, 2016).

The benefit brought by the addition of visual input to aural input for comprehension was explained by Paivio's (1986) Dual Coding Theory, which argues that verbal stimuli (language) and non-verbal stimuli (image) are processed by two different systems that interact. The activation of both systems results in better recall and greater depth of processing – which would explain why L2 learning can be enhanced by combining images with verbal information (Sydorenko, 2010). Imagery has also been shown to have an advantage over verbal input for information processing (Paivio, 2008), and it leads to deep processing of aural input, especially when combined with text (e.g. Jones, 2003). This can be linked to

Mayer's (1997) Cognitive Theory of Multimedia Learning, according to which "people learn more deeply from words and pictures than words alone" (Mayer, 2014: 43).

A number of studies support the idea that language learners can use the imagery associated with videos to assist information processing. Research on listening shows that the presence of images has a positive impact on comprehension (e.g. Jones & Plass, 2002), especially when language proficiency is low (Maleki & Safaee Rad, 2011), while images also help reduce anxiety when confronting unfamiliar topics (Hasan 2000). In a study by Ockey (2007) it was observed that learners used more visual clues from videos than from still images, though both were a helpful support while listening. Baltova (1994) also reported positive effects of visual clues from videos, and found that learners who had access to audio and video almost doubled the comprehension scores of the audio-only group. A recent study by Durbahn, Rodgers and Peters (2019) assessed comprehension of a documentary distinguishing between imagery-based, audio-based and imagery plus audio-based questions. Results showed that, when imagery could be used, learners relied less on the spoken text, whereas for audio-based questions vocabulary knowledge was the factor that played a more important role. A few studies on vocabulary have also found that images provide some sort of semantic support (Pujadas & Muñoz, 2018; Pujadas, 2019; Bisson, Van Heuven, Conklin & Tunney, 2014; Rodgers, 2018a;), suggesting that "imagery can function as a 'sketch pad' for holding information to be worked on by other cognitive mechanisms" (Paivio, 2008: 28).

Besides imagery, another feature of TV series that may also contribute to facilitate comprehension is that they are serial in nature and – compared to documentaries or movies – episodes are rarely watched in isolation. On the basis of the findings from narrow reading (e.g. Schmitt & Carter, 2000), which show that reading texts with similar topics and plot lines may facilitate comprehension, Rodgers and Webb (2011) proposed the concept of narrow

viewing. Narrow viewing refers to viewing a variety of television programmes on the same topic, or of the same genre or series. By targeting programmes of a similar nature (as opposed to watching single episodes randomly), learners can gradually accumulate background knowledge. This background knowledge will benefit top-down processes and, thus, stimulate comprehension of the content (Li, 2014). The more episodes you watch from the same TV series, the more information you have about the recurring characters and locations, making it possible for the viewer to predict or guess what a character will do or say as they accumulate background knowledge, while it also helps viewers to get used to accents or the way characters talk. Rodgers and Webb (2011) also found that related television programmes are likely to contain fewer word families than unrelated programmes, which in turn could progressively facilitate understanding.

Captions, Subtitles and Proficiency

TV series might be too demanding for those learners whose language skills are not high enough to achieve satisfactory comprehension (Webb, 2010; Webb & Rodgers, 2009a). Because of their limited linguistic knowledge, beginner-level learners cannot process input automatically as more advanced learners do (Vandergrift, 2007). They have to consciously decode aural input into meaningful units (bottom-up processing), and "a large proportion of what they hear may be lost, given the speed of speech and the inability of working memory to process all the information within the time limitations" (Vandergrift, 2007: 193). A number of factors can affect this process, including the learners' ability to recognize words and recall their meaning (Buck, 2001) and learners' prior vocabulary knowledge (e.g. Webb & Rodgers, 2009b). Given the need to provide learners with non-adapted, natural samples (Gilmore, 2007), the addition of on-screen text in the form of captions – commonly available nowadays

– may provide access to authentic foreign language material that would otherwise be difficult to comprehend for non-native speakers (Vanderplank, 2016a). Captions may also reduce viewers' anxiety when faced with input that might be beyond their perceived language skills (Vanderplank, 1988) and hence increase learners' motivation (Winke, Gass & Sydorenko, 2010).

Previous research examining comprehension of audio-visual input has consistently shown the positive effects of captioning (or keyword captioning) over non-captioning for viewing comprehension (e.g., Gass, Winke, Isbell & Ahn, 2019; Montero-Perez, Peters & Desmet, 2014; Montero-Perez, Van Den Noortgate & Desmet, 2013; Rodgers & Webb, 2017; Winke et al., 2010). Although some studies have suggested that the presence of on-screen text might impose a cognitive burden for beginner learners (Taylor, 2005) or might be distracting for more advanced students (Bairstow & Lavaur, 2011), the cognitive load that it adds seems not to be so detrimental as to hinder comprehension (Birulés-Muntané & Soto-Faraco, 2016). It has also been observed that aural and verbal textual information are processed in parallel, which would imply that the presence of text does not hinder the processing of the audio (Danan, 2004; d'Ydewalle & Gielen, 1992).

Similar findings are shown by studies comparing captioning and subtitling versus no-text, with either language condition outperforming no-text conditions. However, whether it is captions or subtitles that are more useful in general in an audio-visual context is still a matter of debate, with studies showing mixed results depending on what aspect of the language is being assessed and on learners' proficiency. Captions, in particular, have been shown to aid in various aspects of language learning such as written form recognition (Sydorenko, 2010), aural form recognition (Markham, 1999), form-meaning connection (Winke et al., 2010), meaning recall (Peters, 2019), and speech perception (Mitterer & McQueen, 2009) and

segmentation (Charles & Trenkic, 2014). The majority of studies on comprehension concur, however, that subtitles (in the viewers' native language) facilitate understanding of the content better than captions (Bianchi & Ciabattoni, 2008; Birulés-Muntané & Soto-Faraco, 2016; Latifi, Mobalegh & Mohammadi, 2011; Lwo & Lin, 2012; Markham, Peter & McCarthy, 2001; Markham & Peter, 2003), which is not surprising, since reading the text in your native language logically facilitates understanding. Subtitles are processed automatically, and provide on-line translations (Sydorenko, 2010). Also, learners tend to be better at reading than listening, and they can benefit from seeing difficult content on their native language first (Markham et al., 2001). On the other hand, scant exceptions have favoured captions (e.g. Hayati & Mohmedi, 2011) and some studies have reported inconclusive or conflicting results depending on learners' proficiency level (e.g. Bairstow & Lavaur, 2011; Guichon & McLoran, 2008; Matielo, Oliveira & Baretta, 2017; Vulchanova et al., 2015).

Learner proficiency seems indeed to be a crucial factor for viewing comprehension. Despite the benefits of captions for language learning, they do not compensate for fast speech and unknown vocabulary (Guillory, 1988), especially in the case of low-proficiency learners. Thus, a minimum competency threshold might be necessary to benefit from captioning (Neuman & Koskinen, 1992), whereas subtitles allow understanding of the input regardless of the viewers' proficiency level. If the input is not understood, learning is unlikely to occur, since learners do not pay attention to the precise meaning of the words (Laufer, 2005). This is also supported by Muñoz's (2017) eye-tracking study, which revealed that young, low-proficient viewers spent shorter time fixating on captions than more proficient participants, suggesting that learners who perceived their level of proficiency as too low for comprehension simply did not make the effort to process captions. Yet, detractors of the

subtitles in the L2 context argue that, if students can simply read the L1 text, they will not listen to the L2 audio (Steward & Pertusa, 2004).

