UNIVERSITAT DE BARCELONA

L2 learning from captioned-video viewing in primary school students

Daniela Isabel Avello Garcia

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

UNIVERSITAT DE BARCELONA

L2 learning from captioned-video viewing in primary school students

Doctoral dissertation

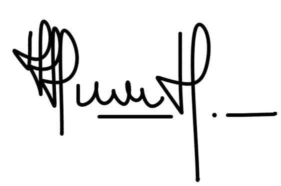
Daniela Isabel Avello García

Doctoral program: Cognitive Science and Language

Theoretical and Applied Linguistics line of research

Thesis director and supervisor: Dr. Carmen Muñoz Lahoz

2022



Abstract

An increasing number of investigations have been conducted in the last decade to explore the effects of audiovisual input on L2 learning (Montero Perez, 2022; Muñoz, 2022). Nevertheless, primary school L2 learners are still an under-researched age group in comparison with university and secondary school students (Montero Perez & Rodgers, 2019). The literature suggests that L1 subtitles would be a more suitable option for primary school learners due to their lower L2 proficiency level and still developing L1 reading skills. However, the question on how suitable and beneficial captioned-video viewing may be for primary school learners from input-limited contexts still needs further evidence to be answered. The present study attempted to fill the gaps in the literature as regards the extent to which extensive captioned-video viewing (11 episodes) fostered vocabulary learning (written-word form recall, and written-word form and meaning recognition) and the development of receptive language skills in six groups of EFL primary school students from Chile (*n*=120; 9-11 years old; years 4 and 5). More specifically, we studied the influence of treatment, learner and input-related factors on the results. To start with, the experimental groups differed in terms of viewing distribution (number of episodes watched a week) and the activities completed at the end of each session (meaning-focused vs. construction-focused activities). As for learners' characteristics, this study assessed the influence of a group of cognitive and language-related factors (L1 and L2) on the results. Finally, we explored the extent to which a set of context and word characteristics predicted vocabulary learning.

The findings that emerged from the statistical analyses were interpreted in light of the literature and also of the learners' perceptions of the viewing experience. Overall, the results revealed significant improvement in vocabulary learning and the development of receptive language skills in both year levels. Still, the treatment appeared to be especially beneficial for fifth graders due to their significantly higher proficiency level in both languages and, possibly, to their cognitive maturity. Notwithstanding this result, the analyses also showed that fourth graders' performance was enhanced by the implementation of shorter lags between episodes. Additionally, the data on learners' perceptions of the treatment provided rich evidence on how the participants from both year levels took advantage of the different modalities to compensate for their knowledge gaps. On the whole, the findings reported in

this dissertation suggest that the use of captioned videos may be suitable and conducive learning in both year levels as long as some specific factors are considered.

Resumen

A través de los años ha incrementado el número de investigaciones que han explorado los efectos del input audiovisual en el aprendizaje de una lengua extranjera (Montero Perez, 2022; Muñoz, 2022). Sin embargo, los estudiantes de primaria han recibido una menor atención en comparación a los alumnos universitarios y de secundaria (Montero Perez & Rodgers, 2019). La literatura sugiere que los subtítulos en la L1 serían una mejor alternativa para los estudiantes de primaria debido a su bajo nivel de proficiencia en la L2. Sin embargo, aún se necesita determinar cuán apropriados y beneficiosos podrían ser los videos subtitulados en la L2 para los estudiantes de primaria provenientes de contextos de menor contacto con el idioma. Por lo tanto, esta investigación se diseñó con el objetivo de contribuir a la literatura en relación a la capacidad de los videos subtitulados en la L2 (11 episodios) de promover el aprendizaje de vocabulario y el desarrollo de habilidades receptivas en seis grupos de estudiantes de primaria de Chile (n= 120; 9-11 años de edad; cuarto y quinto año). Específicamente, se investigó la influencia de diversos factores en los resultados, los cuales estaban asociados tanto al tratamiento como a las características de los estudiantes y del input. En primer lugar, los grupos experimentales se diferenciaron por la distribución de las sesiones (número de episodios vistos a la semana) y los tipos de actividades que los estudiantes debían completar después de cada video (enfoque en comprensión o en construcciones lingüísticas). En relación a las características de los estudiantes, se evaluó la influencia de una serie de factores cognitivos y lingüísticos (L1 y L2) en los resultados. Finalmente, también se exploró la medida en que el aprendizaje de vocabulario era explicado por un grupo de características asociadas a las palabras y a su contexto.

Los resultados que se obtuvieron de los análisis estadísticos se interpretaron en base a la literatura y a las percepciones de los estudiantes respecto a la intervención. En resumen, los resultados indicaron que los estudiantes de ambos niveles mejoraron significativamente en vocabulario y el desarrollo de habilidades receptivas. Sin embargo, los estudiantes de quinto de primaria parecieron obtener mayores beneficios del tratamiento, los cuales podrían ser asociados a su mayor nivel de proficiencia en español y en inglés, y posiblemente a su

mayor madurez cognitiva. A pesar de estos resultados, los análisis también demonstraron que el desempeño de los estudiantes de cuarto de primaria fue potenciado con la implementación de una menor distancia entre episodios. Adicionalmente, las percepciones de los estudiantes sobre el tratamiento demonstraron que los estudiantes de ambos niveles utilizaron las distintas modalidades para compensar por su falta de conocimiento sobre la L2. En general, los resultados publicados en esta tesis sugieren que los videos subtitulados en la L2 podrían ser apropiados y beneficiosos en ambos niveles siempre y cuando se consideren diversos factores en su implementación.

Acknowledgements

First of all, I would like to express my deepest gratitude to Dr. Carme Muñoz for her infinite patience, support and advice along this learning journey. I consider myself extremely fortunate to have had the opportunity of learning from such an exemplary researcher and the outstanding research group she leads. I am deeply indebted to the GRAL research group for their constant encouragement throughout the process. Also, I would like to thank the group of predoctoral researchers for their kindness and support, especially during the difficult moments.

I would also like to extend my thanks to Dr. Annamaria Pinter for her time and valuable advice during my stay at Warwick. I also wish to thank Dr. Raquel Serrano, Dr. Imma Miralpeix and Dr. Andrea Révész for their insightful suggestions in early stages of this investigation. I would also like to express my gratitude to Matthew Pattemore for accepting to record the audios used in the tests. In addition, I would like to recognize the assistance of all those researchers who took the time to reply my e-mails and provide detailed answers to my questions, as well as those researchers who showed interest in my work and provided valuable feedback at the end of each presentation.

I am deeply indebted to the following people, without whom I would not have been able to complete this investigation. Words cannot express how much I appreciate the excellent disposition of the children who participated in this study as experimental, control and pilot groups. Special thanks to the teachers and school authorities who trusted in my work and made me feel welcome as if I had been part of the staff. Likewise, I very much appreciate the support of the teachers from school 2, whose work load was increased as a result of this investigation. I also owe special thanks to all the families involved in this study for giving their consent and showing a supportive attitude during the intervention. Last but not least, I would like to acknowledge the help of the teachers who volunteered to participate with their class in the pilot testing of some of the instruments.

From the bottom of my heart I would also like to thank the unwavering support of the most important people in my life: my daughter and my husband. I will never thank them enough for allowing me to pursue my dream and helping me throughout the process. Finally, I would like to thank my family and all those friends who were always sending me encouraging words and positive thoughts from a distance.

Table of contents

List of figures

List of tables

Introduction

There is a broad consensus among researchers that quality, quantity and intensity of exposure to a target language play a significant role in foreign language learning (Graham et al., 2017; Muñoz, 2008). Yet, despite the sound evidence that has emerged from input-limited contexts such as Spain on this matter (e.g. Muñoz, 2006), there still seems to be a mismatch between research and the short-term policies implemented by local authorities worldwide (Enever, 2018). Policies have, for instance, inclined towards an early start as the key step to achieve higher proficiency in the L2 (Enever, 2018), whereas the literature has shown that formal instruction alone may be insufficient to significantly improve learners' outcomes over the years (Peters et al., 2019). To illustrate, the estimated number of words learned and retained after 6-9 years of instruction may not surpass the 2,000 words, which are not enough to comprehend and use the language in a wide variety of contexts (Webb, 2020).

The growing literature on out-of-school L2 learning (e.g. De Wilde et al., 2019, Lindgren & Muñoz, 2013; Muñoz et al., 2018) has contributed to this debate by demonstrating that the attention should not only be placed on formal instruction but also on the activities that learners engage with outside the classroom (Peters et al., 2019). The evidence has shown that higher out-of-school L2 contact (e.g. video viewing) leads to significant gains in diverse language aspects and skills, such as reading and listening (e.g. Lindgren & Muñoz, 2013), vocabulary knowledge (e.g. De Wilde et al., 2022; Puimège & Peters, 2019a), and grammar (e.g. Muñoz et al., 2018). Therefore, classroom instruction and out-of-school contact may complement each other to boost learners' outcomes in the L2 (Peters et al., 2019). Nonetheless, it is important to consider that the extent to which learners engage with activities in the L2 outside the classroom seems to vary as a function of context (e.g. Lindgren & Muñoz, 2013), L1 and L2 distance (e.g. De Wilde et al., 2022) and learners' characteristics, such as age and gender (Muñoz, 2020b). The evidence to date has shown that primary school learners from input-limited contexts (e.g. dubbing countries) have little contact with the target language at early stages, and it gradually increases over the years (Marzá & Torralba, 2015; Muñoz, 2020b).

Different investigations have consistently reported that viewing is one of the most popular out-of-school activities, which has been certainly found to be conducive to L2 learning (De Wilde et al., 2022; Koolstra & Beentjes, 1999; Lindgren & Muñoz, 2013;

Muñoz, 2020b; Muñoz et al., 2018). Therefore, in dubbing countries, the incorporation of audiovisual input and explicit training in the foreign language classroom may be a good way to encourage L2 learners to use these resources at home (Webb, 2015). However, there is still a lot of work to do in order to take full advantage of audiovisual input and integrate it in the educational system successfully (Donaghy, 2019; Vanderplank, 2016; Zabalbeascoa et al., 2015). An increasing number of investigations have been conducted in the last decade to explore the actual effects of audiovisual input on L2 learning. Overall, the results have consistently shown that L2 learners benefit from captioned-video viewing given that the use of print seems to make the input more accessible (Danan, 2015; Montero Perez, 2022; Vanderplank, 2016). Nevertheless, further research is still required to identify the factors that influence and maximize learning, such as methods, context, input and learner characteristics (Gambier, 2015). Yet, only a handful of studies on captioned-video viewing have been conducted longitudinally (e.g. Pujadas & Muñoz, 2019), and very little attention has been paid to primary school L2 learners in comparison with university and secondary school students (Montero Perez & Rodgers, 2019). Most of the investigations with primary school L2 learners have tested the effectiveness of L1 subtitles (e.g. Black, 2020, 2022; d'Ydewalle & Van de Poel, 1999; Gesa, 2019; Koolstra & Beentjes, 1999), while the use of captions has only been explored by a handful of studies, which have mainly been conducted with late primary school students (e.g. Teng, 2019a, 2019b, 2021, 2022).

Hence, this dissertation attempted to fill the gaps in the literature by exploring the extent to which extensive captioned-video viewing promoted L2 learning in primary school students from an input-limited context where English is taught as a foreign language (EFL). Specifically, it studied the influence of a set of input-, treatment- and learner-related factors on the outcomes. According to the field of developmental psychology, the EFL primary-school learners that participated in this study (9-11 years old) were at the stage of middle childhood, which ranges from the age of 6 to 11/12 years old (Harris & Westermann, 2015; Myles et al., 2019). This is a stage of big changes in social and cognitive skills, where there may also be great variability among individuals. Middle childhood is characterized by the development of L1 literacy skills, as well as the gradual increase in (language) learning awareness and the capacity to learn more explicitly, which influence the L2 teaching and learning process (Myles et al., 2019). Therefore, the literature review will mainly focus on this age range whenever possible.

The literature review is divided into four main sections. Section 1.1 gives an account of the theoretical background and the key terms associated to audiovisual/multimodal input. Section 1.2 presents the main findings concerning L2 learning through audiovisual/multimodal input, while section 1.3 focuses on the role of learner-related factors in L2 learning. Finally, section 1.4 analyzes the role of treatment- and input-related factors in L2 learning.

I. Literature review

1.1Multimodal/audiovisual input: Key terms and theoretical background

In this dissertation, the term multimodal input is used in its broader sense to refer to the type of input whose meaning is conveyed through the integration of different semiotic resources such as sounds, images, gestures and verbal information (written and/or aural text) (Jewitt et al., 2016). Therefore, books with illustrations and/or audio support (e.g. reading-while-listening), as well as videos are addressed as multimodal input. The term audiovisual input is constrained to the use of technological devices to access videos, which, apart from dynamic images and audio, may include the use of text support (e.g. captioned videos). The main types of audiovisual input reported in the literature are as follows (see Montero Perez, 2022):

a) Audio and text in the target language: L2 subtitles/(full)captions/intralingual subtitles/same-language subtitles.
b) Audio in the L2 and text in the L1: L1 subtitles, interlingual subtitles.
c) Audio in the L1 and text in the L2: reversed subtitles.
d) Audio in the L2 and key words in the L2: keyword captions.
e) Audio in the L2 and full captions where a set of target language constructions are highlighted: enhanced captions.

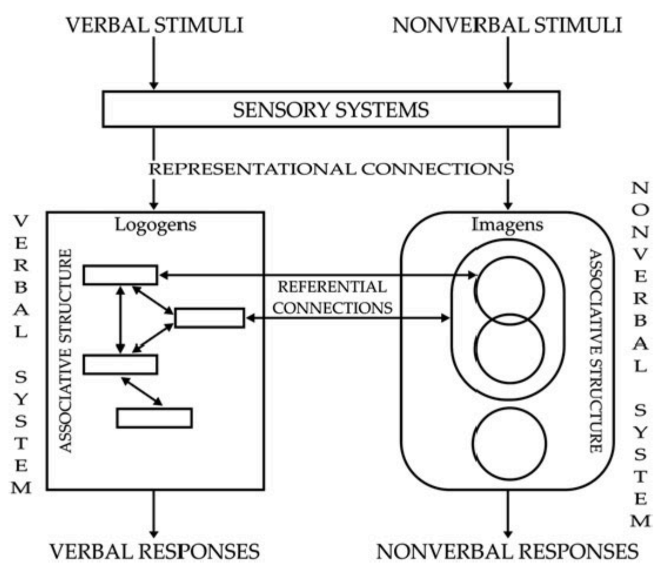
The literature on L2 learners' processing and learning from multimodal input draws on Paivio's Dual Coding theory (Paivio, 1986) and the theory of Multimedia Learning (Mayer, 2014, 2022). Thus, the following sections describe how these two theories that have originally been developed to explain L1 processing and content learning in diverse subject areas (e.g. science) may serve as theoretical support for the use of multimodal input in L2 learning.