A few studies have compared the effects of captions and subtitles on adolescent learners' comprehension, reporting controversial results contingent upon proficiency and age. Bairstow and Lavaur (2011) investigated the comprehension of a 9-minute clip by secondary school students (aged 15-18). They found that, for advanced learners, having on-screen text was distracting, and that the non-captions group outperformed the subtitles and captions groups – who performed similarly. However, for beginner learners it was found that on-screen text had a facilitating effect, and that the subtitles group significantly outperformed the others. They also found that visual and dialogue information was recalled differently depending on viewers' proficiency level and language of the on-screen text. Lwo and Lin (2012) studied the differential effects of captions and subtitles with Grade 8 learners and found that comprehension also depended on learners' proficiency and that differences between the language groups were not significant. They also reported that less proficient students were not overloaded with too much information, and that it seemed they could select what they needed from the input available. Vulchanova et al. (2015) looked at the comprehension of a 22-minute episode by 16- and 17-year-olds. For the older group, they found no significant differences between language groups, but for the younger group those in the captions condition performed better, though the most significant predictor was vocabulary size rather than language of the on-screen text.

Altogether, the above studies suggest that subtitles are generally more useful for comprehension than captions, especially for beginner learners. Subtitles provide on-line translations and allow understanding of the content regardless of the learners' language skills. Captions, on the other hand, can help learners with written and aural form recognition and

with making form-meaning connections, but learners' bottom-up processing may be negatively affected if their vocabulary knowledge is limited. In the case of younger learners, the few existing research findings show the important role played by proficiency but they are inconclusive as regards the effects of captions and subtitles on comprehension. Inconsistent results may also be due to developmental differences.

Mediating Factors in Comprehension through TV

Explicit Focus on Vocabulary

The addition of on-screen text affects attention to input (Winke et al., 2010), since the reading of captions and subtitles is automatic (e.g. Bisson et al., 2014). This attention-drawing function might be seen as positive or negative (i.e., attention depleting). On top of this, in the FL classroom, if learners are required to focus their attention on, for example, vocabulary, this might come at a cost for comprehension, as attentional resources are limited. While research on the use of advanced organizers has shown that providing pre-listening activities has a generally positive effect on comprehension (e.g. Elkhafaifi, 2005) – as they seem to help listeners activate their prior knowledge (top-down processing) (Vandergrift, 2007) – explicit focus on vocabulary yields conflicting results. Chang and Read (2006), who investigated various forms of support for listening comprehension, found that vocabulary instruction was the least effective one – regardless of proficiency level. In another study Lee (2007) explored the effects of textual enhancement on reading comprehension and found that, while vocabulary improved, overall comprehension decreased. This can be seen as a special case of cognitive overload of the verbal channel (see the Cognitive Load Theory in multimedia learning; Mayer & Moreno, 2003; Sweller, 1999) which occurs when learners' cognitive processing exceeds the available cognitive capacity. This is also in line with the

more general Theory of Input Processing (VanPattern, 1996), which states that "learners can do only so much in their working memory before attentional resources are depleted and working memory is forced to dump information to make room for more (incoming) information" (VanPattern, 2002: 757).

Webb and Rodgers (2009b) pointed out, however, that "reaching the target vocabulary size [needed for comprehension] may be too difficult a task for many learners and movies should probably not be used without providing some learning support" (Webb & Rodgers, 2009b: 420). On the basis of an analysis of the lexical coverage of different genre TV programmes, Webb (2010) also suggested that viewers' pre-learning the most frequent low-frequency word-families in those programmes could potentially be more conducive to enhancing their comprehension than just increasing vocabulary size. Pre-teaching vocabulary that the learners will encounter in the input seems, therefore, to provide them with enhanced learning opportunities, as it has been shown in recent studies (e.g. Pujadas & Muñoz, 2019). Nevertheless, to the authors' knowledge, no studies have looked at how comprehension may be affected by having or not having explicit vocabulary instruction.

Familiarity with viewing OV, enjoyment, and engagement

Familiarity with viewing OV audio-visual material may have an impact on the viewing process itself. A European survey (2011) carried out in 33 countries and with 11,000 respondents found that younger people (aged 12-25) preferred subtitling over dubbing, but with a significant exception in dubbing countries (such as Spain), where even young citizens preferred dubbing to subtitles – primarily out of habit. Respondents from subtitling countries were more adept at quickly developing strategies to take advantage of them compared to those coming from dubbing countries (Vanderplank, 1988). This suggests that familiarity

with the use of on-screen text (either in the L1 or the L2) may play a role, and that learners who are used to watching captioned or subtitled OV input might be able to benefit more from it.

Taylor (2005) found that beginner students with little background in reading and listening in the foreign language found it difficult to attend to the three channels and were confused or distracted by the presence of captions, but he emphasized that learners who had only two more years of study were capable of doing so. Pujadas (2019) interviewed a group of secondary school students from the same sample as those in the current study who had been watching TV series in the classroom for 6 months, and found that learners reported a change in their viewing habits at home – moving from dubbed to subtitled TV watching, and from subtitles to captions. Students also reported that by the end of the year they understood the series better as they got used to actors and their voices, a finding reported in previous studies too (e.g. Rodgers, 2013). Although eye-tracking experiments data have revealed that on-screen text is read regardless of the language or the viewers' level of familiarity with it (e.g. d'Ydewalle & Gielen, 1992), that learners read the text does not imply that they do it efficiently.

Besides splitting cognitive resources between comprehension and other aspects of language learning, other factors that might play a role in understanding TV input are attention to and enjoyment from the TV series themselves. A concern might simply be whether learners are paying attention to the input or not, especially when research is classroom-based and TV viewing might be seen as a leisure activity by the students (Vanderplank, 2016b), or because students just do not like the programme. However, results from a survey about attitudes towards TV input in the L2 classroom indicated that – independently of age and

language skills – learners found TV viewing more enjoyable and engaging than traditional listening activities (Pujadas, 2019; Pujadas & Muñoz, 2017).

Testing comprehension

Research in reading and listening comprehension have revealed that differences in testing yield varying degrees of difficulty for test-takers. Such differences have significant effects on comprehension scores depending on input materials, question format and language used – especially for beginner learners (e.g. Shohamy, 1984). In order to design appropriate tasks, we need to establish first what is assessed (construct validity) (Vandergrift, 2007). Buck (2001) proposed a flexible, baseline definition of the listening construct adequate for L2 classroom assessment that describes listening comprehension as: "the ability to: 1) process extended samples of realistic spoken language, automatically and in real time; 2) understand the linguistic information that is unequivocally included in the text; and 3) make whatever inferences are unambiguously implicated by the context of the passage" (Buck, 2001: 114). This definition seems to be appropriate for comprehension through TV input too, considering that even if the addition of visual support may facilitate information processing, we still assess viewers' ability to understand what is being said.

Related to this, Wagner (2002) investigated construct validity of a video-based test, and found evidence for the validation of a two-factor model based on the ability of processing explicit information and implicit information in aural input, instead of the hypothesized top-down and bottom-up factors. This falls in line with previous research that already called attention to these two main types of questions, with numerous variations regarding their nomenclature and possible sub-typologies (Alptekin & Erçetin, 2010; Buck, 2001; Davey & McBride, 1986; Pearson & Johnson, 1978; Rodgers, 2013; Shohamy & Inbar, 1991). Most

commonly, textually explicit or literal questions refer to items that ask for information explicitly stated in the text (information that could be underlined), regarding details or trivial information, and they normally involve bottom-up processing. Textually implicit or inferential questions, on the other hand, ask for information that is found by integrating different pieces of information and making inferences, involving top-down processing. This type of questions can include going beyond the text to understand the central gist or idea or synthetizing information to draw conclusions. Although still in need of more research, studies including different question types indicate that item type has an effect on comprehension scores and that the presence of on-screen text – or the absence of it – interacts with item type (e.g. Rodgers, 2018b; Shohamy & Inbar, 1991). Another aspect to consider is whether questions are audio-based or imagery-based, since differences have also been found between these two types (Durbahn et al., 2019). Item format – that is, how questions are presented – also deserves attention. Response format can significantly affect comprehension scores, as shown, for example, by Cheng's (2004) study, where learners completing multiple-choice items outperformed respondents of open-ended items.

Lexical coverage

Lexical coverage – that is, the percentage of known words in the input – provides an indication of the vocabulary size needed for adequate comprehension of a specific text, together with the vocabulary load that it represents (Webb, 2011; Webb & Rodgers, 2009b). The higher the lexical coverage, the easier it might be for learners to understand the content (Webb, 2011). The lexical coverage of the episodes also plays a prominent role in comprehension, beyond the proficiency level of the learners itself. Research on reading and

listening – and more recently on TV and film viewing – has extensively shown that vocabulary knowledge is a strong predictor of content comprehension (e.g. Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Rodgers, 2013; Van Zeeland & Schmitt, 2013; Webb & Rodgers, 2009b), although disagreement exists on the percentage of lexical coverage needed for adequate understanding of the written, aural or audio-visual input. Research on extensive reading suggests that learners need up to 95% coverage for minimal comprehension and 98% coverage for optimal comprehension (e.g. Laufer & Ravenhorst-Kalovski, 2010), while research on informal listening proposes a less conservative figure, suggesting that a coverage of 90-95% might be enough to understand everyday conversations (e.g. Van Zeeland & Schmitt, 2013). For viewing comprehension, it has also been suggested that 95% might be enough, because of the additional support provided by images (Rodgers & Webb, 2011), and that less coverage might be needed for adequate comprehension compared to unassisted reading (Durbahn et al., 2019). In a pioneer study assessing comprehension of several consecutive TV episodes, Rodgers (2013) found that comprehension improved with increased lexical coverage in most but not all episodes, which indicates the need to take into account differences on episodes' lexical coverage when assessing comprehension across different videos.

Aims and Research Questions

The present study addresses the gaps identified in the sections above by exploring the benefits of captions and subtitles on content comprehension of a TV series by adolescent FL learners over a period of eight months. The study also examines the effect that having students' attention pre-directed towards specific items in the input might have on information

processing, by comparing two instructional conditions (with or without pre-teaching of target vocabulary). Moreover, it addresses the role that learners' individual differences may play, as well as the impact that factors related to the episodes selected for the treatment and the testing instruments might have on comprehension. Specifically, the study poses the following research question and sub-questions:

To what extent does the language of the on-screen text affect comprehension of TV series? Is comprehension of TV series also affected by:

a) Other instruction-related factors (i.e. explicit focus on vocabulary items)?

b) Learner-related factors (i.e. general proficiency, vocabulary size, familiarity with OV, attention to and enjoyment from the series)?

c) Test-related factors (i.e. item type and item format)?

d) Episode-related factors (i.e. lexical coverage)?

Method

Participants

The original pool of participants were 106 secondary school learners (65 females, 41 males) in Grade 8 (13-14 years old) from a state school in the area of Barcelona. They were Catalan-Spanish balanced bilinguals and they had a beginner to low-intermediate proficiency level in English (Pre-A to B1 according to the Common European Framework of Reference), and a mean vocabulary size of 1,959 words (as measured by the X_Lex test). Prior to the intervention, around 55% of participants reported watching movies or TV series in English with L1 subtitles on a weekly basis and around 15% with L2 captions or no subtitles. More than 50% reported finding subtitles useful or very useful and only 4% considered them to be useless or annoying.

Participants had been randomly distributed in four classes by the school. Although all students took part in the intervention, only those who had 85% attendance or more were included in the analysis, leaving a total of 88 (56 females, 32 males). Two of the classes were assigned to the captions condition (n = 46), and the other two were assigned to the subtitles condition (n = 42). One class in each language group was taught target vocabulary. According to the language of the on-screen text and whether they had instruction or not, the groups were the following: captions with instruction (n = 24), captions without instruction (n = 22), subtitles with instruction (n =22), and subtitles without instruction (n = 20).

Audio-visual Materials

The TV series selected was *Fresh off the Boat* (Khan et al., 2015), which was found suitable for various reasons: the length of the episodes was adequate for a 1-hour lesson (the average running time was 20 minutes); it was a sitcom, a format with which participants are familiar through watching similar TV series; it was serial in nature, which allowed participants to gather information about the characters as they continued watching new episodes (Rodgers, 2013); it was not strongly accented; its content was appropriate for this particular age group; and it was engaging (the main character was the same age as the participants, and they could identify with him). Also, at the time the intervention took place *Fresh off the Boat* had not been aired in Spain, which minimized the possibility that participants had watched any of the episodes before.

From the first and second season of the TV series, 24 consecutive episodes[1] were selected for the treatment. By the end of the intervention participants had received 515 minutes of

exposure to audio-visual input and had been exposed to a total of 69,350 tokens. Subtitles (in Spanish) were manipulated by the first author to ensure – to the extent possible – comparability with captions in terms of length and number of encounters with the target vocabulary as well as translation accuracy. The 24 episodes chosen for the intervention were analysed using the RANGE software (Nation & Heatley, 2002). The analysis of the lexical profile showed that, overall, the series reached 93.84% coverage at the 2,000 word-level and 95.70% coverage at the 3,000 word-level plus proper nouns and marginal words[2]. Research on informal listening has suggested that a coverage of 90 - 95% might be enough (Noreillie, Kestemont, Heylen, Desmet & Peters, 2018; Van Zeeland & Schmitt, 2013), so the series was considered adequate. Participants in the present study had a mean vocabulary size of almost 2,000 words[3] (which for this series represented a coverage of around 94%) and they had the additional support of the on-screen text, which ensured that input was challenging enough to promote learning but not overwhelming (Krashen, 2003).

Instruments

   Testing learner-related factors. Initial proficiency was assessed by means of the Oxford Placement Test (OPT) and vocabulary size was measured using the X_Lex tool[4] (Meara & Milton, 2003). Two questionnaires were administered to collect information on learners' viewing habits, attitudes towards and impressions about viewing of TV series in general. A first questionnaire was administered prior to the intervention to assess learners' familiarity with viewing OV TV series or movies. Participants reported on a 6-point scale how often they watched OV with subtitles, captions or without text, from *never* to *more than 6 hours per week*. For the analysis, data was recoded and learners where divided into three categories with a similar number of participants in each group: low-frequent, mid-frequent and high-

frequent viewers. A second questionnaire was given after the intervention to gather data on participants' perceived attention to and enjoyment from the series *Fresh off the Boat*. Attention and enjoyment data were self-reported, and participants had to express how attentive they had generally been during the screenings and how much they liked the TV series, using a 5-point Likert scale from 1 (*not at all*) to 5 (*a lot*). The variables attention and enjoyment were also recoded into three categories – low, mid and high – to have a more balanced number of learners in each group for the analysis.

Testing comprehension. Comprehension was assessed by means of post-viewing tests, administered after each of the 24 episodes. Tests were presented in Spanish as its main purpose was to assess learners' content comprehension, and the use of the L1 ensured avoiding errors attributable to poor comprehension of the questions (Vandergrift, 2007). Each test consisted of 10 items, including 5 *multiple-choice* items (MC) and 5 *true/false* items (TF). Using a variety of question formats provides a more balanced assessment (Buck, 2001) and participants were already familiar with these two item formats. Also, both provide a quick and reliable method for testing understanding of the content. MC items had 3 options (1 correct and 2 distractors). All items were designed in a way that the information given by the images of the video alone was not sufficient to answer the question, and the two distractors in the MC items did not provide clues to respond to other questions. Comprehension items also included two types of questions: *textually explicit* items (when the information is explicitly stated in the text, and it could be underlined in the script)[5], and *inferential* items (when the information is found by combining or deducing from different pieces of information, integrating them to understand the central gist or idea). The operationalization of item type was based on an adaptation from Davey and McBride (1986), Alptekin and Erçetin (2010), and Rodgers (2013). It needs to be noted that the distinction between textual explicit and inferential items is based on how the information is retrieved,

rather than on the type of information asked (i.e. global, detail) (see Appendix). Table 1 shows the distribution of the total 240 comprehension items.