1.1.1 Dual Coding theory

The Dual Coding theory (Paivio, 1986) claims that humans process verbal and non-verbal information through two different but interconnected channels (referential connections). Due to its important educational implications, this framework has increasingly been adopted to explain why the synergy between verbal and non-verbal information (e.g. imagery) may significantly support learning in diverse educational areas and daily life (Clark & Paivio, 1991; Paivio, 2014). On the whole, this theory claims that the simultaneous processing and encoding of verbal (logogens) and non-verbal (imagens) information generates dynamic referential and associative processes that foster learning and recall (Paivio, 2014) (see Figure 1). Thus, as Clark and Paivio (1991) emphasize, learning from both, verbal and non-verbal information seems to be better than "…learning from a verbal code alone" (p. 165).

Figure 1.

Structural model of Dual Coding: Representational units and their referential and associative interconnections



Note. Paivio (2014).

Within the verbal system, Clark and Paivio (1991) explain that the information may be encoded through different modalities such as visual, auditory and articulatory. Likewise, the non-verbal system includes a variety of sensory modalities such as imagery, sounds, actions, gestures and olfactory stimuli that may provide even richer information in a single instantiation. In fact, Clark and Paivio (1991) use the idiomatic expression 'pictures are worth a thousand words' to illustrate the extent to which the complexity of a single image may support comprehension, learning and further recall (p. 152). The later activation of the referential connections between the verbal and non-verbal systems works in both directions. When a word is activated, the referential connections activate the non-verbal system and its representation, whereas the encountering of an object activates its verbal representation. Furthermore, within each system, the associated structure activates representations that are connected to the stimulus (e.g. when the word 'apple' is encountered, other food items are also activated).

Clark and Paivio (1991) claim that the strength of the referential connections between verbal and non-verbal representations may be mediated by concreteness given that this factor determines the extent to which words and phrases are more or less difficult to evoke. Likewise, there may also be great interindividual variation when it comes to processing, visualizing and retrieving images. Therefore, some learners may need additional support during the imaging process (e.g. while reading a passage). By way of illustration, the study conducted by Center et al. (1999) with L1 young learners (second grade) showed that explicit visual imagery training prior to the exposure to aural input compensated for students' poor decoding skills to support comprehension, resulting in higher reading and listening skills at the end of the treatment.

Sadoski and Paivio (2013) also give a theoretical account of reading skills development through their Dual-theoretical model of reading. Overall, this theoretical model contends that the reading process is essentially multimodal (Sadoski, 2008; Sadoski & Paivio, 2007). Research on L1 reading skills has shown that the simultaneous processing of verbal and non-verbal input (e.g. using images or visualizations) does not only support text comprehension (Sadoski, 2008) but also text decoding and the learning of word spelling (e.g. Sadoski et al., 2004). Yet, at earlier stages, poor decoding skills may hinder comprehension and learning from multimodal input. To illustrate, the encountering of many unknown and graphophonemically irregular constructions may slow down the associational and referential processing between verbal and non-verbal codes. This is attributed to the fact that early readers focus their effort and attention on lower linguistic levels: phonological and orthographic, so they do not have enough cognitive resources available to process imagery and enhance comprehension (Sadoski et al., 2004; Sadoski & Paivio, 2013).

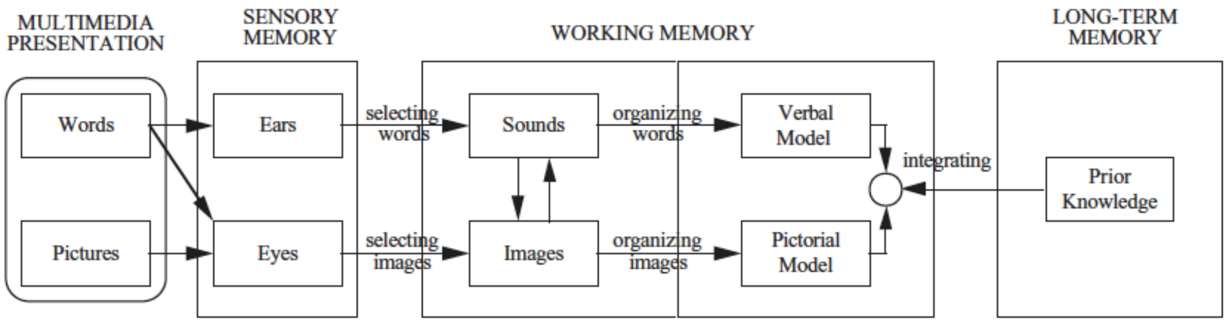
The literature on multimodal input has relied on the Dual Coding theory to explain why verbal input (written text and/or audio) and imagery reinforce each other, and their simultaneous processing may aid L2 learning. Consistent with this theory, the study conducted by Durbahn et al. (2020) indicated that the lexical demands of audiovisual input appear to be lower with the presence of supportive imagery than in listening-only condition, which might be explained by the fact that visual information fosters comprehension (i.e. the use of gestures, facial expressions, body language and concrete objects). Therefore, imagery may be addressed as a compensatory mechanism for lower L2 proficiency learners (Peters & Muñoz, 2020b; Wright, 2010). Nonetheless, as suggested by Sadoski and Paivio (2013), the evidence has also shown that lower L2 proficiency readers spend longer time processing print, so they may have less time and cognitive resources available to process images. Yet, the simultaneous processing of aural and written input seems to facilitate decoding to devote greater attention to images (e.g. Serrano & Pellicer-Sánchez, 2019). The processing of written and aural input, and images has also been studied in light of the Cognitive Theory of Multimedia Learning (Mayer, 2014), which is partly built on Dual Coding Theory. Some of its principles indicate which conditions are required to benefit from multimodal input.

1.1.2 Cognitive Theory of Multimedia Learning

The Cognitive Theory of Multimedia Learning (Mayer, 2014, 2022) claims that multimedia may foster learning effectively as long as the functioning of the human mind is considered. Therefore, the multimedia design should facilitate learning and prevent students' cognitive overload (Kalyuga & Sweller, 2014). As shown in Figure 2, the Cognitive Theory of Multimedia Learning consists of three cognitive processes that are essential for learning: selecting relevant verbal and pictorial information from the input, organizing the information in working memory to create coherent mental representations (Baddeley, 2007), and integrating these representations with previous knowledge (long-term memory). This theory draws on three basic assumptions: dual channels (Paivio, 1986), limited cognitive capacity (i.e. a limited amount of information can be processed simultaneously) (Baddeley, 2007; Kalyuga & Sweller, 2014), and active processing (i.e. cognitive engagement) (Mayer, 2022).

Figure 2.

Cognitive theory of multimedia learning



Note: (Mayer, 2022, p.62).

Mayer (2022) states that there are three kinds of demands in learners' cognitive system while processing multimodal input, from which a series of principles have also been elaborated (see Table 1). First of all, extraneous processing refers to the elicitation of cognitive procedures that are irrelevant to the target instructional objective, which is associated with poor materials design. Secondly, essential processing has to do with the complexity of the instructional materials. Therefore, learners should be capable of selecting the target information from the input to succeed at the following stages. Finally, generative processing has to do with learners' motivation and capacity to integrate verbal and pictorial information with their prior knowledge to create coherent representations and promote learning. Hence, when selecting, adapting, creating, and implementing multimodal resources, instructors should "…reduce extraneous processing, manage essential processing, and foster generative processing" (Mayer, 2022, p. 69). In this regard, a series of representative techniques (principles) have been formulated to enhance multimedia learning in a variety of subject areas (e.g science). Table 1 defines some of the principles that may be most relevant and/or controversial for L2 research on audiovisual input.

Table 1.

Cognitive theory of Multimedia learning: Some representative techniques

Representative techniques	Definition
Multimedia principle	Learning from words and pictures is more effective than learning from words alone.
Coherence principle	People learn better from multimedia instructional messages when extraneous words and images are excluded.
Signaling principle	People learn better when cues are added that highlight or spotlight the key information in a multimedia lesson and its organization.
Redundancy principle	People learn better when the same information is not presented in more than one format.
Split-attention principle	People learn better when words and pictures are physically and temporally integrated.
Temporal-contiguity principle	People learn better when narration and corresponding graphic appear simultaneously.
Spatial-contiguity principle	People learn better when printed text is placed near corresponding graphic.
Pre-training principle	People learn better from a multimedia message when they know the names and characteristics of the main concepts.
Modality principle	People learn better from a multimedia message when the words are spoken rather than written.
Personalization principle	People learn better when the words of a multimedia presentation are in conversational style rather than formal style.
Image principle	People do not necessarily learn better when the speaker's static image is on the screen.
Animation principle	People learn better from dynamic graphics than static graphics.

Note. (Mayer & Fiorella, 2022, pp. 3-16).

The evidence on L2 learning through multimodal input does not seem to support the redundancy and modality principles. Therefore, considering that this theory was primarily developed to explain content learning, the applicability of its principles may not necessarily be extrapolated to L2 learning. The literature suggests that listening comprehension may be a daunting challenge for L2 learners given that spoken input seems to be hard to decode under online processing pressure (Kormos et al., 2019; Newton & Nation, 2021). As a result, empirically-based suggestions have been raised to foster comprehension and L2 learning (Newton & Nation, 2021; Rodgers, 2016). As mentioned earlier, studies on audiovisual input and reading-while-listening programs with different age groups have proven that simultaneous exposure to aural and written input results in a synergy that facilitates decoding, speech segmentation, comprehension, and learning, especially in the case of non-transparent languages like English (Charles & Trenkic, 2015; Kormos et al., 2019; Montero Perez et al., 2013; Teng, 2019a, 2019b; Toscano-Fuentes, & Julián de Vega, 2018). In other words, the use of bimodal verbal input does not seem to burden L2 learners' working memory capacity. Instead, as Mayer et al. (2020) acknowledge, these two principles are reversed in the case of L2 learners given that print stays longer and may be revisited. Consequently, these exceptions gave rise to the subtitle principle (Mayer et al., 2020), which encourages the use of either written text or bimodal verbal input over the exposure to spoken-only condition.

With respect to the spatial- and temporal-contiguity principles, they contend that words and imagery should be presented near each other and simultaneously to enhance learning (Mayer, 2022). Therefore, it should not be assumed that imagery aids comprehension and learning per se since the quality of its support depends on the extent to which it clearly illustrates the meaning conveyed by the verbal input and how they are integrated (Black, 2020; Pellicer-Sánchez, 2022; Rodgers, 2020). The study carried out by Suarez et al. (2021) with adult learners of English demonstrated that learners' outcomes as regards L2 vocabulary learning through captioned-video viewing may change as a function of TV genre and their characteristics (e.g. number of shots and close-ups, captions appearance and pace). Their findings also indicated that the use of highly supportive imagery may even counteract the effects of L2 proficiency on L2 vocabulary learning through captioned-video viewing.

Thus far, the literature reviewed in this section suggests that the use of multimodal input does not necessarily lead to success. There are different factors to consider to match the requirements proposed by the Dual Coding Theory and the Cognitive Theory of Multimedia learning. Therefore, the following sections do not only move on to identify the actual effects of audiovisual input on L2 learning but also the factors that mediate learning: input-, treatment- and learner-related factors.

1.2 L2 learning through audiovisual input

Overall, the increasing literature on audiovisual input has provided robust evidence of its positive effects on L2 learning (Montero Perez, 2022). Nevertheless, as mentioned earlier, it is also important to bear in mind that the extent to which learners watch TV in a foreign language since an early age depends on contextual factors and national policies (Black, 2022; De Wilde et al., 2019; Lindgren & Muñoz, 2013; Muñoz et al., 2018). In the case of input-limited contexts, where only dubbing and L1-original-version broadcasting are the norm, learners' exposure to TV series and movies in a foreign language is restricted to their access to video sharing websites (e.g. YouTube), pay-per-view TV channels, and streaming platforms such as Netflix or Amazon Prime. Yet, in many cases, video settings may be manipulated by the viewers who may choose their preferred language for audio and/or subtitles. In other words, viewers still have the option to stick to their L1. This factor may explain primary school learners' little experience with OV (original version) TV series and the use of on-screen text (Marzá & Torralba, 2015). Equally important, the evidence suggests that children and families are not fully aware of the potential benefits of audiovisual input and on-screen text on L2 learning (Black, 2022). Therefore, further research and pedagogical interventions may be required to increase the use of audiovisual input as a language learning tool and help both, L2 instructors and learner-viewers make informed decisions.

As far as the language of on-screen text is concerned, L1 subtitles have been found to be more suitable for lower proficiency learners (e.g. Muñoz, 2017a; Pujadas & Muñoz, 2020). Specifically, their use has mainly been associated to video comprehension (e.g. Birulés-Muntané & Soto-Faraco, 2016; Casulleras & Miralpeix, 2017), word meaning learning (e.g. Peters et al., 2016), and leisure viewing given that L1 subtitles may be less cognitively demanding for L2 learner-viewers (Mariotti, 2015; Vanderplank, 2015). Notwithstanding, the empirical evidence also suggests that, overall, the use of L1 subtitles may lead to either lower (e.g. Mitterer & McQueen, 2009; Peters et al., 2016) or similar (e.g. Fievez et al., 2020; Pujadas & Muñoz, 2019) L2 gains in comparison with L2 captions in language aspects other than comprehension (Montero Perez, 2022). Still, it is important to acknowledge that much of the literature on the effects of captions has devoted its attention to vocabulary learning (see Montero Perez, 2022) rather than other language aspects such as grammar and pronunciation.

Considering that this dissertation examines the role of captions and not L1 subtitles on L2 learning, the following sections primarily focus on the findings from studies that have tested the effects of captioned-video viewing. Yet, due to the lack of investigations with primary school learners, the sections that explicitly refer to this age group will also review studies on reading-while-listening, L1 subtitles, as well as some investigations conducted in L1 contexts. To date, researchers have been reluctant to use L2 captions with primary school learners due to their still-developing reading skills in L1 and L2 (Vanderplank, 2015) and their low proficiency level. In fact, Vanderplank (2015) suggests that primary school learners are unable to cope with the speed of L1 subtitles until the age of 10. The scant research with foreign language primary school learners has shown that this threshold might also be applicable to L2 captions, as long as the materials match their characteristics (e.g. Alexiou & Yfouli, 2019; Montero Perez et al., 2013; Tragant & Pellicer-Sánchez, 2019).

1.2.1 Effects of captioned-video viewing on L2 learning

As stated earlier, the growing body of research on captioned-video viewing has inclined to support the use of on-screen text to promote L2 learning. Captions have been found to scaffold the viewing process by making aural input comprehensible. Yet, comparisons among studies should be done with caution because of the methodological differences adopted by each investigation (Montero Perez, 2022; Montero Perez et al., 2013).

1.2.1.1 Viewing comprehension and listening skills development

The literature suggests that the burden of listening comprehension tasks may be quite high for L2 learners since, most of the time, word meanings and ideas must be extracted at a speed that listeners cannot control for (Kormos et al., 2019; Newton & Nation, 2021). In addition, there are multiple factors that influence comprehension, such as vocabulary knowledge, the availability of contextual support (e.g. images, gestures, intonation, written-text support) and background knowledge on the main topics addressed in the input (Newton & Nation, 2021; Rodgers & Webb, 2011). The existing body of research on captioned-video viewing has shown evidence of the positive effects of onscreen text on viewing comprehension, particularly when assessed by means of receptive tasks, such as multiple-choice tests (Montero Perez et al., 2013). By comparing the captions vs. no-captions conditions, the seminal study conducted by Price (1983) with adult-ESL viewers showed initial evidence of the significant effects of captions on viewing comprehension. Likewise, Baltova's (1999) doctoral dissertation with secondary school learners of French confirmed the effectiveness of captions to enhance content comprehension and promote learning. The facilitating effects of captions are associated to speech segmentation given that the identification of word boundaries in the stream of speech may be a struggle for L2 listeners (Charles & Trenkic, 2015). Moreover, the use of on-screen text seems to be key for aural word recognition (Bird & Williams, 2002; Birulés-Muntané and Soto-Faraco, 2016), especially when it comes to the learning of a language with opaque orthography (Toscano-Fuentes & Julian de Vega, 2018), and when learners' exposure is mainly restricted to written input. The use of captions has also been found to help learners tune in to unfamiliar accents (e.g. Mitterer & McQueen, 2009), which is an outcome that may be relevant for input-limited contexts, where learners are mainly exposed to the target language through formal instruction (e.g. teacher-talk) and the target language variety employed in the materials selected for the course.

Yet, even when most findings support the use of captions to enhance speech decoding and comprehension (Montero Perez et al., 2013; Winke et al., 2010), the extent to which the use of L2 on-screen text facilitates comprehension at beginner stages may be uncertain. Viewers' capacity to use captions efficiently appears to be mediated by L2 proficiency level (e.g. L2 vocabulary knowledge) and input characteristics (e.g. vocabulary coverage) to prevent learners' cognitive overload (Mayer, 2014; Pujadas & Muñoz, 2020; Rodgers & Webb, 2017; Teng, 2019b, 2021; Webb, 2021). The first investigations with lower proficiency learners to examine the use of captions showed conflicting results (e.g. Markham, 1989 vs. Taylor, 2005). The investigation carried out by Markham (1989) with ESL university students of different proficiency levels (elementary, intermediate and advanced) revealed that the participants exposed to captions obtained significantly higher comprehension scores regardless of their proficiency level. Still, in line with more recent investigations (Pujadas and Muñoz, 2020; Rodgers & Webb, 2017), their comprehension scores were also influenced by the characteristics/complexity of each video. In contrast to Markham (1989), the study conducted by Taylor (2005) with university students did not show significant differences in comprehension between the captions vs. no-captions conditions, albeit viewers' positive attitude towards the use of text support. The lower proficiency learners (fewer years of formal instruction) seemed to struggle to comprehend the videos, while the use of captions increased the cognitive demands resulting in negative effects on comprehension, especially at the free recall task (in L1 English).

It is important to point out that Markham's (1989) and Taylor's (1989) investigations differed in a number of respects. To start with, Markham's (1989) study was conducted in an ESL context (naturalistic setting), therefore, the participants were constantly exposed to the target language. By contrast, in Taylor's (2005) investigation, the participants had been formally instructed for a short period of time, between 8 months and 5 years. Hence, it is not surprising that the low-proficiency group was unable to process the audiovisual materials with greater ease. This is why Taylor (2005) concluded that some minimum amount of L2 instruction, including listening, reading and viewing practice, may be required to process multimodal input more effectively. With respect to the testing instruments, Markham (1989) used two multiple-choice tests in the L2, whereas Taylor (2005) used two instruments in learners' L1: a free recall task and a multiple-choice test. It seems that the use of measures that assess receptive comprehension has a lower level of difficulty (Montero-Perez et al., 2013), especially when the instrument is in learners' L1.

The longitudinal study carried out by Pujadas and Muñoz (2020) with eighth graders from Spain (13-14 years old) compared learners' performance on different test items that assess receptive comprehension (in learners' L1): true-and-false and multiple-choice formats to measure either textually explicit and inferential comprehension. The participants that watched the episodes with captions scored higher at the true-and-false items, especially in the ones that measured inferential comprehension. As Pujadas and Muñoz (2020) explained, learners may not have fully comprehended the textually-explicit information conveyed in the captions, so they used top-down processes to take advantage of the multimodal input and fill knowledge gaps so as to figure out the main ideas from the episodes. This falls in line with the literature that posits that L2 learners may rely on multiple factors, such as their L1 skills and their previous knowledge on the topics addressed in the input to compensate for their low proficiency level and enhance comprehension (Cho & Krashen, 1994; Hwang & Nation, 1989; Krashen, 1996, 2004; Schmitt & Carter, 2000; Yamashita, 2002).

The evidence presented thus far appears to highlight the critical roles of L2 proficiency and materials selection (Gass et al., 2019). Beginner learners may not be able to process captioned videos with ease if their complexity does not match their proficiency level

(Montero Perez et al., 2013; Muñoz, 2017a; Winke et al., 2010). The study conducted by Teng (2019b) with primary school learners of English in Hong Kong (year 6) adapted two age-appropriate videos to compare three captioning conditions: full captions, keyword captions and no captions. This study also examined the effects of repeated viewing (twice) on comprehension. The results confirmed the effectiveness of full captions over keyword captions and no-captions in global and detailed comprehension, which were measured by means of a written recall protocol and a multiple-choice test in the L2, respectively. These outcomes fall in line with the literature that has demonstrated that the simultaneous exposure to aural and written input facilitates decoding and allows learners to pay more attention to the content and images (Pellicer-Sánchez, 2022). Overall, full L2 captions and repeated viewing resulted in higher comprehension (detailed and global comprehension), which is a finding that has also been obtained with adult learners (Majuddin et al., 2021; Winke et al., 2010). Nevertheless, when the video was watched only once, the participants showed a better performance in global rather than detailed comprehension, which is congruent with Pujadas and Muñoz's (2020) findings. Similarly, the lower-proficiency learners obtained comparable gains in keyword and full captions conditions, however, the latter required a second viewing to score higher at the detailed comprehension questions. As Teng (2019b; 2022) concluded, learners' greater effort to cope with the speed of captions prevented them from paying attention to specific information (i.e. cognitive overload). In other words, learners first watched for global comprehension, while a second viewing allowed them to focus on details. This is partially consistent with the study conducted by Linebarger (2001) with very young L1 learners (second graders) in the US, which found that the absence of L1 captions allowed the participants to attend to non-essential elements of the story.

The evidence presented in this section suggests that captions support listening comprehension as long as the input matches learners' proficiency level. In addition, at lower proficiency levels, repeated viewing seems to be a good strategy to boost comprehension (Teng, 2019b, 2022). Nonetheless, most of the studies have tested comprehension of the same videos used in the intervention (e.g. Pujadas & Muñoz, 2020; Rodgers, 2013; Rodgers & Webb, 2017; Teng, 2019b) and there is little evidence on the extent to which extensive viewing may support the development of L2 listening skills (i.e. generalization of learning), as in the case of reading-while-listening, where the use of bimodal verbal input has been shown to lead to significant improvement over time (e.g. Chang, 2011). The sub-study conducted by Lindgren and Muñoz (2013) as part of the ELLiE project (Early Language Learning in Europe) on out-of-school exposure showed evidence of the effects of sustained viewing on the development of receptive language skills (listening and reading). They collected data from primary school learners (4th graders, 10-11 years old) across seven European countries (Croatia, England, Italy, the Netherlands, Poland, Spain and Sweden) by means of a questionnaire, a listening and a reading task. The results indicated that viewing films in the foreign language was a strong predictor of learners' performance at the listening and reading tasks.

The few existing experimental studies on listening skills development have examined the effects of captioned-video viewing on speech perception, namely bottom-up processing, after a short intervention (1 or 2 videos). The 25-minute video used by Mitterer and McQueen (2009) with L1-Dutch learners of English was reproduced with a strong regional accent of English (Australian or Scottish) that was unfamiliar to the participants. The researchers measured learners' capacity to reproduce (orally) a set of fragments from the same episode, as well as novel phrases from the same TV series/movie. To this aim, the participants were assigned to one of the three subtitling conditions: L2 captions, L1 subtitles and no subtitles. The results showed that the participants that watched the video with captions showed superior performance when reproducing both, known and unknown extracts. Therefore, the use of captions allowed the participants to retune to the unfamiliar accent and improve in aural word recognition.

By drawing on Mitterer and McQueen's (2009) study, Charles and Trenkic (2015) conducted an experiment with university learners of English (international students) in the UK. The participants had to watch two documentaries (30 minutes each) in a period of two weeks under one of the three input conditions: L2 audio with captions, L2 audio without captions, and only text. By following a pretest-posttest design, the researchers administered a shadowing task (oral repetition) that consisted of utterances that were encountered in the episodes, as well as fragments that were not part of the treatment and included other speakers. In line with Mitterer and McQueen's (2009) findings, the L2 captions condition showed significantly higher improvement at posttest. The students were not only able to segment and reproduce the utterances that had been encountered in the videos, but also the ones that had not.

Charles and Trenkic's (2015) results were partially replicated in the study by Birulés-Muntané and Soto-Faraco (2016) with intermediate learners of English from Spain (university level). They employed a listening task (fill-in-the-gaps), whose aural stimuli were new extracts from the same TV series used for the study. The comparisons between the three viewing conditions (L2 captions, no captions, L1 subtitles) indicated that the L2 captions group showed greater improvement in auditory perception after watching a single one-hour episode. It is important to note that spelling accuracy was not part of the assessment. All in all, the generalization of learning found in the three studies suggests that captioned-video viewing fosters the development of L2 listening skills as a result of a short intervention. Yet, given that these studies were carried out with adult learners, it is uncertain whether younger learners would make significant progress in such little time. The literature has proven that in foreign language contexts, primary school students are less efficient learners than teenagers and adults (Holmes & Myles, 2019; Muñoz, 2008; Van Lommel et al., 2006).

1.2.1.2 Reading skills development

L2 reading is a highly complex task that integrates lower- and higher-level reading processes (e.g. word recognition and general interpretation, respectively), which rely on multiple factors to ensure adequate levels of comprehension, such as learners' knowledge of the L2 (e.g. vocabulary, syntax, orthography, phonology and morphology), background knowledge on the main contents addressed in the input, general comprehension abilities, presence of contextual support (e.g. images and contextual cues) and L1 reading skills (Grabe & Jiang, 2018; Koda, 2007; Nassaji, 2014; Perfetti et al., 2007; Sparks, 2021). Yet, even when the literature has consistently shown evidence of the influence of L1 reading skills on L2 reading (Birch & Fulop, 2021; Koda, 2007; Llanes, 2018; Nassaji, 2014; Perfetti et al., 2007; Tragant et al., 2019), research suggests that L2 factors might be stronger predictors of L2 reading (Alderson et al., 2016; Jeon & Yamashita, 2014; Proctor et al., 2005; Sparks, 2021; Verhoeven and van Leeuwe, 2012; Yamashita, 2002). Overall, it may be safer to convey that L2 reading cannot be fully detached as a language or reading problem, given that among the multiple variables that play a role in the process, L1-L2 distance and the differences between languages as regards their writing system may have either a positive or negative influence on L2 reading skills development, especially at earlier stages (Birch & Fulop, 2021).

As a result, the complexity of the L2 reading process might potentially explain why L2 reading has not been found to be a highly popular activity outside the L2 classroom (De Wilde et al., 2019; Muñoz, 2020b; Peters, 2018; Riveros, forthcoming), which is a key limitation when considering that the reading practice is crucial to show significant improvement over time (Birch & Fulop, 2021; Grabe, 2009; Grabe & Stoller, 2020). Reading is largely associated with the reading of books or printed texts but, although this is not an activity that should be replaced, there are other actions that may well increase the amount of practice and contribute to the development of L2 reading skills (Riveros, forthcoming). The investigation by Lindgren and Muñoz (2013) reported above (section 1.2.1.3) lends support to the use of audiovisual input to enhance L2 reading skills development in foreign language contexts. Unfortunately, most of the studies with primary school students that have explicitly focused on reading skills development through captioned-video viewing have been conducted in L1 contexts. This is problematic since some of the key differences between L1 and L2 readers have to do with language proficiency and the amount of exposure to print, which are crucial to become familiar with L2 orthographic patterns and the automatization of low-level reading skills (e.g. word recognition, syntactic parsing, and meaning encoding) (Grabe, 2009; Grabe & Jiang, 2018). Therefore, further research is strongly required to determine whether the findings emerging from L1 contexts may be fully translated to the learning of a foreign language.

Up to now, L1 studies have promoted the use of bimodal verbal input as a tool that supports reading skills development (Kothari et al., 2002). However, the processing patterns and the specific aspects that may benefit from this activity seem to depend on the extent to which lower-level reading skills are automatized. By comparing four conditions that examined the use of L1 captions and oral narration, Linebarger's (2001) study with second graders (7-9 years old) in the US demonstrated that the use of onscreen text led to greater gains in terms of word recognition. Even when the use of print allowed the participants to identify the key information from the clips, their greater cognitive effort to process the written input affected their capacity to focus on details or less relevant elements from the story. The fact that learners remembered more information in the listening-only condition is consistent with the modality principle (Mayer, 2014), which has been found to be applicable in L1 contexts (Mayer et al., 2020). As also suggested by Sadoski and Paivio (2013), when the text decoding process is effortful, learners devote greater attention to lower linguistic levels, leaving less cognitive resources available to process other elements, such as images and gestures, and make referential connections.

In a new study, Linebarger et al. (2010) examined reading skills development in second and third graders in the US (native and second language learners), who were considered 'at risk' of experiencing reading failure due to their limited proficiency level. After watching six episodes of different animated cartoons (L1 captions vs. no captions conditions), the viewers exposed to on-screen text scored significantly higher in terms of word recognition. A key finding of this study was that L1 captions supported target word comprehension and inferential comprehension, whereas literal comprehension was unaffected by the use of print. Therefore, unlike Linebarger's (2001) outcomes, learners' cognitive capacity did not seem to get overloaded by the use of text support. In addition, the L1-captions group improved as regards non-word reading (English patterns) but not oral reading fluency since six episodes may not have been enough to make significant progress. Differences between investigations may be associated to learners' stage of reading skills development (Minucci & Cárnio, 2010) (see Table 2) and the characteristics of the materials (e.g. language complexity and image support). As Linebarger et al. (2010) hypothesized, there must be a stage where captions are neither too challenging nor too easy to follow, so they may successfully aid comprehension and foster reading skills development.