Table 1. *Comprehension items by format and type of question*

	Textually explicit	Inferential	Total
Multiple-choice	59	61	120
True/False	60	60	120
Total	119	121	240

Procedure

The intervention took place during a whole academic year and was embedded as a part of the normal English lessons. While initial tests of proficiency were administered by the first author, the 24 viewing sessions were conducted by the school teachers, who received training prior to the start of the academic year. The complete intervention – including proficiency tests, vocabulary tests at the beginning and end of the terms and questionnaires – extended over 32 sessions[6].

As stated above, four intact classes took part in the study, with each one assigned to a different experimental condition according to language of the on-screen text and vocabulary instruction. Viewing sessions followed a similar structure and each session lasted around 50 minutes. Only the two groups with explicit instruction on target vocabulary started with a pre-viewing task aimed at teaching five target items and three distractors appearing in the episode. Students had 5 minutes to complete the activities, which included matching exercises, word searches, fill-in-the-blanks tasks and crosswords, and were corrected orally by the teacher. Students were not asked to memorize the words nor were they provided with any strategies to use during the viewing. Then, all four groups watched the episode (with

either captions or subtitles) and completed a vocabulary task and a content comprehension task, which were only given after the viewing. The post-viewing tasks were not corrected in class. Comprehension was assessed in the same way across the four groups, regardless of the experimental condition.

Results

Preliminary Analysis

Table 2 shows the descriptive statistics for proficiency (OPT scores), vocabulary size (X_Lex scores) and familiarity with OV (self-reported data from the questionnaire) prior to the intervention. The table reports the mean scores per each language group. As can be observed, there were no significant differences between the groups in terms of proficiency (F (1,77) = .861 p = .356), vocabulary size (F (1,74) = .203; p = .653) or familiarity with viewing OV (F (1,83) = .015; p = .904).

Table 2. *Mean values (SD in brackets) of initial proficiency, vocabulary size and familiarity with OV*

	n	Proficiency (OPT)	n	Vocabulary size (X_Lex)	n	Familiarity with OV
Subtitles	37	92.76 (13.85)	38	1,988 (486)	41	2.02 (.79)
Captions	42	95.79 (15.00)	38	1,931 (601)	44	2.05 (.81)
Mean		94.37 (14.47)		1,959 (544)		2.04 (.79)

All 240 comprehension items from the 24 tests were scored dichotomously (1 = correct / 0 = incorrect). Once tests scores were obtained, the difficulty index was calculated to assess how easy or hard the items were in relation to the total correct responses within the sample

(Del Rincón, Arnal, Latorre & Sans, 1995). From the 240 items, 40% were very easy, 42% had an easy to medium level of difficulty and 18% were considered hard or very hard. However, an item discrimination index used as a validation measure (Kelly, 1939) showed that 67.1% were very good, 12.5% were good, 11.3% were regular, and 9.2% were poor discriminators. Poor discriminators were not eliminated after checking that they were homogeneously distributed across the 24 tests. The mean discrimination index per test was *good* for 2 tests and *very good* for the other 22.

The analysis of the scripts of the 24 episodes showed that, as mentioned before, overall the series reached a lexical coverage of 95.70% – which is the general threshold for adequate comprehension (e.g. Van Zeeland & Schmitt, 2013) – at the 3k word-level, plus marginal words and proper nouns. Exploration of the data showed, however, that coverage provided by the first 3,000 words of the BNC/COCA word lists ranged from 93.98% to 96.73% between the episodes. Although the difference seems small (2.75%), research has shown that even a small increase in lexical coverage can already be beneficial for comprehension (Laufer & Ravenhorst-Kalovski, 2010), and since it could not be assumed that all episodes were equally difficult, the percentage of lexical coverage per episode was also included as an episode-related factor in the analysis.

Table 3 shows the number (and percentage in brackets) of correct and incorrect responses for the 240 comprehension items, separated by language condition. As can be observed, overall the subtitles group had 17.8% more correct responses than the captions group.

Table 3. *Number of correct and incorrect responses per language condition*

	Correct responses	Incorrect responses
Subtitles	7,834 (82.0%)	1,716 (18.0%)

Captions	6,768 (64.2%)	3,772 (35.8%)
Total	14,602 (72.7%)	5,488 (27.3%)

Factors affecting TV Viewing Comprehension

Our research question aimed at examining the effect of several variables on comprehension scores, including the following intervention-related, learner-related, test-related and episode-related parameters:

Figure 1. *Independent factors*

A Generalized Linear Mixed Model (GLMM) with repeated measures was calculated using SPSS 21.0 with comprehension score (at item-level) as the outcome variable, and 10 variables (see Figure 1) as fixed factors, including all two-way interactions. This type of statistical test was found particularly appropriate because of several reasons: it does not require normal distribution nor homogeneity of variances; there was an acceptable ratio of observations to independent variables; and there was no multicollinearity. GLMM also allows the inclusion of learner variables, intervention variables and item variables in a single model. In this type of analysis, a particular score (correct or incorrect) is defined by the combination of participant, item and response. The GLMM was based on 17,310 observations.

To arrive at the best fitting model, we entered all the explanatory variables in the model and removed one by one all non-significant interactions and main effects ($p < .10$). Table 4 presents the final fitted model, and Table 5 shows the significant main effects for the

categorical variables. The model revealed that there were four factors that significantly ($p <$ .05) contributed to the model: language of the on-screen text, vocabulary size, item format and lexical coverage; and one factor that contributed marginally: type of instruction. Three significant interactions emerged: between language and vocabulary size, between language and item format and between language and item type.

Table 4. *Results from the GLMM: influence of intervention-related, learner-related, video-related and test-related factors on comprehension scores*

Terms	Coeff	SD	t	Sig	Exp Coeff	95% CI for Exp Coeff	
						Lower	Upper
Language	1.610	.4512	3.569	<.001	5.004	2.067	12.116
Instruction	-.189	.1107	-1.708	.088	.828	.666	1.028
VS	.001	.0001	4.499	<.001	1.001	1.000	1.001
Format	-.187	.0447	-4.187	<.001	.829	.760	.905
Coverage	.173	.0241	7.162	<.001	1.188	1.134	1.246
Language*VS	.000	.0002	-2.025	.043	1.000	.999	1.000
Language*Format	.184	.0715	2.570	.010	1.202	1.045	1.383
Language*Type	.279	.0716	3.900	<.001	1.322	1.149	1.522

Language = Language of the on-screen text; Instruction = Type of instruction; VS = vocabulary size; Coverage = Lexical coverage; Format = Item format; Type = Item type

Table 5. *Results from GLMM: influence of fixed main effects*

Factor	Group	$\bar{X}$ (SE)	df	F	*p value*
Language	Subtitles	.826 (.013)	1, 17300	76.355	<.001
	Captions	.646 (.016)			
Instruction	Yes	.728 (.015)	1, 17300	2.923	.087
	No	.764 (.015)			
Format	MC	.737 (.011)	1, 17300	7.081	.008
	TF	.755 (.011)			

MC = multiple-choice; TF = true-false; TE = textually explicit; IN = inferential

Intervention-related variables. The GLMM analysis showed that there was a significant main effect of language of the on-screen text ($p < .001$), indicating that an average learner's score would be 14% higher if they were in the subtitles groups when all other factors were held constant (see Table 5). There was an interaction effect between language condition and three other parameters – vocabulary size, item format and item type –, which suggested that the effect of these three variables needs to be explained in relation to language (see below). There was a tendency for comprehension scores to depend on type of instruction ($p = .088$), indicating that the two groups who received explicit instruction on target vocabulary items tended to score lower than the two groups without instruction.

Learner-related variables. The model showed that vocabulary size was the only learner-related variable in our study that emerged as a predictor of comprehension ($p = .004$), while general proficiency did not appear to be a significant predictor (although the two variables correlated significantly; $r = .334$ $p < .001$). The interaction effect found between language group and vocabulary size ($p = .043$) indicated that the effect of vocabulary size depended on the language of the on-screen text, and that vocabulary size was only significant for the captions group. Figure 2 illustrates the relationship between these two variables, showing the participants' average percentage of comprehension per vocabulary size and language condition.