The features of the audiovisual materials implemented and the extent to which they match learners' characteristics (e.g. age and L1 reading skills development) (see Table 2) seem to predict the effects of onscreen text on reading skills development. In a study with fourth and fifth graders in India, Kothari et al. (2002) examined the effects of L1 captions on L1 reading skills development from watching song video clips. Specifically, they measured syllable and word reading ability to assess the effects of a 3-month experiment (less than 18 hours in total). By comparing the experimental conditions (L1 captions vs. no captions vs. control group), the results revealed that the captions condition resulted in significantly higher gains, especially in the case of monosyllable words.

Table 2.

Ehri's (2005) phases of word reading development

Phases	Description
Pre-alphabetic phase	Non-readers. Learners are unable to match phonemes and graphemes to read words. They may read words by remembering visual features.
Partial alphabetic phase	They know the names or sounds of alphabet letters and use these to remember how to read words. They are only able to link some letters and sounds in words. They tend to confuse words that have similar spelling patterns. They lack knowledge of the alphabetic system, especially vowels. They struggle when decoding unfamiliar words. They tend to invent parts of word spelling. Inaccurate segmentation of words into phonemes.
Full alphabetic phase	They can learn sight words by forming complete connections between letters in spellings and phonemes in pronunciations. They know the main grapheme–phoneme correspondences and they can segment pronunciations into phonemes that match up to the graphemes they see. Learners do not struggle with phonological segmentation anymore.
Consolidated alphabetic phase	Readers retain increasingly more sight words in memory. As they become familiar with letter patterns that recur in different words, the grapheme–phoneme connections in these words become consolidated into larger units. These include spellings of rimes, syllables, morphemes, and whole words that have become unitized. The letters that form each word are not processed separately anymore.

Note. (Ehri, 2005, pp. 173-176)

Although it is true that the ultimate aim of reading instruction is to attain high levels of comprehension, the literature has shown that the instruction and development of lower-level reading skills are crucial to fulfill this objective (Ijalba & Obler, 2015). The interaction between higher- and lower- level reading processes is key to build coherent mental representations (Alderson et al., 2015; Grabe & Stoller, 2013; Nassaji, 2014). Therefore, even when research suggests that, at early reading stages, captions support the development of lower-level reading skills at the expense of viewing comprehension (Linebarger, 2001), the use of on-screen text might still be seen as a contribution to the learning process. In other words, despite its limitations, the use of captions would equally foster the automatization of orthographic and phonological processing, which may eventually result in higher levels of comprehension and motivation to read (Toscano-Fuentes & Julián de Vega, 2018). That being the case, captions may have the potential to break the vicious circle of low-achievers' reluctance to read (Birch & Fulop, 2021) and, to a certain extent, counteract learners' lack of exposure to L2 print (De Wilde et al., 2019; Muñoz, 2020b; Peters, 2018). Previous research has shown evidence of teachers and young learners' positive attitude towards the use of captions and L1 subtitles in both, naturalistic L1/second language settings and foreign language contexts (Black, 2022; Koskinen et al., 1985; Marzá & Torralba, 2015; Zabalbeascoa et al., 2015). These high levels of motivation have also been echoed in studies that have implemented reading programs with audio support (Tragant & Vallbona, 2018; Tragant et al., 2019).

Yet, it is important to bear in mind that, in contrast to early research on the area, the recent literature on the potential effects of subtitles on reading skills development (e.g. Black, 2021) does not intend to antagonize or favor the use of subtitles (either in L1 or L2) over the reading of static texts. For instance, the pioneer longitudinal study conducted by Koolstra et al. (1997) with primary school learners (2^{nd}-4^{th} graders) demonstrated that the regular use of interlingual subtitles (L2 audio and L1 subtitles) at home led to significant gains as regards L1 decoding skills but not reading comprehension, which seems to be partially consistent with the study carried out by Linebarger (2001) a few years later. However, this study also indicated that a greater engagement with subtitled viewing was associated to the *inhibition* of the development of L1 reading comprehension, as well as "… a *television-induced reduction* in leisure-time book reading and a *television-induced depreciation* of reading…" (p.147). At present, research aims at exploring the learning potential and limits of both activities by acknowledging their differences and similarities, which would make them complementary rather than incompatible (Webb, 2015). In fact, the outcomes of Koolstra et al.'s (1997) study may have reflected learners' avoidance of reading as a result of their poor skills instead of the negative effects of viewing on reading.

As for the reading of dynamic texts, research has shown that this activity may be particularly challenging in either language (L1 or L2) (e.g. Hefe, 2013), especially for young learners due to their strong reliance on reading speed and language proficiency (Muñoz, 2017a; Newton and Nation, 2021). Thus, investigations should not only study the reading aspects that profit from text support, but also identify the stage at which L2 learners may be prepared to cope with captions to significantly benefit from them and enjoy the experience. Some investigations on the use of L1 and L2 subtitles have shed some light on young learners' processing of onscreen text. The study conducted by Minucci and Cárnio (2010) found that for most second graders, the reading of L1 subtitles (no audio) was a struggle that required greater levels of visual attention since, at that age, most children are still developing their decoding skills (Ehri, 2005; see Table 2). Through an interview, Black (2022) obtained similar results when exploring 8-9-year-olds' experience with interlingual subtitles (L2 audio, L1 subtitles). Despite their positive attitude towards the treatment, several participants found that the subtitles were too fast. As for the processing of L2 captions, the eye-tracking study conducted by Tragant and Pellicer-Sánchez (2019) with fifth graders (10-11 years old) revealed that the audiovisual material employed for the purpose of their study was suitable to encourage L2 learners to read. Learners' capacity to process L2 captions with greater ease may have been the result of their higher stage as regards reading skills development, as well as the characteristics of the video, whose script was very easy to read (Flesch Reading Ease score=97,9) (see Table 2). Teng's (2019b) study with sixth graders also confirmed that once certain level of automaticity in lower-level reading skills is attained, children are capable of using captions to boost comprehension.

Hence, in view of all that has been mentioned so far, it seems that captions might have the potential to support the development of L2 reading skills. However, that being the case, learners seem to be prepared to cope with the speed of captions at around the age of 9 or 10. At a younger age, children are still developing their L1 decoding skills, which explains why the reading of L1 subtitles is found to be an effortful process by then (Black, 2022; Vanderplank, 2016). Equally important, researchers and L2 instructors should take into consideration the characteristics of the audiovisual materials to warrant they match learners' L2 proficiency level (Alderson et al., 2016; Grabe, 2009; Sparks, 2021).

1.2.1.3 Vocabulary learning

As previously stated, most of the literature on captioned-video viewing has paid particular attention to vocabulary learning. Nevertheless, comparisons among studies should be done with due caution given that they differ in terms of learner and word characteristics, treatment conditions, and most importantly, the specific knowledge dimensions under study (see Table 3), as well as the instruments administered to measure gains (Montero Perez, 2022; Montero Perez et al., 2013). As discussed earlier, the simultaneous processing of audio and onscreen text compensates for learners' L2 knowledge gaps to facilitate text decoding and improve the processing of imagery (Pellicer-Sánchez, 2022; Pellicer-Sánchez et al., 2020). For instance, in a study with Dutch-Belgian secondary school students (16-17 years old), d'Ydewalle and Pavakanun's (1997) comparison of 81 experimental conditions (audio: L1, foreign language, no audio; subtitles: L1, foreign language, no subtitles; 9 target languages) demonstrated that the simultaneous processing of audio and subtitles in a foreign language (captions) leads to higher gains (e.g. vocabulary) than each mode in isolation (in a foreign language). Therefore, learners' effective use of L2 audio and captions may not only trigger the noticing of new target word forms, but also the filling of unknown word knowledge dimensions (see Table 3). Some studies on reading-while-listening have also lent support to the use of bimodal verbal input by either demonstrating that it enhances vocabulary learning (Brown et al., 2008; Webb & Chang, 2012, 2015) or showing that audio and text lead to comparable gains in contrast with the reading-only condition, where learners may process the text at their own pace (Tragant et al., 2019).

Overall, the literature has shown that captioned-video viewing leads to significant vocabulary gains in terms of both, receptive and productive vocabulary knowledge (Montero Perez et al., 2013). Still, the use of audiovisual materials leads to higher gains in receptive vocabulary knowledge, which is an outcome that may be associated to the lower cognitive demands required to accomplish this task (González-Fernández & Schmitt, 2020; Montero Perez, 2022). Equally important, meaning-learning seems to be more cognitively demanding than word-form learning given that the former draws on learners' capacity to integrate the meaning cues provided by each modality while viewing (Gesa, 2019; Mayer, 2014, 2022; Montero Perez et al., 2014; Peters et al., 2016; Pujadas & Muñoz, 2019; Suárez & Gesa, 2019). Although learners' vocabulary gains may be found to be significant, the average number of items acquired in incidental conditions has been found to be relatively low (Montero Perez, 2022; Webb, 2020), this is why some studies have examined the synergy between video-viewing and vocabulary pre-teaching or tasks to maximize learning (e.g. Pujadas & Muñoz, 2019; Suarez & Gesa, 2019; Teng, 2022). Added to that, previous vocabulary knowledge seems to be one of the strongest predictors of vocabulary learning from audiovisual input. The so-called 'the rich get richer principle' or 'The Matthew effect' indicate that the greater vocabulary knowledge, the greater vocabulary gains (Montero Perez, 2022; Montero Perez et al., 2013; Stanovich, 1986), which is one of the variables that will be further explained in section 1.3.2.

Table 3.

What is involved in knowing a word.

Table 2.1 What is involved in knowing a word

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express the meaning?
Meaning	Form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	Constraints on use (register, frequency, . . .)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Note: In column 3, R = receptive knowledge, P = productive knowledge.

Note. (Nation, 2020, p.16).

As regards primary school students, the investigations on vocabulary learning through audiovisual input have mainly explored the effects of L1 subtitles. For instance, the investigation conducted by d'Ydewalle and Van de Poel (1999) with L1-Dutch Belgian primary school participants between third and sixth grade (8-12 years old) studied the effects of subtitled-viewing on word-meaning recognition, among other language aspects. To this aim, the participants had to watch a 10-minute video either in L2 Danish or L2 French by using reversed or L1 subtitles. The results revealed significantly higher vocabulary gains in L2 Danish in comparison with L2 French due to the greater language distance between Dutch and French. In other words, the similarities between two languages (e.g. in terms of cognateness) facilitated learning (De Wilde et al., 2022; Muñoz et al., 2018). In addition, the outcomes indicated that the participants scored higher when the L2 was presented in the soundtrack than in the form of onscreen text (reversed subtitles). Therefore, these results showed evidence of the fact that young learners make a greater effort when reading in another language without audio support and need plenty of practice to obtain greater benefits from L2 onscreen text.

The benefits of interlingual subtitles (L2 soundtrack and L1 subtitles) were also confirmed by Koolstra and Beentjes (1999) in their seminal study with fourth and sixth graders from The Netherlands. Specifically, they explored L2 vocabulary learning (aural-word form and meaning recognition) from viewing a 15-minute documentary under two experimental conditions: L1 subtitles vs. no subtitles. The results revealed that the use of on-screen text led to higher learning gains than the listening-only condition. In addition, the two experimental groups outperformed the control group that watched the video in L1 Dutch. In this study, sixth graders outperformed fourth graders regardless of the experimental condition. Still, viewing at home had a greater influence on the outcomes in both year levels. Similarly, Lekkai's (2014) investigation with primary school learners (9-12 years old; 4th-6th grade) from Greece also yielded fourth graders' significantly lower scores in word-meaning recognition from watching a single video in Italian, which was a language they had never been instructed to learn. Specifically, the participants were expected to gain vocabulary knowledge under one out of three treatment conditions: L2 Italian audio-L1 Greek subtitles, L2 Italian audio-no subtitles, and L1 Greek audio-L1 Greek captions (control group). The results also revealed significantly higher gains when L1 subtitles were available.

In a recent study with young EFL learners (year 6) from Spain, Gesa (2019) also explored the effects of interlingual subtitles on vocabulary learning. Nevertheless, this investigation differed from previous studies in that the researcher implemented an extensive treatment to finally measure word-form and meaning recall. Specifically, the participants had to watch 24 episodes of two different TV series over three school terms (once a week), and the viewing sessions were accompanied by pre-viewing and after-viewing vocabulary activities to promote intentional vocabulary learning. Comparisons between the experimental and the control group (only completed the vocabulary learning activities) did not reach statistical significance until term 3. In other words, gains in the first two terms were associated to the vocabulary activities rather than the subtitled-viewing experience. In addition, the results showed significant but small immediate vocabulary gains, while the delayed-posttest indicated that there was little vocabulary retention. Overall, scores on word-form recall were found to be higher than meaning recall despite the availability of direct L1 translations. The limited gains may be attributed to the testing demands (i.e. the vocabulary dimensions elicited) and learners' low proficiency level, as well as the fact that the learning process may be less efficient when using interlingual subtitles in the case of young learners given that they have to split their attention between the two sources of verbal input which are available in different languages. In addition, as Gesa (2019) pointed out, the significantly higher gains obtained in the third term might also be attributed to the fact that the TV series used at this stage of the investigation displayed more supportive imagery. Therefore, this assumption further supports the importance of input characteristics (Suarez et al., 2021) and imagery in L2 learning (Rodgers, 2020).

With respect to the use of L2 captions, very few studies have explored their effects with primary school students, and most of them have been conducted in late primary school years (fifth and sixth grade). To start with, Galimberti and Miralpeix (2018) assessed vocabulary learning (meaning recognition, and word-form and meaning recall) in L1 Italian learners of English (sixth grade). The participants had to watch a 22-minute episode under one of the experimental conditions: L2 captions, L1 subtitles or no subtitles. The results yielded small but significant gains regardless of the treatment condition. The overall recall scores (word form and meaning recall) indicated that the captions group scored higher than the participants that were exposed to L1 on-screen text or the listening-only condition but their difference only reached significance in the case of L1 subtitles. With respect to meaning recognition, the use of L1 subtitles led to higher gains than the use of captions and the listening-only condition. Yet, the between-groups differences were not found to be statistically significant. Thus, as suggested by Gesa (2019), the use of L1 subtitles may not contribute significantly to meaning learning in this age group.