Figure 2. *Mean percentage of comprehension per vocabulary size and language*

group
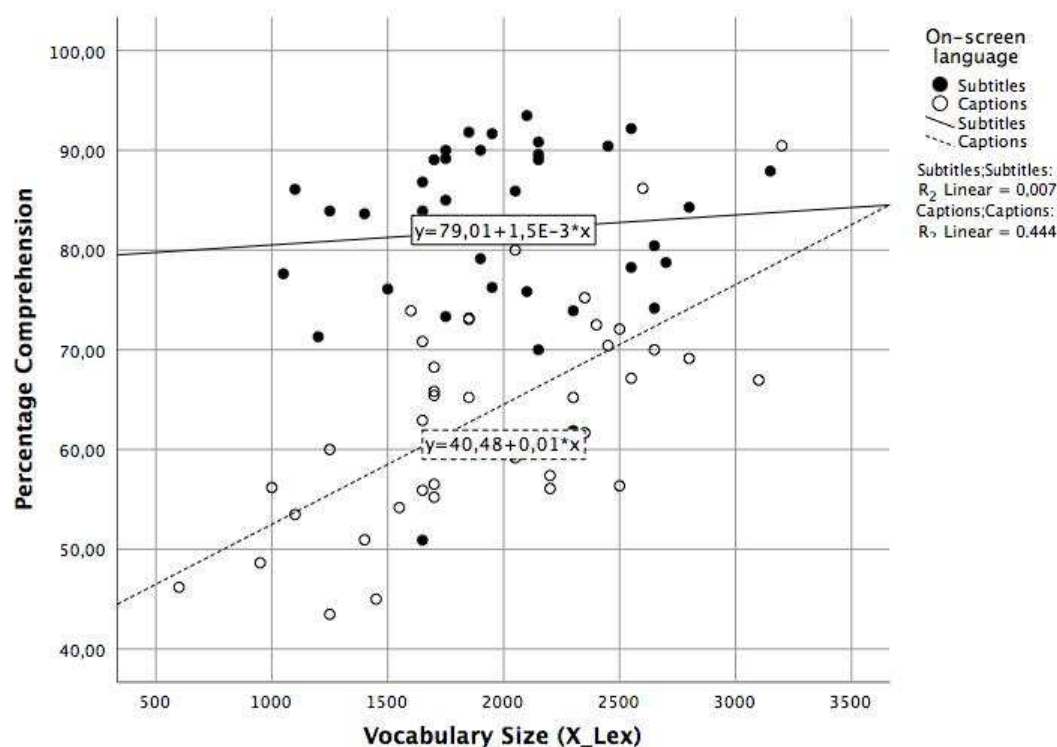
Attention and enjoyment did not appear to contribute to the final fitted model, but it was observed that, when introduced in the model individually, they emerged as significant predictors ($p < .001$ and $p = .003$ respectively), with higher attention and enjoyment associated with higher comprehension gains. Although no interaction was found with language condition, further exploration showed that attention and enjoyment were significantly higher in the subtitles group compared to the captions group ($F (3,82) = 6,581$; $p < .001$ and $F (3,82) = 8,753$; $p < .001$ respectively). Familiarity with viewing OV did not contribute to explaining comprehension scores.

Test-related variables. With respect to test-related variables, the GLMM indicated that item format was a strong predictor of comprehension scores while item type was not. Both, however, interacted significantly with language condition although they had a different impact on comprehension depending on the language of the on-screen text. Table 6 presents

the mean comprehension scores per item format and item type when divided by language condition.

Table 6. *Results from the GLMM: interactions. Categorical factors*

Interaction			$\bar{X}$ (SE)	df	F	*p value*
Language* Format	Subtitles	MC	.826 (.013)	1, 17300	.004	.952
		TF	.826 (.013)			
	Captions	MC	.624 (.017)	1, 17300	17.571	<.001
		TF	.667 (.017)			
Language* Type	Subtitles	TE	.835 (.013)	1, 17300	5.125	.024
		IN	.817 (.014)			
	Captions	TE	.628 (.017)	1, 17300	11.687	.001
		IN	.663 (.017)			

*MC = multiple-choice; TF = true-false; TE = textually explicit; IN = inferential*

For item format, learners in the subtitles groups performed equally in both formats, while in the captions groups learners performed significantly better in the true-false items than the multiple-choice items ($p < .001$). For item type, the subtitles condition performed better in the textually explicit items than in the inferential ones ($p = .024$), while inversely the captions group performed better in the inferential items ($p = .001$). In sum, the significant main effect of item format ($p < .001$) indicates that, independently of the language condition, true-false items had significantly more correct responses than multiple-choice items. On the other hand, the interaction between item type and language indicates that correct responses in one type or the other depended on the language of the on-screen text.

Episode-related variables. The lexical coverage of the episodes also emerged as a strong predictor of comprehension ($p < .001$), with no interaction with language of the on-screen text, indicating that episodes with higher lexical coverage received a higher number of correct responses (see Table 4). Figure 3 shows the mean raw score for the 24 episodes by

each language group (captions and subtitles). Independent samples t-tests revealed that the subtitles groups significantly outperformed the captions group in all 24 episodes – ranging from $p < .001$ to $p = .021$. While comprehension scores vary from episode to episode, no linear progression from the first to the last session can be observed[7].

Figure 3. *Mean raw comprehension score per episode and on-screen language group*
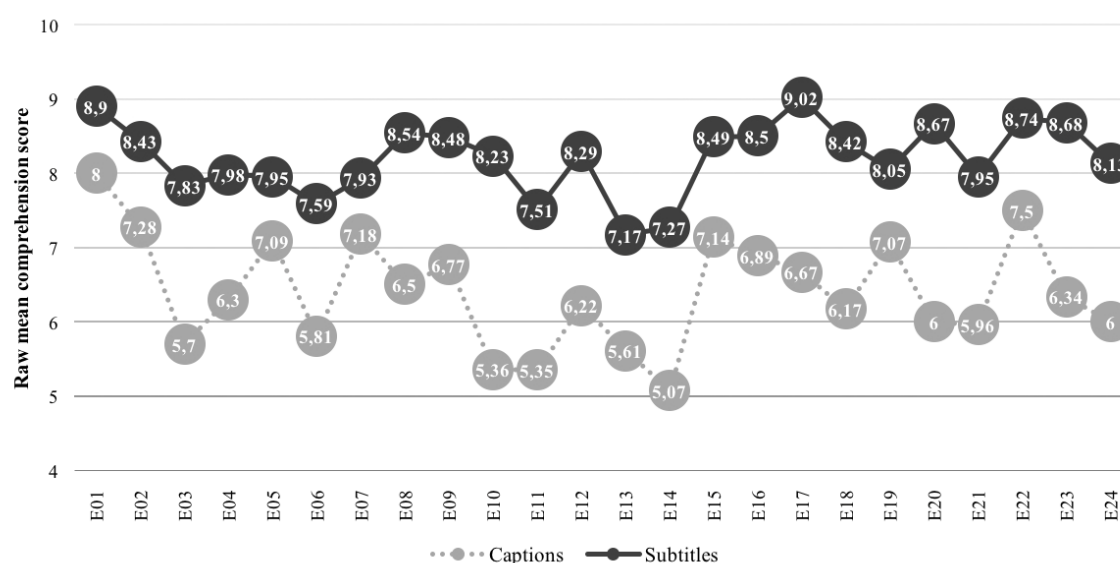
Discussion

This study explored TV viewing comprehension by adolescent learners through exposure to 24 episodes over a period of 8 months and, in particular, how comprehension was affected by the use of captions or subtitles, alongside variables related to the instructional focus, the learner, the lexical coverage of the episodes and the test items.

Factors related to the intervention

Results showed that language of the on-screen text was a significant predictor of comprehension scores, with the subtitles group significantly outperforming the captions group, as expected. Previous studies have also showed the advantages of having L1 text for comprehension in audio-visual media (Bianchi & Ciabattoni, 2008; Birulés-Muntané & Soto-Faraco, 2016; Latifi et al., 2011; Lwo & Lin, 2012; Markham et al., 2001; Markham & Peter, 2003)[8]. Results also showed that having explicit instruction on target vocabulary had a small negative effect on overall comprehension – a drawback also found in previous studies (Lee, 2007) –, indicating that learners at this age and proficiency level may find it hard to split their attention between the two demands (VanPattern, 2002). This suggests that research assessing comprehension performance when participants are also asked to pay attention to language forms (e.g. vocabulary, grammar) might need to take into account the depleting effects that explicit attention to specific aspects of the language might have on students' performance. Yet, it has also been found that – in this context – directing learners' attention towards target vocabulary renders significant improvement in vocabulary recall (Pujadas & Muñoz, 2019), hence trade-offs between content comprehension and learning specific language aspects, such as vocabulary, deserve further attention and exploration.

Factors related to the learner

Although general proficiency did not emerge as a predictor in our study, it was found that vocabulary size was positively related to comprehension scores, with larger vocabulary related to higher comprehension. This falls in line with results from other studies that also found that learners' vocabulary knowledge was a good predictor of comprehension (e.g. Vulchanova et al., 2015; Montero-Perez, Peters, Clarebout & Desmet, 2014, Rodgers, 2013) – which concurs with findings on vocabulary acquisition research (e.g. Peters & Webb,

2018). The interaction between vocabulary size and language indicated, however, that vocabulary size was a significant predictor only in the captions condition. This may suggest that learners in the subtitles condition relied more on reading the L1 text than on listening to the L2 audio (Steward & Pertusa, 2004; Vandergrift, 2007), thus making prior L2 lexical knowledge less relevant for comprehension when having subtitles available. This finding underlines the value of L1 subtitles as a scaffold for lower-proficiency learners to access multimodal authentic input.

Our results partially concur with those of Lwo and Lin (2012), who found that Grade 8 learners – of same age as our participants – benefitted better from subtitles than captions. In their study, however, Lwo and Lin acknowledged that the subtitles group was more proficient, a setback not found in the present study, in which both groups were comparable in terms of initial proficiency and vocabulary size. On the other hand, these results contrast with results from two previous studies with adolescent learners, in which it was found that there were no significant differences between language conditions (Vulchanova et al., 2015), and that the on-screen text had a distracting effect for more advanced students (Bairstow & Lavaur, 2011). Yet, participants in those studies were older and, probably, more proficient. It is possible that, with an increase in proficiency and vocabulary size, the difference between our language groups would have been smaller.

The other three learner-related factors – familiarity with viewing OV, attention to, and enjoyment from the TV series – did not appear to predict comprehension outcomes. It is likely that participants in our sample did not have as much prior experience viewing OV as for this factor to have a significant effect on comprehension scores. Although attention and enjoyment did not emerge as significant predictors, both were significantly higher in the subtitles groups. It might have been the case that language condition overpowered these other

two parameters in the analysis, or it might be that higher attention to and enjoyment from the TV series were a result of the language condition – they were higher *because* learners were viewing the programmes with the L1 text.

Factors related to the episodes and the test items

Another parameter that had a significant effect on comprehension was the episodes' lexical coverage, which has been shown to be a strong predictor in past research in listening comprehension (e.g. Hu & Nation, 2000) and video comprehension (e.g. Rodgers, 2013). Even if the difference between episodes was relatively small, episodes with higher lexical coverage had a higher percentage of correct responses. While the complexity of the plot or the familiarity with the topic of individual episodes (which sometimes included culture-bound references such as Thanksgiving) might have also played a role, episode lexical coverage appears to be a reliable and robust predictor for comprehension, independently of other factors such as on-screen text language or learners' vocabulary size. The fact that no clear pattern of improvement could be observed over time also suggests that comprehension was indeed episode-dependent. On the other hand, it is interesting to note that, independently of the language of the on-screen text, 74% of students reported (in the end-of-intervention questionnaire) that they understood the series better by the end of the intervention, suggesting that their comprehension was starting to improve though this was not yet detected by the measures used in the study.

Regarding test characteristics, item format was revealed to predict comprehension scores, with TF items having more correct responses than MC items. Yet, once language of the on-screen text was taken into account, the difference was only significant for the captions groups, suggesting that the availability of the L1 rendered item format unimportant. The

language of the on-screen text also mediated responses by item type. While overall comprehension scores were not affected by item type, once language of the on-screen text was taken into account, for the subtitles groups it was found that recalling textually explicit information was easier than recalling inferential information. This falls in line with findings in the listening research literature showing that processing scattered information is harder than recalling information from just one location, and that recalling exact content tends to be easier than recalling the gist or main idea (Buck, 2001). However, for the captions group it was the inferential items that received significantly more correct responses. It could be the case that – for the captions group – answering textually explicit items demanded that learners understood details that they might have missed due to the fast speech rate of the series and their low L2 linguistic skills (i.e., they could not use bottom-up processing successfully), whereas for inferential items the fact that they can gather information from different parts compensates for a missed piece of data (i.e., they were more successful at using top-down processing).

Conclusions

This study contributes to the area of foreign language learning through audio-visual input with results from a unique extensive classroom intervention. It is the first study analysing learners' exposure to authentic input over an extensive period of 8 months and to include vocabulary instruction and language of the on-screen text as mediating variables in comprehension. It is also one of the few studies using several full-length TV programmes. The results of the present study confirm previous findings in the field regarding the higher efficiency of L1 subtitles over L2 captions for content comprehension at this level of proficiency, while corroborating the importance of vocabulary size when L2 captions are

present – more demanding than subtitles for beginner-level learners. The study also suggests that explicit attention to target vocabulary items may have depleting effects on comprehension scores, which underlines the need to align the cognitive demands of tasks to learners' processing skills. Another valuable finding of the study concerns the interaction between item type and language of the on-screen text, suggesting that learners process textually explicit information and inferential information differently depending on the support they receive from the language available on the screen. The influence of item format and item type on comprehension also highlights the importance of taking into account item-related characteristics in the analysis. Finally, results corroborate the key role of lexical coverage as a strong predictor of comprehension, in line with findings from prior corpus-driven research and the few experimental studies existing in this area (e.g. Rodgers, 2013).

This study has several pedagogical implications. Firstly, it demonstrates the advantage of subtitles over captions for comprehension in a context with limited exposure to English and for adolescent participants with limited proficiency in the target language (average level between A1 to A2) and limited vocabulary size (around 2,000 words). Only students with larger vocabulary size could cope with captions. In a school classroom setting – where students may have different levels of L2 proficiency – the use of subtitles would engage the weakest students at the beginning while offering all learners the benefits of listening to authentic input and raising their motivation. It might be worth contemplating the possibility of combining both types of on-screen text support, moving gradually from subtitles to captions as learners get used to the characters, the voices and the overall topic of the TV series. Secondly, findings suggest that teachers also need to be aware of the fact that focusing on vocabulary may have a detrimental effect on content comprehension. However, pre-teaching vocabulary has been shown to be an effective way of acquiring it (Pujadas & Muñoz, 2019), and increasing vocabulary is – ultimately – an efficient way of supporting

comprehension of audio-visual input. Thirdly, the association of comprehension and input lexical coverage suggests the need to align the vocabulary load of the audio-visual input to learners' language skills. Finally, findings fit the principles of extensive viewing outlined by Webb (2015) (e.g. listening comprehension was supported by captions, subtitles and pre-teaching activities; input was of an appropriate level), and endorse TV viewing in the classroom as a starting point for extensive viewing out of the school.