Teng (2019a) also studied the effects of captioned-video viewing with sixth graders from China. His investigation examined the effects of different L2 captioning conditions (full captions, keyword captions, no captions) and word frequency (1/3 repetitions) on vocabulary learning (written-word form recognition, and meaning recognition and recall) after watching a 25-minute video. The analyses revealed that the full captions condition significantly outperformed the keyword captions and the no captions conditions. Moreover, within the full captions condition, the participants scored higher in terms of word-form recognition, followed by meaning recognition and meaning recall. With respect to word frequency, three repetitions led to significantly higher gains regardless of the captioning condition. All in all, this investigation shows evidence of the contribution of the synergy between audio and text to vocabulary learning (Montero Perez, 2022; Vanderplank, 2016). In a subsequent study, Teng (2022) compared receptive vocabulary learning gains (form, meaning and use) in sixth graders under four different captioning conditions: full captioning, keyword captioning, glossed keyword captioning and glossed full captioning. The results indicated that the differences between full and keyword captioning found in the previous study was kept but, in this study, the glossed full captioning condition led to significantly higher vocabulary gains. In other words, having access to word meaning was found to boost vocabulary learning in form, meaning and use, which is consistent with the results of studies that have assessed the use of glosses with university students (Fievez et al., 2021 Montero Perez et al., 2018). Similarly, the implementation of graphic organizers prior to video viewing also contributed to vocabulary learning significantly. However, the captioning conditions had a stronger effect on the outcomes in comparison with the use of graphic organizers.

Pujadas and Muñoz (2019) carried out one the few longitudinal studies with school students that have explored the effects of captioned-video viewing on vocabulary learning (24-week treatment). The four sub-groups of secondary school students from Spain (13-14 years old) watched one episode a week of the TV series *Fresh off the Boat*. Vocabulary gains were tested as regards written-word form and meaning recall. The groups were randomly assigned to one of the four treatment conditions: L2 captions+no vocabulary pre-teaching, L2 captions+vocabulary pre-teaching activities, L1 subtitles+no vocabulary pre-teaching, L1 subtitles+vocabulary pre-teaching activities. The results indicated that the participants exposed to L2 captions and vocabulary pre-teaching activities obtained significantly greater gains from the treatment. Although the use of vocabulary pre-teaching activities resulted in greater gains regardless of the subtitling condition (L1 or L2), the captions group was found to score higher than the L1 subtitles group. With respect to the students that were not subjected to vocabulary pre-teaching activities, the L1-subtitles group outperformed the captions group, especially in terms of meaning-recall, which was expected considering that L1 subtitles provide direct translations of the unknown words. Yet, the use of pre-teaching activities seemed to be more effective as regards the learning of written-word forms given that proficiency appeared to be a stronger predictor of meaning learning. Overall, this study confirms that low-proficiency school learners (eighth graders) do benefit from captioned-video viewing, especially if they are complemented with pre-viewing tasks. In addition, it demonstrates that meaning learning may be more cognitively demanding given that learners have to attend and connect verbal and non-verbal input efficiently to work out the meaning of unknown words. Still, it is important to point out that even though the participants from this study were early secondary school students, the outcomes of this investigation may not be directly extrapolated to primary school learners due to multiple factors, such as stage of cognitive development (Muñoz, 2008), reading skills (Muñoz, 2017a), greater exposure to L2 input (Muñoz, 2020b), and years of L2 instruction.

In a study with 8 and 9-year olds (L1 Greek-L2 English), Alexiou and Yfouli (2019) explored the effects of extensive captioned-video viewing on meaning recognition and oral-word form recall. To this aim, the participants had to watch ten episodes of the animated cartoon *Charlie and Lola*. Specifically, in each session, the participants had to watch an episode twice (at the beginning and at the end of the class). As expected, the gains were higher in meaning recognition than aural-word form recall. Interestingly, the younger participants scored significantly higher than the 9-year-olds. Yet, this investigation was conducted at language school and it lacked a pretest, therefore, previous knowledge and learners' background information did not seem to have been considered.

Taken together, the studies cited in this section suggest that video viewing with different types of subtitles leads to significant vocabulary learning gains. Yet, most of the studies have been conducted with late primary school learners (Teng, 2019a, 2022), while the one that did assess younger students' outcomes from captioned-video viewing lacked a pretest (Alexiou & Yfouli, 2019). In addition, the simplification of the captions (keyword captions) did not seem to equal the outcomes obtained in full captions condition (Teng, 2019a, 2022), suggesting that viewing might be less cognitively demanding when processing bimodal verbal input with full captions, leaving more cognitive resources available to notice and learn unknown vocabulary items (Kormos, 2017). Likewise, the use of glosses or previewing vocabulary activities seem to maximize vocabulary learning (Pujadas & Muñoz, 2019; Teng, 2022). All things considered, further research is strongly required to determine the extent to which captioned-video viewing is beneficial for younger foreign language learners.

1.2.1.4 Grammar learning

The scant evidence on the effects of audiovisual input on grammar learning has indicated that, in contrast to vocabulary studies, a single video may not be sufficient to foster the learning of L2 grammar constructions (Van Lommel et al., 2006; Y'dewalle & Van de Poel, 1999). Therefore, sustained exposure to audiovisual input (Kuppens, 2010; Muñoz et al., 2018) and the support of explicit instruction may be required to obtain significant gains over time (Van Lommel et al., 2006). A Usage-based perspective for language learning emphasizes that frequency of occurrence of target language constructions is key to promote incidental learning due to learners' sensitivity to regularities in the input at different language levels (Ellis, 2002; Rieder, 2003). Target language knowledge is gradually acquired in the form of linguistic constructions of different levels of abstraction and complexity that range from concrete words and formulaic chunks (e.g. thank you), to abstract classes (e.g. nouns and passive form) and mixed constructions (e.g. question formation) that barely detach lexis from grammar (Madlener, 2015; Muñoz et al., 2018, pp. 8-9). In addition, the learning process is influenced by input characteristics such as frequency distributions, saliency and task essentialness of the target linguistic constructions (Ellis, 2002; Ellis & Collins, 2009; Madlener, 2015; Socket & Kusyk, 2015).

As regards frequency types, it is necessary to make a distinction between token and frequency type (Ellis & Collins, 2009). The former is defined as the number of times a linguistic unit appears in the input (e.g. hello), whereas the latter refers to the number of times a category from a specific word-level or syntactic construction occurs (e.g. regular verbs in the verb slot). Although it is true that high token frequency promotes chunk entrenchment (strong mental representation), there is no consensus on the number of times a target item should be encountered in the input to promote intake (Uchihara et al., 2019). Concerning frequency type, the findings suggest that skewed distributions (centered on prototypical exemplars) may be more effective at enhancing pattern abstraction and productivity in comparison with high frequency types (Ellis & Collins, 2009; Madlener, 2015). However, this is not something that can be manipulated when using authentic materials such as TV series or movies. Some studies have created or adapted their input in order to control for different variables (E.g. Cintrón-Valentín et al., 2019; Lee & Révész, 2018; Madlener, 2015).

The few investigations conducted with young learners have been implemented with interlingual or reversed subtitles but not captions. The study by d'Ydewalle and Van de Poel (1999) described above also examined the learning of syntax and morphology (L2 French and L2 Danish) with primary school learners from Belgium (between the ages of 8 and 11). They found that watching a 10-minute video was insufficient to learn syntax and morphology in either language or subtitling conditions (i.e. very small gains). Likewise, Van Lommel et al. (2006) taught five Esperanto grammatical rules to a group of primary (11 years old) and another group of secondary school students (17 years old) from Belgium through a 25-minute cartoon (one episode) with either L1 or reversed subtitles. As in d'Ydewalle and Van de Poel's (1999) study, the gains obtained in incidental conditions were low, whereas the use of pre-teaching activities and the explicit instruction to learn grammar from the video were found to foster intake, especially in the case of secondary school students. Given the low gains obtained in these seminal studies with young learners, the more recent investigations with university students have tested the effects of longer treatments (e.g. Muñoz et al., 2021; Pattemore & Muñoz, 2020, 2022a) and input enhancement (e.g. Cintrón-Valentín et al., 2019; Cintrón-Valentín & García-Amaya, 2021; Lee & Révész, 2018, 2020), which consists of highlighted target language constructions that aim at getting viewers' attention (Schmidt, 2001).

Muñoz et al. (2021) compared the learning of target grammar construction under two experimental conditions (L2 captions vs. no-captions) with EFL learners (university students, A1-C2 CEFR levels) in Spain. The participants had to watch 10 episodes of the TV series *The Good Place* (one episode per session). The results indicated that even though both groups benefitted from the treatment, the use of L2 captions enhanced grammar learning. In addition, frequency effects were found to be moderated by the processing of onscreen text, suggesting that captions make the target grammar constructions more salient in the input. Furthermore, the treatment seemed to be particularly more beneficial for intermediate students since, as suggested by the authors, the ones that had an advanced level of English had less room for learning (higher prior knowledge). Pattemore and Muñoz (2020) also contributed to the body of research on grammar learning by indicating that the results were also influenced by working memory capacity but not language learning aptitude. However, the effects of working memory were specifically found in the group that was not exposed to captions, indicating that the use of onscreen text facilitated input processing and moderated the influence of this cognitive factor. Therefore, in line with the literature on vocabulary learning, captions appear to foster the learning of grammar constructions.

A later investigation carried out by Pattemore and Muñoz (2022a) expanded on these findings by examining the effects of three different experimental conditions at pretest, posttest and delayed posttest: L2 Captions, no-captions and enhanced captions. Also, they compared the learning of three types of grammar constructions: fully-filled (i.e. a chunk), partially-filled (e.g. with a slot as in look after [somebody]) and fully-schematic (e.g. Noun modifier+Noun) (Fried, 2015). The results revealed that the participants improved significantly over time regardless of the captioning condition. However, at delayed posttest, the results indicated that the participants that watched the videos with unenhanced captions scored higher than the ones that did it under enhanced captions or no-captions conditions, whose scores were found to be comparable. As Pattemore and Muñoz (2022a) hypothesized, the use of captions may have been enough to get viewers' attention on the target constructions. In addition, the learning of fully-filled constructions led to lower gains over time, which was attributed to other factors such as frequency and saliency, or the fact that their use is less flexible than in the case of partially-filled and fully-schematic constructions, resulting in higher learning burden. In comparison with their previous investigation (Muñoz et al., 2021), frequency of occurrence may have been overpowered by other factors given that it was not found to predict learning in any of the experimental conditions (captions or no-captions).

The participants in Lee and Révész (2018) (EFL university students from Korea, B1-B2 CEFR level) were subjected to three treatment sessions where they had to complete nine multimodal input-based tasks. They consisted of sentences that were simultaneously processed by the participants in audio and print to decide which static picture displayed on the screen (1/3) accurately represented each sentence. The participants were assigned to one of the two captioning conditions: regular L2 captions or textually enhanced captions. They used an eye tracker in order to examine how learners processed the input while completing the tasks and measure the amount of attention the participants devoted to the target language construction (third-person pronominal anaphoric reference). Learning was measured through a written and oral grammaticality judgement test before and after the treatment. The eye-tracking data indicated that textual enhancement directed learners' attention to the referential antecedent but not necessarily on the personal pronouns, given that in the latter, the behavior of the enhanced-captions group did not seem to differ from the unenhanced captions group. In terms of grammar learning over time (accuracy), both groups appeared to benefit from the treatment but the results revealed that textual enhancement yielded higher gains.

Then, in a follow-up study, Lee and Révész (2020) made multiple changes to their research design. First of all, the students had to watch 24 news clips in a single session (2-hour long). At the end of each video, the participants had to decide whether the title and the category assigned to the clips were appropriate or not. This time, apart from the enhanced and unenhanced captions conditions, they included a control group that watched uncaptioned videos in order to measure the effects of onscreen text on the learning of two target grammatical constructions (present perfect and past simple) that are frequently encountered in news reports. In addition, apart from using an eye-tracker to examine attention allocation, they administered three instruments at three testing times (pretest, posttest and delayed posttest): an oral production test, a written production test, and a fill-in-the-blank test. Apart from confirming the results obtained in Lee and Révész (2018), the analyses showed that the use of captions led to significantly higher gains in comparison with the uncaptioned condition regarding the use of present perfect at the written production test. Their performance was also superior at the oral production and the fill-in-the blank tasks but the between-groups differences did not reach statistical significance. As for the captioning groups, the enhanced captions condition was found to score higher than the unenhanced captions group in all the testing measures. As the researchers explained, input enhancement may have allowed the participants to reflect more on the target language construction, enabling them to transfer and use the explicit knowledge acquired in new contexts. The fact that the learners only improved significantly as regards the use of present perfect and not past simple was not associated to the higher complexity of the latter but their high prior knowledge, which may have resulted in a ceiling effect.

The study conducted by Cintrón-Valentín et al. (2019) with learners of Spanish as a foreign language (university level) from the USA compared the learning of vocabulary and four different target grammatical constructions under three experimental conditions: enhanced captions for vocabulary, enhanced captions for grammar constructions and no-captions. To this aim, they created previewing audiovisual materials to explicitly teach the four target language constructions before video viewing. The researchers created one video for each grammatical construction in order to control for frequency of occurrence and randomize the encounters. Learning was only measured by means of the administration of immediate posttests (vocabulary recognition and translation, together with grammar recognition and translation), as well as delayed posttests (grammar translation test). With respect to vocabulary learning, both captioning conditions (enhanced and unenhanced) scored significantly higher than the no-captions group. Yet, the students exposed to enhanced captions outperformed the participants in the unenhanced captions condition. As for grammar learning, the outcomes of the three groups did not differ significantly in the grammar recognition task, whereas in the production test, the groups exposed to captions scored higher than the uncaptioned condition in two out four target grammatical constructions. Cintrón-Valentín et al. (2019) concluded that saliency and learning are mediated by multiple factors, such as the characteristics of each target language construction, learners' prior knowledge, the number of grammar rules learned simultaneously and the support of contextual cues. In a subsequent investigation, Cintrón-Valentín and García-Amaya (2021) obtained similar results.

All in all, the investigations reviewed in this section suggest that the effects of textual enhancement are not clear-cut, which may be associated to the use of different audiovisual materials (Pattemore and Muñoz, 2022a) and the target language constructions under study (Cintrón-Valentín et al., 2019; Cintrón-Valentín and García-Amaya, 2021). In addition, as regards the use of authentic materials, even when the literature suggests that the use of captions may be enough to foster grammar learning (Pattemore and Muñoz, 2022a), further evidence is required to determine whether the use of captions may enhance learning in less proficient participants and younger learners. The findings obtained with primary school learners (Van Lommel et al., 2006; Y'dewalle & Van de Poel, 1999) need to be complemented by investigations that implement captioned videos and, ideally, extensive interventions.