The study has some limitations that need to be acknowledged. First of all, although there was a considerable number of observations, our sample size was relatively small. Also, the school setting made it not feasible to have a control group. Another shortcoming was that attention to and enjoyment from the TV series were self-reported and only assessed at the end of the intervention. Finally, the study did not take into account other factors that might have had an influence on comprehension, such as learners' working memory, the topic complexity of the individual episodes, the role of the imagery or the location of the necessary information within the episode – items tend to be easier when the information is presented at the beginning or when it is repeated (Buck, 2001). Further research including these variables, other TV genres and other proficiency levels would provide valuable information and help obtain a more comprehensive picture of the factors involved in TV viewing comprehension.

---

[1] Three non-consecutive episodes were skipped because school teachers considered they contained inappropriate scenes for 13-year-olds, or because they did not contain enough frequent vocabulary to teach in the groups who received vocabulary instruction. However, the missing episodes did not hinder the understanding of the overall story arch.

[2] Nation (2006) suggested that proper nouns have a minimal learning burden. Also Webb and Rodgers (2009b) showed in their study of the lexical coverage of movies that if learners knew proper nouns and marginal words (e.g. ah, oh, huh) they could reach 95.76% coverage with the most frequent 3,000 word families. In the present study, proper nouns make up 3.11 % of the running words, adding more coverage than words from the 3,000

word-family (1.62%). Considering that characters and locations reoccur throughout the episodes, it seems safe to assume learners are familiar with most of the proper nouns.

[3] The RANGE software (used to analyse the lexical coverage of the input) and the X_Lex test (used as the measure for learners' vocabulary size) are based on different word-lists, but a validation study by Miralpeix (2012) has shown that the results of the Levels Test and the X_Lex are comparable. Also, while the RANGE software indicates the coverage by word-level, the X_Lex test provides a total score of out 5,000 words (by adding up the knowledge in each of the 5 word-bands). However, it seems logical to assume that – out of a score of 2,000 words – most of the words would be from the first or second thousand word-bands, even if some of the words do indeed belong to the fourth or fifth bands (Miralpeix, personal communication, December 18, 2018). An analysis of the X_Lex results in each word band in the present study also revealed this tendency (75% of the words were from the first three bands).

[4] The X_Lex test is a computerized test in check list format that measures learners' L2 receptive vocabulary knowledge of the most frequent 5,000 words in the language. The test randomly presents 120 items selected from the first five frequency bands, plus a number of non-words to control for guessing.

[5] Because comprehension tests were in Spanish, textually explicit questions were formulated using paraphrases and synonyms rather than literal excerpts from the audio, since this might have prompted learners in the subtitles group to just choose the options containing the vocabulary they were seeing in the subtitles (Taylor, 2005). With this, we tried to avoid lexical overlap, which tends to be a predictor of easy items, since test-takers tend to select options that contain vocabulary that they recognize from the input (Buck, 2001). This is also why the term "textually explicit items" is preferred over "literal items".

[6] Note that participants were taking part in a larger classroom-based intervention and completed a series of tasks unrelated to comprehension that will not be addressed here (see Pujadas & Muñoz, 2019).

[7] This was further confirmed by the results of a linear regression showing that the percentage of variance explained by this factor was extremely small and insignificant.

[8] Although it would seem that having access to the native language would lead to 100% understanding of the dialogue, research shows that it is uncommon. In studies comparing the use of captions and subtitles – at different levels of proficiency –, the mean comprehension for the subtitles groups were 67% (Markham & Peter (2003), 72% (Bianchi & Ciabattoni, 2008), 72% (Latifi, et al., 2011), 82% (Markham et al., 2001), and 93% (Birulés-Muntané & Soto-Faraco, 2016). There may be several reasons for this (e.g., factors related to the level of detail of the questions, working memory), but this issue falls beyond the scope of the study.

References

Almeida, P., & Costa, P. (2014). Foreign language acquisition: the role of subtitling. *Procedia-Social and Behavioral Sciences*, *141,* 1234-38.

Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, *33*(2), 206-19.

Bairstow, D., & Lavaur, J.M. (2011). Audiovisual information processing by monolinguals and bilinguals: effects of intralingual and interlingual subtitles. *Approaches to Translation Studies*, *36*, 273-93.

Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review*, *50*(3), 507-31.

Bianchi, F., & Ciabattoni, T. (2008). Captions and subtitles in EFL learning: an investigative study in a comprehensive computer environment. In Baldry, A., Pavesi, M., & Taylor Torsello, C. (Eds.) *From didactas to ecolingua* (pp. 69-90). Edizioni Università di Trieste.

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PloS One, 11*(6): e0158409.

Bisson, M., Van Heuven, W., Conklin, K., & Tunney, J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning, 64*(4), 855-77.

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

Chang, A., & Read, J. (2006). The effects of listening support on the listening performance of EFL learners. *TESOL quarterly*, *40*(2), 375-97.

Charles, T., & Trenkic, D. (2015). Speech segmentation in a second language: The role of bimodal input. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning: Principles, strategies and practical experiences*, (pp. 173-198). Bern: Peter Lang

Cheng, H. (2004). A comparison of multiple-choice and open-ended response formats for the assessment of listening proficiency in English. *Foreign Language Annals*, *37*(4), 544-53.

d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In Rayner K. (Eds.) *Eye movements and visual cognition* (pp. 415-427). Springer Series in Neuropsychology.

Danan, M. (2004). Captioning and subtitling: undervalued language learning strategies. *Meta: Journal Des Traducteurs, 49*(1), 67-77.

Davey, B., & McBride, S. (1986). Effects of question-generation training on reading comprehension. *Journal of Educational Psychology*, *78*(4), 256-62.

Del Rincón, D., Arnal, J., Latorre, A., & Sans, A. (1995). *Técnicas de investigación en ciencias sociales*. Madrid: Dykinson.

Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, *100*(S1), 19-47.

Durbahn, M., Rodgers, M., & Peters, E. (2019). The relationship between vocabulary and viewing comprehension. Advance online publication: Doi: https://doi.org/10.1016/ j.system.2019.102166

Elkhafaifi, H. (2005). The effect of prelistening activities on listening comprehension in Arabic learners. *Foreign Language Annals, 38*(4)*,* 505-13.

Ellis, N. (2013). Second language acquisition. In Trousdale, G., & T. Hoffmann (Eds.) *Oxford Handbook of Construction Grammar* (pp. 365-378). Oxford: Oxford University Press.

European Commission (2011). Study on the use of subtitling: the potential of subtitling to encourage foreign language learning and improve the mastery of foreign languages. Final Report.

European Commission. (2017). Media use in the European Union. European Union, standard Eurobarometer 88 Project.

Gass, S., Winke, P., Isbell, D., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology 23*(2), 84-104.

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, *40*(2), 97-118.

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System*, *36*(1), 85-93.

Guillory, H. (1998). The effects of keyword captions to authentic French video on learner comprehension. *CALICO Journal, 15*(1-3), 89-108.

Hasan, A. (2000). Learners' perceptions of listening comprehension problems. *Language, Culture and Curriculum*, *13*(2), 137-53.

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), 181-92.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403-30.

Huang, H., & Eskey, D. (1999). The effects of closed-captioned television on the listening comprehension of intermediate English as a second language (ESL) students. *Journal of Educational Technology Systems*, *28*(1), 75-96.

Jones, L. (2003). Supporting listening comprehension and vocabulary acquisition with multimedia annotations: The students' voice. *CALICO Journal*, *21*(1)*, 41-65.

Jones, L., & Plass, J. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, *86*(4), 546-61.

Kelley, T. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*, 17-24.

Khan, N., Kasdar, J., Melvin, M., Blomquist, R., Huang, E., & McEwen, J. (2015). Fresh off the Boat [TV series] Los Angeles, CA: ABC

Krashen, S. (2003). *Explorations in Language Acquisition and Use*. Portsmouth, NH: Heinemann

Latifi, M., Mobalegh, A., & Mohammadi, E. (2011). Movie subtitles and the improvement of listening comprehension ability: Does it help? *The Journal of Language Learning and Teaching*, *1*(2), 18-29.

Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook,* 5(1), 223-50.