1.2.1.4 Learners' perceptions of captioned-video viewing

The sections above have undoubtedly confirmed the L2 learning potential of sustained exposure to audiovisual materials and onscreen text (L1 and L2) as long as they match learners' characteristics. Still, the empirical evidence is incomplete if learners' views and experiences are disregarded (Pinter, 2022). Therefore, it is of paramount importance to explore students' levels of enjoyment, their willingness to exploit these materials either inside and/or outside the L2 classroom, the challenges encountered throughout the process, their learning perceptions, viewing self-efficacy and the strategies they use to cope with the viewing process. As regards self-efficacy, it "…is essentially a self- evaluation of how able you feel to carry out a specific task in a specific situation successfully" (Irie, 2022, p.100). Self-efficacy influences the extent to which you are willing to face the challenges encountered when performing a specific task (Irie, 2022). Indeed, the literature has shown that captioned-video viewing encompasses multiple challenges, especially in the case of lower proficiency learners, who may need to use multiple strategies and make a greater cognitive effort while processing the input to achieve appropriate levels of comprehension (Bravo, 2008; Vanderplank, 2016). Therefore, in this investigation, viewing self-efficacy refers to the extent to which L2 learners feel capable of staying on the viewing task by relying on the benefits of multimodal input to fill L2 knowledge gaps and enhance comprehension (Pellicer-Sánchez, 2022). The existing evidence to date has mainly been collected from adult L2 learners, while most of the investigations conducted with younger learners have explored their perceptions as regards the use of L1 subtitles.

In the UK, Vanderplank's (1988) investigation with international university students of different levels of proficiency (from low-intermediate to advanced) provided initial evidence of learners' perceptions on L2 learning from captioned-video viewing. The experiment consisted of 9 sessions where different BBC TV programs (different genres) were shown. Throughout the process, the researcher elicited as much information as possible as regards learners' perceptions of their viewing experience, such as their reliance on onscreen text, the specific words and expressions learnt, their reactions, and the strategies used over time. The results indicated that, at first, the participants considered that the use of onscreen text was highly distracting, however this perception changed over time since a higher amount of practice seemed to be required to process audio, text and imagery simultaneously. In fact, the processing of multimodal input did not seem to lead to cognitive overload, and learners' prior experience with L1 subtitles appeared to facilitate the processing of L2 captions, which is in line with other studies on the area (e.g. Koolstra & Beentjes, 1999; Vanderplank, 2019). Overall, the participants reported that captioned-video viewing contributed to English language learning. With respect to the specific language gains, the participants indicated that the viewing experience triggered the learning of words and expressions, word spelling and the comprehension of different English accents. Specifically, the use of audio and text seemed to facilitate the learning process by reducing viewers' levels of anxiety and making the unknown language items more salient in the input. The fact that the Arabic-speaking participants were found to struggle more in order to follow the speed of captions may be associated to their lower L2 proficiency level and the distance between the L1 and L2 orthographic systems (Hamada & Koda, 2008), which may have affected the L2 reading process.

In Spain, Pattemore et al.'s (2020) investigation explored university students' perceptions of L2 learning through the viewing of 10 episodes of a TV series under different captioning conditions: full L2 captions, Enhanced L2 captions and no captions. At the end of the intervention, the participants mainly reported the learning of words and expressions (47,79% and 61,76%, respectively), whereas just a small percentage of the subjects selected grammar (11,76%) and pronunciation (13,23%) as their language learning gains. Yet, this investigation (see Pattermore & Muñoz, 2020) did find significant grammar learning from captioned-video viewing but the participants seemed to be less aware of the outcomes, which might be explained by the fact that this language aspect tends to be explicitly instructed in regular language courses, whereas the viewing treatment attempted to foster incidental grammar learning. As for the captioning conditions, the participants that were exposed to enhanced captions seemed to be more aware of the learning potential of audiovisual materials, which is why their learning reports were visibly higher in the case of vocabulary and grammar learning. With respect to their preferred subtitling mode, the participants that were inclined to use L1 subtitles prior to the experiment appeared to keep that choice, which was associated to their lower proficiency level and the fact that L1 subtitles have been found to be more suitable for less proficient viewers. Concerning the participants that initially selected L2 captions, their subsequent reports suggested that after sustained exposure to L2 captions, there seems to be a turning point where viewers may feel confident enough to watch uncaptioned videos, which was congruent with Vanderplank's (2019) experiment with adult learners.

Pujadas (2019) also carried out an investigation in the Spanish context, however, her participants were secondary school EFL learners (13-14 years old; Pre-A to A2/B1 proficiency level). The treatment consisted of 24 episodes of an American TV series under two main experimental conditions: L1 subtitles and L2 captions. Learners' attitude towards the use audiovisual input and onscreen text was elicited through questionnaires administered prior to the intervention, right after the experiment and eight months after the posttests, as well as individual interviews with some of the participants. The results suggested that learners' viewing habits changed over time as a function of practice. At initial stages, L1 subtitles seemed to be more suitable to enhance comprehension but then, as a result of the treatment, many participants felt confident enough to either use captions or stop relying on the onscreen text. Still, most of the students that indicated that the text support was not required to improve comprehension or found it distracting were shown to have a higher level of proficiency (A2-B1). As in Vanderplank (1988) and Pattemore et al.'s (2020) investigations, learners believed that their ability to process the input and achieve comprehension improved over time, which is a finding that was also confirmed through the interviews. In addition, around half of the participants (52.3%) found the viewing experience relaxing, which is in line with Vanderplank's (1988) findings. With respect to the perceived L2 gains, the language aspects that were selected the most were form-meaning connections, vocabulary learning, and listening skills development. This finding is congruent with the outcomes of the studies cited above, which suggest that some gains are more salient for the learners regardless of the actual quantitative outcomes (Vanderplank, 1988; Pattemore et al., 2020). Concerning the type of onscreen text, the participants that were exposed to L2 captions seemed to be more aware of the potential L2 gains given that they showed a significantly higher overall feeling of learning. Finally, the interviews indicated that in comparison with learners' regular English lessons, the viewing sessions were more enjoyable and led to higher L2 gains, such as vocabulary learning.

Bravo (2008) elicited data on learners' perceptions of captioned-video viewing from a group of EFL learners of similar characteristics as in Pujadas' (2019) study (secondary school students, 13-14 years old, A2-B1 CEFR level). The 77 EFL learners from Portugal had to watch 10 episodes of the TV series *The Fresh Prince of Bel-Air* (one episode a week), and at the end of the treatment they were administered a questionnaire that allowed the researcher to identify the advantages and disadvantages (challenges) of the viewing experience. Overall, the use of captions was found to support the learning of sound-symbol correspondences, spelling, pronunciation, as well as words and expressions, and sentence construction. Indeed, some participants indicated that being pushed to read in the L2 resulted in L2 learning. In addition, learners reported that the use of onscreen text helped them focus their attention on the input and comprehend what was not identified in the audio. As regards the challenges encountered along the viewing experience, they were mainly associated to the fact that the audiovisual input is processed under time pressure, which affects their capacity to follow L2 captions. This is why, at first, the reading of captions seemed to be challenging and confusing. In addition, the participants reported that the comprehension of L2 input takes a higher amount of time, suggesting that the task may be more cognitively demanding. All in all, the viewing experience does not seem to be exempt from challenges, but the number of advantages reported seems to outweigh the disadvantages, especially when learners are willing to stay on task to improve their capacity to process the input over time.

With students in a similar age range (13-16 years old) as in Pujadas (2019) and Bravo (2008), Zabalbeascoa et al. (2015) explored learners' attitude towards the use of L1-subtitled videos in the English classroom by using a variety of TV programs and collaborative after-viewing tasks over the term. The didactic experience was found to promote collaborative learning and not only increase learners' autonomy, but also their levels of motivation and engagement in the EFL class. As a result, classroom management was also shown to improve over time. Some of participants even reported watching more videos at home as a result of the treatment, which is consistent with Pujadas' (2019) investigation. In addition, some students indicated the use of L1 subtitles and imagery facilitated video comprehension. As expected, only the participants that were more proficient in the L2 expressed their preference towards the use of L2 captions over L1 subtitles. With respect to the challenges experienced throughout the process, the less skilled L1 readers were found to struggle to focus on the videos for a long time since the materials may have been too cognitively demanding for them. The teachers also indicated that the integration of audiovisual resources in the curriculum may be challenging for language instructors since they need plenty of time to develop tasks that may appropriately fit in the syllabus. This finding did not seem to match the outcomes of the investigation carried out by Koskinen et al. (1985) with language teachers in the US. Their results revealed that teachers may be able to create a wide variety of viewing tasks if they receive appropriate guidance and support.

With respect to the studies conducted with primary school learners, the evidence is restricted to the use of L1 subtitles. In Spain, Marzá and Torralba (2015) examined learners' viewing habits and attitude towards the use of onscreen text and the viewing of a 22-minute subtitled cartoon. Specifically, the participants were between 9 and 11 years old, and the data was collected through a survey, discussion groups and teacher's observation notes. The results indicated that learners' exposure to subtitled materials increases with age (Muñoz, 2020b), especially in the case of the students born in immigrant families, who seem to be more accustomed to the use of text support. Overall, the levels of acceptance towards the use of L1 subtitles was high, but they were also shown to be influenced by students' viewing habits, subtitles readability and the extent to which the show was found to be enjoyable. The youngest participants (fourth graders) reported higher levels of difficulty to process the different input sources simultaneously (audio, text and imagery) for 20 minutes, which may be explained by their under-developed L1 literacy skills and their little experience with subtitled videos. In fact, the observations showed that the participants that struggled the most were found to lose their focus after ten minutes.

Collectively, the evidence presented in this section suggests that learners' reading skills and sustained exposure to onscreen text are key to follow the speed of subtitles/captions and process audiovisual input with greater ease. This was confirmed by the investigation conducted by Black (2022) with 17 primary school learners of Spanish (aged 8-9 years old) in the UK. After watching two clips with L1 subtitles, the students participated in semi-structured interviews and workshops. As expected, the participants were shown to have little experience with subtitled videos and some of them found that the subtitles went too fast. Yet, despite the difficulties, the students showed high levels of enjoyment and a very positive attitude towards the treatment. Overall, they felt capable of performing the viewing task and reported that the amount of effort required to comprehend the videos was still appropriate. In addition, most of the participants were willing to watch more videos at home and perceived the viewing experience as an opportunity to learn languages and about other cultures.

In view of all that has been mentioned so far, the challenges encountered over the viewing process may not prevent them from enjoying the learning experience, as long as they feel capable of using certain strategies to compensate (to a certain extent) for their knowledge gaps and stay on task. The promotion of viewing self-efficacy seems to be key to reach the amount of practice required to facilitate the viewing process. Yet, it is important to point out that the materials implemented in these investigations were carefully selected to match the characteristics of the sample groups (Montero Perez et al., 2013).

The key studies cited on L1/L2 learning from viewing in young learners are summarized in Table 4.

Table 4.

Summary: Key studies on L1/L2 learning from viewing in young learners

Researchers	Main focus	Participant characteristics	Audiovisual materials and relevant methodological procedures	Results	***Types of onscreen text***
Alexiou and Yfouli (2019)	Vocabulary learning from captioned-video viewing (words and chunks=109 in total). Receptive and productive form-meaning mapping: meaning recognition and oral-word form recall.	8-9 year-old participants from a language school in Greece (21 in total).	10 episodes of the animated cartoon Charlie and Lola. One per session, which was watched twice (3 weeks). No pretest.	Significant gains over time. Receptive > productive vocabulary knowledge. Gender did not affect the outcomes. 8-year olds > 9-year-olds in receptive vocabulary knowledge. **Absolute vocabulary gains** Meaning recognition (pointing the correct picture): 62.17/109 Oral-word form recall (using a picture as prompt): 35.57/109	L2 captions
Black (2020)	Children's processing of AV content with two types of L1 subtitles: standard and integrated interlingual subtitles (integrated=closer to the speakers).	17 children 8-9 years old	Two clips from an animated film: La Llorona **Conditions:** standard and integrated interlingual subtitles (L2 audio and L1 subtitles). Use of an after-viewing content comprehension test (5 multiple choice questions).	Percentages of fixation time and count on the images: integrated subtitles > regular subtitles. Subtitling conditions did not affect comprehension. 57-86% of fixation time on standard subtitles. This was mainly attributed to two factors: age and lack of subtitled-viewing practice. Regardless of the subtitling condition, mean fixation durations are longer on images than on subtitles. Low percentage of skipped subtitles.	L1 subtitles
Black (2022)	Opinions and experiences of children watching L1 subtitled videos.	17 children 8-9 years old	Two clips from an animated film: La Llorona Semi structured interview Participatory workshops L2 audio-L1 subtitles	Little prior experience watching subtitled videos. Learners' positive attitude towards the use of subtitled-videos. High levels of enjoyment. Most of them reported making a low effort to process the videos. Participants perceived viewing as a learning experience (e.g. learning languages and other cultures). Willing to watch subtitled videos at home. The participants learned to recognize different types of subtitles. Subtitles were not a problem for the participants but many of them found the subtitles too fast.	L1 subtitles
Bravo (2008)	Learners' reflection on the viewing process and the identification of advantages	77 EFL learners from Portugal (ninth grade)	10 episodes of the TV series 'The Fresh Prince of Bel-Air'. One episode a week.	**Advantages:** Captions as support to: Connect sound-symbol correspondence.	L2 captions

	and disadvantages of captioned-video viewing.	13-14 years old	Condition: L2 captions (the participants exposed to L1 subtitles did not participate in this part of the investigation). A2-B1 CEFR level A questionnaire was administered at the end of the intervention.	Learn how words are spelled. Learn words and expressions. Being forced to read in English triggers learning. Help learners focus their attention on the input. Comprehend what was not identified in aural input. Recognize sentence patterns. Learn pronunciation. **Disadvantages:** Processing audiovisual input under time pressure. E.g. processing captions. Comprehension takes time. Captions may be challenging and confusing at early stages. Reading in English seems to be more difficult.	
Casulleras and Miralpeix (2017)	Viewing comprehension and vocabulary learning (word recognition)	11-year-old primary school students.	5-month treatment An episode a week of the animated cartoon Curious George. Conditions: L1 and L2 subtitles.	Comprehension: L1 subtitles > L2 captions. Word recognition L2 captions > L1 subtitles.	L2 captions L1 subtitles
d'Ydewalle and Van de Poel (1999)	The effects of interlingual subtitles on the learning of vocabulary (meaning recognition), morphology and syntax.	3rd, 4th, 5th and 6th grade (8-12 years old). Belgium (Dutch speaking). L2 French instructed since fifth grade.	10-minute long still motion movie. Conditions: reversed subtitles (either in L2 French or L2 Danish), L1 subtitles (Audio in L2 French or L2 Danish), control group (video in L1).	Input in Danish > Input in French (distance) Formal instruction (French) did not play a role. Similarities between languages enhanced learning (Language distance). Greater gains in vocabulary. Small gains morphology and syntax. Learners tended to perform better when the target language was in the audio. Reading in a foreign language takes longer. In terms of vocabulary, there was no significant difference between year levels in L2 Danish. The significantly higher scores of fifth and sixth graders in L2 French was expected since French instruction starts in fifth grade. Non-significant difference between third and fourth graders. **Absolute vocabulary gains** Meaning recognition *Danish vocabulary test* Visual test: (M=50,5%) Auditory test: (M=54%) *French vocabulary test* Visual test: (M=60%) Auditory test: (M=56%)	L1 subtitles Reversed subtitles.
Galimberti and Miralpeix (2018)	The effects of different subtitling conditions on	6th graders (12 years old)	A 22-minute episode: The Suit Life of Zack and Cody	No significant difference in terms of comprehension. Slight differences between groups: L1 subtitles > L2 captions > no captions.	L2 captions L1 subtitles

	vocabulary learning and **comprehension**. Form and meaning recall. Meaning recognition.	Italy L1 Italian L2 English	**Subtitling conditions**: L2 captions, L1 subtitles, no subtitles.	Small but significant vocabulary gains in all the conditions. **Overall recall (differences were not significant when assessing gains separately).** L2 captions > No captions (non-significant) L2 captions > L1 subtitles (significant). **Recognition** L1 subtitles scored higher but differences were not significant. L1 subtitles > L2 captions > No captions. Non-significant differences were associated to the hypothesis that younger learners learn more from the audio than from the subtitles. **<u>L2 captions (absolute gains)</u>** Overall word recall M=1.41 Word-form recall M=1.35 Word-meaning recall M=1.00	
Gesa (2019)	**Vocabulary learning** through subtitled videos (interlingual subtitles). Word form and meaning recall.	Year 6 (11 years old) Catalonia, Spain. Unfamiliar with subtitled materials.	Terms 1 and 2: The Suit Life of Zack and Cody (16 episodes). Term 3: Wizards of Waverly Place (8 episodes). Around 94-95% coverage at K2. **Conditions:** vocabulary activity+subtitled video/Vocabulary activity without video viewing.	The experimental group scored significantly higher than the control group in the third term, not earlier. Overall, small gains and low retention. Word meaning > word form The complexity of the input may have been too high for the target group. The participants seemed to struggle to cope with the input (L2 audio and L1 subtitles) and learn from it. They might have felt overwhelmed with the subtitles. **<u>Relative gains</u>** <u>Written-word form recall</u> Term 1: 19.91% Term 2: 16.53% Term 3: 18.69% <u>Meaning recall</u> Term 1: 9.66% Term 2: 10.17% Term 3: 11.71%	L1 subtitles
Koolstra and Beentjes (1999)	Aural word and meaning recognition. Out-of-school viewing	Grades 4 and 6 Dutch The Netherlands	15-minute documentary **Conditions:** L2 English soundtrack and L1 Dutch subtitles/L2 English soundtrack without subtitles/L1 Dutch soundtrack (control)	Learning with and without subtitles. Better results in the condition with L1 subtitles. Subtitles > no subtitles > Control group Students were listening to the audio. Year 6 > Year 4 (no interaction between age and condition). Positive relationship between subtitled- viewing at home and vocabulary knowledge (in both year levels). Greater influence of this factor than year level.	L1 subtitles

				4th graders knew some words prior to formal language instruction. Aural-word form recognition (absolute gains)=69.94% Meaning recognition (absolute gains)=71.60%	
Koolstra, van der Voort, and van der Kamp (1997)	Relationship between subtitled television viewing at home and reading and decoding skills development.	1,050 Dutch primary school children in Grades 2 and 4 at the outset of the research (8-10 years old).	Administration of questionnaires and other instruments such as: L1 reading skills. Word decoding skills. Intelligence.	Reading comprehension was negatively associated with the time spent watching television. The students that had a more negative attitude towards reading spent a greater amount of time watching television. There was a television's inhibitory effect on reading comprehension. The students with higher reading comprehension were increasingly engaged with leisure-time reading. The opposite direction was found in poor readers. Subtitled-viewing promoted the development of decoding skills. No effects of subtitled-viewing on reading comprehension.	L1 subtitles
Koskinen, Wilson and Jensema (1985)	Teachers and students' perceptions of the use of captioned videos to improve learners' reading skills (comprehension, vocabulary and oral reading fluency).	Teachers and primary school students between grades 2 and 6 (US). Remedial L1 readers or ESL students.	Video excerpts. Use of questionnaires to elicit their perceptions. Students participated in workshops. Teachers were trained to use captioned videos.	Positive evaluation of students and teachers. Teachers=very good and excellent. 90% of the students said that they would like to learn through captioned videos. Teachers reported students' high interest in watching videos. Teachers found creative ways of using the videos.	L1 captions (L2 captions in the case of ESL students)
Kothari, Takeda, Joshi and Pandey (2002)	To examine the effects of L1 intralingual subtitles on **reading skills development** (syllable and word reading ability).	Grades 4 and 5, India.	Captioned-song videos. 3-month experiment (less than 18 hours) **Conditions:** L1 captions/no-captions/control group.	Syllable and word reading improvement, especially in the case of monosyllable words.	L1 captions
Lekkai (2014)	Incidental receptive L2 **vocabulary** learning through subtitled-audiovisual input. Target (unknown) language: Italian.	9-12 years old, Greece. From fourth to sixth grade.	15-minute cartoon (1 episode watched twice). **Conditions:** L2 audio and L1 Greek Subtitles/no subtitles/audio and subtitles in Greek (control group). **Tests:** multiple choice (target word and translation). Aural word-form recognition (yes/no). *No pretest*	Significant higher gains in the subtitles condition. Subtitles > no subtitles > control group Fourth graders < fifth and sixth graders. **Absolute gains:** Meaning recognition: 25.58/30	L1 subtitles and L1 captions (control group).

Linebarger (2001)	L1 reading skills development through captioned-video viewing	76 children L1 English End of 2nd grade (7-9 year olds) The US	Five 4-6-min clips from the Nickelodeon series Pinwheel (5 short sessions). Modified to be suitable for year 1 students. Slower pace than regular videos. **Conditions:** 4x4 captions/no captions. Audio/no audio.	Captions > no captions In word decoding and retention. Yet, the use of audio did not lead to greater gains in the captions conditions. Non-significant differences between conditions on oral reading fluency (insufficient amount of exposure). There was improvement from first to second reading but the scores were not maintained. Captions helped concentrate on and comprehend the critical elements in the story. There was no significant difference between aural and captions conditions. When captions were not present, the participants remembered incidental/non-essential elements of the story. As for the identification of the main idea in the story, the absence of captions led to better results. The participants spent too much time on captions decoding (due to their underdeveloped reading skills). Aural input was required to use the target words more often in an oral recall task. Captions did not have an effect. The researcher suggests that children may be instructed to maximize the processing and learning from captions given that in this investigation, the participants were not given any kind of instruction on how captions had to be processed.	L1 captions
Linebarger, Taylor and Greenwood (2010)	The effects of captioned-video viewing on the development of L1/L2 reading skills. L2= ESL	70 second and third graders in the US. The participants were considered 'at risk readers' and half of them were minority language speakers (ESL). 79% were reading below second-grade level.	Six 30-minute programs **Conditions:** captions vs. no captions. **Measures:** General word-recognition (oral reading). Target word recognition (oral reading). Word-meaning recall. Comprehension: literal (climax) and inferential question (main idea). Nonsense word fluency. Oral reading fluency.	Captions > no captions in most of the measures. Captions led to higher scores in: Episode-word recognition. Meaning recall. Inferential comprehension. Non-word reading. Captions > no-captions in episode-word recognition but only in the participants that had been labelled as at risk or moderately at risk. Captions > no captions in general word recognition but the difference was not significant. Episode-word recognition= low retention. Gains in general word-recognition=greater retention. Literal comprehension remained unaffected by the use of captions. Non-significant improvement in oral reading fluency (amount of practice was insufficient to transfer gains to ORF).	L1 captions (L2 captions in the case of ESL students)
Marzá and Torralba (2015)	Examining habits and attitudes towards subtitled cartoons.	118 Spanish school children (9-12 years old)	22-minute cartoon Zeke's Pad Self-reported data (survey): habit, readability of subtitles	Incidental language learning (e.g. vocabulary) and multicultural awareness. Older participants were mainly familiar with subtitles, especially those born into immigrant families.	L1 subtitles

		Proficient in Catalan (L1 or L2).	and acceptance of the subtitled cartoon. Discussion groups Teacher's observation notes	Very high acceptance at all levels. Influenced by three conditions: habit, readability and enjoyable show. 4th graders reported some difficulties in switching attention between image and subtitle and in having to read for the whole duration of the show (22 minutes). Older participants more familiar with subtitled-viewing.	
Muñoz (2017a)	Influence of age and proficiency on eye movement with L1 & L2 subtitles.	19 primary school learners: grades 5 and 6 9 adolescents (13-16 years old) 12 adults (19-41 years old). Spain	Extracts from 2 episodes of "The Simpsons" **Conditions:** L2 soundtrack with either L1 subtitles or L2 captions.	Greater use of audiovisual materials in the classroom may encourage the participants to watch TV in the L2 at home. Children differed from the older groups in most measures. The primary school children skipped subtitles much less than adolescents and adults. Children also made more fixations on subtitles and spent a longer total time on them than adolescents and adults. Reading in the L2 is more cognitively challenging. Beginner and intermediate level learners spent more total time on subtitles in the L2 than in L1.	L1 subtitles L2 captions
Neuman and Koskinen (1992)	The influence of Captioned-video viewing on vocabulary acquisition (written-word recognition and conceptual recall, contextual word use (recognition), meaning recognition. Combination of word-related and video variables that contributed to vocabulary gains (captions condition). Relationship between students' linguistic proficiency and vocabulary acquisition. -Word-related variables: frequency, conceptual difficulty, importance of the word to the development of the Science concept, and visual support for each word.	129 bilingual year-7 and year-8-students (ESL in US). (2 or 3 years below grade level in terms of educational development)	3 units of Science segments, twice a week for 12 weeks (designed for 8-12 year olds) **Four conditions:** L2 captions/no captions/reading-while-listening/textbook only (control).	Captions: better results in all measures of word knowledge. Contextual support (imagery in video conditions) was a significant predictor of vocabulary learning. Captioned-video viewing enhanced content comprehension. Higher proficiency level=greater vocabulary gains. Captions (absolute gains) Word recognition test=22.15/30 Contextual word use=20.85/27 Meaning recognition=56.56/90	L2 captions (ESL context)

Pujadas (2019)	Learners' perceptions on their viewing experience. How their preferences changed over time.	Eighth graders (13-14 years old) from Spain (Catalonia). Pre-A to A2/B1 proficiency level.	Watching 24 episodes of the TV series Fresh off the Boat (one episode a week). Instruments: 2 questionnaires prior to the intervention (preferences in terms of onscreen text, their use, perceived L2 gains, out-of-school exposure). A questionnaire administered right after the treatment (changes in viewing habits, attitude towards the intervention, perceived learning). A questionnaire administered eight months after the intervention (long-term changes as a result of the intervention). Individual interviews with sample participants.	***After the intervention:*** Most of the participants that reported needing L1 subtitles or L2 captions to enhance comprehension were between the lowest levels of English (Pre-A1 and A1 level). Overall, the participants that indicated that onscreen text was not required to improve comprehension or found the text support distracting had higher levels of proficiency (A2-B1). ***Learning gains and perceptions:*** 73.3% of the participants increased their levels of comprehension throughout the intervention. 30.2% found the activity motivating. 52.3% felt relaxed during the viewing experience. Overall, a higher number of participants reported learning form-meaning connections (67,4%), vocabulary learning (47,7%) and listening skills development (46,5%). The learning gains reported by the participants were higher in the case of the participants exposed to L2 captions. Differences between L2 captions and L1 subtitles conditions reached significance in: form-pronunciation mapping, vocabulary learning, and retention of words and phrases. Overall positive feeling of learning: L2 captions (66.7%) > L1 subtitles (46.3%) Learning about cultural aspects: L2 captions (24.4%) < L1 subtitles (63.4%) Would like to continue watching videos in class: L2 captions (73.3%) < L1 subtitles (90.2%) ***Interviews:*** 17/17 participants found the viewing experience more enjoyable than regular classes. 15/17 believed that the viewing experience led to higher learning gains than their regular classes. 5/8 participants in the captions condition listened to the audio and read the captions for confirmation or supporting comprehension. 15/17 reported vocabulary learning. Only 3/17 reported learning chunks and how to structure sentences. 17/17 said that the viewing task became easier over time.	L1 subtitles L2 captions
Pujadas and Muñoz (2019)	Vocabulary learning (written-word form and meaning recall) from video viewing.	Eighth graders (13-14 years old) from Spain (Catalonia). Pre-A to A2/B1 proficiency level.	Watching 24 episodes of the TV series Fresh off the Boat (one episode a week). **Conditions:** Captions vs. L1 subtitles Vocabulary pre-teaching (previewing activities): yes/no.	The captions+vocabulary pre-teaching group obtained higher gains from the treatment. Recalling form > Recalling meaning in all the conditions Pre-teaching activities led to higher gains independently from the subtitling conditions. Higher proficiency=Higher gains. **Vocabulary pre-teaching:** Captions > subtitles in meaning and form-recall.	L1 subtitles L2 captions

				No-vocabulary pre-teaching: Subtitles > Captions (slightly better in meaning recall). **Interaction between proficiency and activity:** *<u>Form recall</u>* Vocabulary pre-teaching > No pre-teaching in all proficiency levels. *<u>Meaning recall</u>* Vocabulary pre-teaching = No pre-teaching (similar at all proficiency levels. Proficiency may have played a greater role. **<u>Relative gains</u>** <u>Form recall</u> Captions+vocabulary pre-teaching: 30.10% Captions+focus on meaning: 13.02% L1 subtitles+vocabulary pre-teaching: 21.53% L1 subtitles+focus on meaning: 14.30% <u>Meaning recall</u> Captions+vocabulary pre-teaching: 14.54% Captions+focus on meaning: 5.97% L1 subtitles+vocabulary pre-teaching: 8.45% L1 subtitles+focus on meaning: 8.34%	
Pujadas and Muñoz (2020)	Viewing comprehension	Eighth graders (13-14 years old) from Spain (Catalonia). Pre-A to A2/B1 proficiency level.	Watching 24 episodes of the TV series Fresh off the Boat (one episode a week). **Conditions:** Captions vs. L1 subtitles Vocabulary pre-teaching (previewing activities): yes/no.	L1 subtitles > Captions in comprehension. Explicit vocabulary instruction had a small negative effect on comprehension because of the higher cognitive demands (splitting their attention between viewing comprehension and intentional vocabulary learning). Significant effects of vocabulary size in the captions condition. Lexical coverage was a significant predictor of comprehension.	L1 subtitles L2 captions
Teng (2019a)	Vocabulary learning through different captioning conditions and number of repetitions (frequency). Vocabulary knowledge assessed: written-word form recognition, meaning recall, meaning recognition.	6th graders ESL learners (Chinese)	One video (25 minutes)-15 target words. **Conditions:** Different captioning conditions: full captions, keyword captions, no captions. Frequency of occurrence of the target words (1/3). *Only posttest.*	Full captions > keyword captions > no captions. 3 encounters > 1 encounter (regardless the type of captions). **In full captions condition:** Word recognition > meaning recognition > meaning recall. **Absolute gains:** *<u>Form recognition:</u>* Full captioning (1 repetition): M= 11.78/15 Full captioning (3 repetitions): M= 13.14/15 Keyword captioning (1 repetition): M=7.08/15 Keyword captioning (1 repetition): M=10.12/15 *<u>Meaning recall:</u>* Full captioning (1 repetition): M= 4.57/15	L2 captions Keyword captions