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a foreign language*, *22*(1), 15-30.

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.

Lwo, L., & Lin, M. (2012). The Effects of Captions in Teenagers' Multimedia L2 Learning. *ReCALL, 24*(2), 188-208.

Maleki, A., & Safaee Rad, M. (2011). The effect of visual and textual accompaniments to verbal stimuli on the listening comprehension test performance of Iranian high and low proficient EFL learners. *Theory and Practice in Language Studies*, *1*(1), 28-36.

Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, *32*(3), 321-28.

Markham, P., & Peter, L. (2003). The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *Journal of Educational Technology Systems*, *31*(3), 331-41.

Markham, P., Peter, L., & McCarthy, T. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals*, *34*(5), 439-45.

Matielo, R., de Oliveira, R., & Baretta, L. (2017). Intralingual subtitles, interlingual subtitles, and video comprehension: insights from an exploratory study. *Letrônica*, *10*(2), 758-74.

Mayer, R. (1997). Multimedia learning: Are we asking the right questions? *Educational psychologist*, *32*(1), 1-19.

Mayer, R. (2014). *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.

Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52.

Meara, P., & Milton, J. (2003). *X_Lex: the Swansea levels test*. Berkshire: Express Publishing.

Miralpeix, I. (2012, March). *X_Lex and Y_Lex: A validation study*. Paper presented at the 22nd Lexical Studies Conference, Newtown, UK.

Mitterer, H., & McQueen, J. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PloS one*, *4*(11): e7785.

Montero-Perez, M., Peters, E., & Desmet, P. (2014). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL*, *26*(1), 21-43.

Montero-Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*(1), 118-41.

Montero-Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*(3), 720-39.

Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System*, *67*, 77-86.

Muñoz, C. (in press) Boys like games and girls like movies. Age and gender differences in out-of-school contact with English. *Revista Española de Lingüística Aplicada*.

Nation, P. (2006). How large a vocabulary is needed for reading and listening?. *Canadian modern language review*, *63*(1), 59-82.

Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, *1*(1), 2-13.

Nation, P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts [software]. Retrieved from: https://www.victoria.ac.nz/lals/about/staff/paul-nation

Nation, P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking. ESL & Applied linguistics professional series*. Abingdon: Routledge.

Neuman, S., & Koskinen, P. (1992). Captioned television as comprehensible Input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly, 27*(1), 94-106.

Noreillie, 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. *ITL-International Journal of Applied Linguistics*, *169*(1), 212-31.

Ockey, G. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing, 24*(4), 517-37.

Paivio, A. (1986). *Mental representations: a dual coding approach*. Oxford: Oxford University Press.

Paivio, A. (2008). Mind and its evolution: a dual coding theoretical approach. *Psychol Rec*, *59*, 295-300.

Pearson, P., & Johnson, D. (1978) *Teaching reading comprehension*. New York: Holt, Rinehart, & Winston.

Peters, E. (2019), The Effect of Imagery and On-Screen Text on Foreign Language Vocabulary Learning From Audiovisual Input. *TESOL Q, 53*, 1008-1032.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551-77.

Pujadas, G. (2019) *Language learning through extensive TV viewing. A study with adolescent EFL learners* (Unpublished doctoral dissertation). University of Barcelona, Barcelona.

Pujadas, G., & Muñoz, C. (2017, April) *Learning through subtitles. Learners' preferences and task perception*. Paper presented at the 2017 International Conference on Task-Based Language Teaching, Barcelona, Spain.

Pujadas, G., & Muñoz, C. (2018, October) *What words do we learn better through TV series? The effects of age, proficiency and type of instruction*. Paper presented at the Second Language Research Forum, Montreal, Canada.

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: a study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, 1-18.

Rodgers, M. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions* (Unpublished doctoral dissertation). Victoria University of Wellington.

Rodgers, M. (2016). Extensive listening and viewing: the benefits of audiobooks and television. *The European Journal of Applied Linguistics and TEFL, 5*(2), 43-57.

Rodgers, M. (2018a). The images in television programs and the potential for learning unknown words. *Approaches to Learning, Testing, and Researching L2 Vocabulary ITL - International Journal of Applied Linguistics 169*(1), 191-211.

Rodgers, M. (2018b, October). *The effects of captions on facets of viewing comprehension of an authentic television program.* Paper presented at the Second Language Research Forum, Montreal, Canada.

Rodgers, M., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly, 45*(4), 689-717.

Rodgers, M., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English-language television programs. *CALICO Journal*, *34*(1), 20-38.

Schmitt, N., & Carter, R. (2000). The lexical advantages of narrow reading for second language learners. *TESOL Journal*, *9*, 4-9.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language testing*, *1*(2), 147-70.

Shohamy, E., & Inbar, O. (1991). Validation of listening comprehension tests: The effect of text and question type. *Language testing*, *8*(1), 23-40.

Stewart, M., & Pertusa, I. (2004). Gains to language learners from viewing target language closed-captioned films. *Foreign language annals*, *37*(3), 438-42.

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.

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

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals*, *38*(3), 422-27.

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), 457-79.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language teaching*, *40*(3), 191-210.

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal*, *42*(4), 272–281.

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), 235-50.

Vanderplank, R. (2016b). *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning*. New York: Springer.

Vanderplank, R. (2019). 'Gist watching can only take you so far': attitudes, strategies and changes in behaviour in watching films with captions. *The Language Learning Journal*, 1-17.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research.* New Jersey: Ablex.

VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, *52*(4), 755-803.

Vulchanova, M., Aurstad, L. M., Kvitnes, I. E., & Eshuis, H. (2015). As naturalistic as it gets: subtitles in the English classroom in Norway. *Frontiers in Psychology*, *5*, 1510.

Wagner, E. (2002). Video listening tests: A pilot study. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, *2*(1), 1-39.

Webb, S. (2010). Pre-learning low-frequency vocabulary in second language television programmes. *Language Teaching Research*, *14*(4), 501-15.

Webb, S. (2011). Selecting television programs for language learning: Investigating television programs from the same genre. *International Journal of English Studies*, *11*(1), 117-35.

Webb, S. (2015). Extensive viewing: Language learning through watching television. In Nunan, D., & Richards, J. (Eds.) *Language learning beyond the classroom* (pp. 175-184). Abingdon: Routledge.

Webb, S., & Rodgers, M. (2009a). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335-66.

Webb, S., & Rodgers, M. (2009b). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407-27.

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

Appendix

Examples of textually explicit items (TE), in multiple-choice (MC) and true-false (TF) format

With what does Eddie trip over when he breaks his arm?

        a)   With a mechanical bull.

        b)   With a white rug.

        c)   With a cord.

SCRIPT: *"Okay! Hey, boys, we can't let mom know **Eddie tripped over the cord of the mechanical bull and broke his arm.** (…) Now tell these nice people the truth about how you broke your arm. Eddie. You were at Cattleman's **and you tripped... On the cord to the new mechanical bull."***

T / **F**   Neither Jessica nor Louis knew that to be a relator you needed a licence.

SCRIPT: *"[Jessica] apparently, you need a license to sell houses! Did you know this?! [**Louis**] **I did know this. I told you this, and you ignored me.**"*

Examples of inferential items (IN), in MC and TF format

Why do Jessica and Louis ask Nicole to babysit their children?

        a)   Because Eddie asked for it to be able to talk to her.

        b)   Because their gramma is too old to take care of the children.

        c)   **Because Nicole is the cheapest option.**

SCRIPT: *"[Eddie] Babysitting? Why can't grandma watch us? [Jessica] Because last time, she charged me $100 an hour. All right, we've got to get going. Nicole, some ground rules... no scary movies."*

T / F   Evan has a sever gambling problem

SCRIPT: *"[Eddie:] I can't believe you lied to me. I thought you were getting picked on. [Evan:] I'm addicted, okay? I'm addicted to playing Pogs. I don't even get what you see in that noise. I love the thrill of taking risks, laying it all on the line."*