				Full captioning (3 repetitions): M= 7.01/15 Keyword captioning (1 repetition): M=2.01/15 Keyword captioning (1 repetition): M=4.89/15 *Meaning recognition:* Full captioning (1 repetition): M= 7.91/15 Full captioning (3 repetitions): M= 10.15/15 Keyword captioning (1 repetition): M=5.71/15 Keyword captioning (1 repetition): M=7.05/15	
Teng (2019b)	Viewing comprehension (global and detailed comprehension). Repeated viewing.	6th graders from Hong Kong (11.57 years old on average)	Two videos (short stories): video 1 watched once (10'28'') and video 2 watched twice (10'20''). **Conditions:** full captions, keyword captions and no captions **Measures:** written recall protocol and a multiple-choice test.	Full captions: more effective for high proficiency learners (global and detailed comprehension). No significant difference between full captions and keyword captions in low proficiency learners. High-proficiency > Low-proficiency in all conditions. Higher performance in full captions condition when watching a video twice (Global and detailed comprehension). Higher performance in full captions condition when watching once in terms of global comprehension, but not detailed comprehension. Watching video twice > watching video once (all captioning conditions). Low proficiency level: watching once did not lead to higher gains in detailed comprehension in the full captions conditions. They must have struggled to follow the captions. Full captions supported global comprehension.	L2 captions Keyword captions
Teng (2022)	Vocabulary learning through different captioning conditions. Vocabulary knowledge assessed (receptive): form, meaning and use. Use: in which sentence the target word was used correctly.	6th graders (11-12 years old) ESL learners (Chinese)	4 short storytelling videos on YouTube (16 minutes). 20 target words. **Conditions:** Advance-organizer strategy before viewing (yes/no). It intended to reduce the cognitive load. Full captions/Keyword captions/Glossed full captions/Glossed keyword captions.	Glossed full-captions condition led to greater vocabulary gains. The use of advance-organizer strategy contributed to vocabulary learning. It supported the subsequent processing of images and verbal input. Glossed full captions made the target words salient in the input. Greater effect of captions than the use of advance-organizer strategy. Glossed-full captions conditions+advance-organizer strategy led to higher gains. Word recognition > meaning recognition > word use. Without advance organizer: glossed condition > non-glossed condition. **Absolute gains *(M)*** *Form* Full captioning:6.900/20 Keyword captioning: 5.100/20 Glossed keyword captioning: 8.933/20 Glossed full captioning: 11.033/20 Full captioning+ graphic organizer: 15.167/20 Keyword captioning+ graphic organizer: 13.33/20	L2 captions Keyword captions

				Glossed keyword captioning+ graphic organizer:16.967/20 Glossed full captioning+ graphic organizer: 19.033/20 *Meaning* Full captioning: 5.033/20 Keyword captioning: 2.867/20 Glossed keyword captioning: 6.967/20 Glossed full captioning: 8.933/20 Full captioning+ graphic organizer: 13.167/20 Keyword captioning+ graphic organizer: 10.967/20 Glossed keyword captioning+ graphic organizer: 15.033/20 Glossed full captioning+ graphic organizer: 16.867/20 *Use* Full captioning: 2.933/20 Keyword captioning: 1.167/20 Glossed keyword captioning: 4.933/20 Glossed full captioning: 6.933/20 Full captioning+ graphic organizer: 10.967/20 Keyword captioning+ graphic organizer: 8.933/20 Glossed keyword captioning+ graphic organizer: 12.967/20 Glossed full captioning+ graphic organizer: 14.633/20	
Tragant and Pellicer-Sánchez (2019)	Comparing children's processing of multimodal input.	5th graders (10-11 years old). Students had little prior experience with subtitled materials.	An episode of Charlie and Lola and the same story in audiobook format.	No significant difference between both input conditions in terms of comprehension. Learners spent more time processing the text than the images in both formats. Images did not distract learners from reading the text. Negative relationship between average fixation duration and L2 vocabulary knowledge in the case of the video but not the book. The dynamic nature of images in the video condition increased learners' attention to this mode. Video: longer average fixations on the images than on the text. Book: longer average fixations on the text than on the book. High individual variability in the video condition.	L2 captions
Van Lommel, Laenen and d' Ydewalle (2006)	Learning L2 **grammar** through subtitled-audiovisual input (5 grammar rules) Unknown language: Esperanto.	6th grade (primary school) (11 years old) 6th grade (secondary school) (17 years old)	25-minute cartoon (1 episode). **Conditions** Experiment 1 a) Presentation of rules (yes/no)	Movie only condition: no incidental grammar learning. Strong effect of rule presentation (particularly in the case of older participants). In experiment 1, the movie helped identify some of items in the test. However, the movie did not help figure out the rules. Children learned more when they were presented the rules in advance and when they were explicitly told to learn from the video.	L1 subtitles Reversed subtitles.

		Belgium.	b) Use of a video with reversed subtitles (L1 audio, L2 subtitles) (yes/no) Test: translating sentences into Esperanto (Multiple choice). Experiment 2 (L2 audio-L1 subtitles). a) Presentation of rules (yes/no). b) Presentation of rules depending on the target construction (within-participants). c) Instructing participants to learn from the video (yes/no).	A single video is not enough to learn grammar.	
Zabalbeascoa, González-Casillas and Pascual-Herce (2015)	Students' perception of the use of L1-subtitled videos in the English class. Listening comprehension.	Secondary school students from Barcelona (2 schools). Age: 13-16 years old.	Wide variety of videos, as well as previewing and post-viewing tasks.	**The treatment was found to be beneficial:** +autonomy +motivation and engagement Classroom management was improved. Promoted collaborative learning. Treatment groups > control group in listening comprehension. **Challenges detected:** Low achievers (L1 reading skills): struggled to focus on the videos. The teachers addressed the challenge of designing viewing tasks on their own and integrating them to the syllabus. **Other findings:** Some students continued watching videos at home. Higher proficiency learners preferred L2 captions. Images and L1 subtitles supported comprehension.	L1 subtitles

1.3 The role of learner-related factors in L2-learning from audiovisual input

The number of studies on the role of individual differences (IDs) in L2 learning has grown exponentially over the years. Their importance lies in the fact that "…learner traits and characteristics may have an impact on learning processes, behaviors, and outcomes" (Li et al., 2022, p.3). In second language acquisition (SLA), the literature has moved from identifying the L2 learners that should be instructed to identifying the factors that explain interindividual variation in L2 learning (R. Ellis, 2022). In this investigation, we focused on the influence of age (primary school learners' outcomes), L2 proficiency (vocabulary knowledge, and listening and reading skills), L1 reading skills, and cognitive factors (working memory and visual processing speed).

1.3.1 Primary school learners' characteristics

Primary school education coincides with the developmental period of middle childhood (6-11/12 years old), which is associated to fast and marked changes regarding physical, socio-emotional and cognitive development. Throughout this stage, children's cognitive processes become gradually more efficient due to the increase in their reasoning and problem-solving skills, as well as self-regulation, executive functions and working memory capacity (Delgiudice, 2018). Accordingly, the multiple changes undergone by children along their primary school years influence their L2 learning process and experience, moving progressively from implicit to more explicit learning mechanisms (Holmes & Myles, 2019). Although primary school learners enjoy the implementation of fun activities, they become gradually more aware of their learning process and their limitations, which is why their sense of progress and actual achievements are directly associated to their levels of motivation (Muñoz, 2017c; Myles, 2022).

The evidence has shown that younger children are good at learning languages intuitively and implicitly, however, their rate of success relies heavily on their amount of exposure to the target language (Muñoz, 2006; Muñoz & Spada, 2018). Therefore, in foreign language contexts, where children's exposure is limited to the L2 classroom, younger learners have been found to be less efficient than older learners (in late primary or secondary school years), who appear to benefit from their advanced cognitive development and the stronger foundations of their L1 literacy skills (Andringa, 2022; Muñoz, 2006; Muñoz &

Spada, 2018; Singleton & Pfenninger, 2019). Indeed, in incidental vocabulary learning conditions, older instructed learners have been found to acquire words more efficiently (Kim & Webb, 2022a) due to their greater sensitivity to frequency effects (Uchihara et al., 2019) and cognateness (Muñoz, 2020a).

Given that primary school students are not yet fully autonomous learners, they strongly depend on local policies and instructional programs to have access to the quality and plentiful input required to make greater progress over time (Holmes & Myles, 2019). Considering that in most regular language programs instruction takes place for only a few hours a week, schools may explicitly encourage learners to do a series of informal activities that might help them compensate for their lack of exposure (e.g. TV viewing and gaming) (Webb, 2015). Research on out-of-school contact has consistently demonstrated that in contexts where there is plentiful access to L2 input since an early age (e.g. Belgium and Denmark), children already show significant L2 gains prior to formal instruction (De Wilde et al., 2019, Muñoz et al., 2018; Puimège & Peters, 2019a; Prophète et al., 2022). By contrast, in L2 input-limited contexts with established dubbing tradition, children and families do not seem to be aware of the advantages of informal activities in the target language (Black, 2022), which is why their extramural exposure remains low in early primary school (Marzá & Torralba, 2015). Thus, in contexts where children are not widely exposed to informal activities in the target language, research is strongly required in order to explore how young learners engage with these activities inside the classroom and identify the factors that should be considered to ensure that their implementation actually leads to L2 learning.

As regards captioned-video viewing, some of the characteristics of primary school learners anticipate that there are important factors to consider before its actual implementation in L2 classrooms. Although the literature suggests that the use of bimodal verbal input facilitates decoding (Pellicer-Sánchez, 2022), primary school learners are still developing their cognitive and L1 literacy skills (Holmes & Myles, 2019). To illustrate, the starting age at which they might be able to cope with the speed of captions for comprehension and learning purposes is still unclear (see section 1.2.1.2). Vanderplank (2016) mentions the age of 10 as a possible threshold for L1 subtitles, but further research is still needed to shed light on this issue given that the literature has already shown positive experiences with 8-9-year olds (Alexiou & Yfouli, 2019). Although it is true that young learners make a greater cognitive effort when processing both, subtitles and captions (Muñoz, 2017a), it is also important to consider the input demands. Based on the findings obtained with adult L2 learners, viewers may benefit from the use of audiovisual input and compensate for knowledge gaps depending on the TV genre and the extent to which imagery supports comprehension (Durbahn et al., 2022; Suárez et al., 2021). Yet, in view of their developing cognitive skills, it is also important to explore children's capacity to integrate verbal and non-verbal input effectively, which seems to be crucial in the processing of multimodal input (Pellicer-Sánchez, 2022). Equally important, episode duration is also a factor to consider when working with young learners given that the great cognitive resources allocated on viewing seem to prevent them from staying on task for more than ten minutes, particularly in the case of low achievers (Marzá & Torralba, 2015; Zabalbeascoa et al., 2015).

1.3.2 L2 proficiency and vocabulary knowledge

The extent to which L2 learners benefit from audiovisual input seems to be strongly predicted by learners' proficiency level (Gesa & Miralpeix, 2022). L2 proficiency has been strongly associated to learners' vocabulary knowledge and its key role in L2 comprehension (Montero Perez, 2020; Miralpeix & Muñoz, 2018; Stæhr, 2008), which is why this section focuses on the role of vocabulary knowledge on L2 learning through viewing. The empirical evidence has demonstrated that a minimum level of vocabulary knowledge is required in order to show appropriate levels of comprehension in different modalities (Durbahn et al., 2020, 2022; Pellicer-Sánchez & Webb, 2022; Schmitt et al., 2011; van Zeeland & Schmitt, 2012). Once this threshold is surpassed, L2 learning is likely to occur due to the less effortful comprehension process and the availability of enough cognitive resources to notice unknown target language constructions (Kim & Webb, 2022a; Lin & Siyanova-Chanturia, 2015; Montero Perez, 2020). By the same token, more proficient L2 learners have richer knowledge of the L2 stored in long-term memory to figure out the meaning of unknown language constructions (Kim & Webb, 2022a; Montero Perez, 2020). Drawing on this tendency, the strong positive correlation between L2 knowledge and L2 gains has also been addressed in the literature as Matthew effect or the-rich-get-richer (Stanovich, 1986).

Although the lexical coverage that ensures appropriate levels of comprehension in viewing has been found to be less demanding than in reading-only and listening-only conditions (80%; Durbahn et al., 2022), the majority of studies on audiovisual input have identified L2 proficiency (Gesa & Miralpeix, 2022) and vocabulary knowledge as significant predictors of L2 learning in different age groups (e.g. Alexiou, 2015; Montero Perez et al. 2013, 2018; Peters et al., 2016; Peters & Webb, 2018; Pujadas & Muñoz, 2019). Yet, the extent to which vocabulary knowledge predicts word learning has also been found to vary among investigations. To illustrate, Fievez et al.'s (2020) investigation with low-intermediate vocational school learners of French explored the role of L2 vocabulary knowledge on incidental vocabulary learning (word recognition, meaning recall and meaning recognition). In this investigation, vocabulary knowledge was found to be a significant, albeit weak predictor of word learning, which was contrasted with the stronger effects obtained in previous investigations (e.g. Montero Perez et al., 2018; Peters & Webb, 2018). This outcome was attributed to the small variability in L2 vocabulary knowledge scores obtained by the participants (Fievez et al., 2020). Among the few exceptions where vocabulary knowledge has not been found to be significant (e.g. Frumuselu, 2015; Rodgers, 2013), the study conducted by Suárez et al. (2021) with university students found that, overall, vocabulary size significantly increased the odds of word learning through captioned-video viewing; yet, this factor did not reach statistical significance when watching a documentary, which was associated to the moderating effects of imagery (Suárez et al., 2021).

Viewers' proficiency level has also been found to influence their reliance on onscreen text. As expounded in previous sections, the processing of L2 audio may be quite challenging for L2 learners, especially at low L2 proficiency levels. Regardless of listeners' familiarity with each individual word of a stream of speech, lexical segmentation has been shown to be a struggle, especially when the audio is produced by native speakers or in unfamiliar accents (Charles & Trenkic, 2015; Mitterer & McQueen, 2009). This struggle might be explained by the fact that, at earlier stages, formally instructed foreign language learners may have greater knowledge of the L2 in written representation, which is why the use of bimodal verbal input has been studied as a synergy that may facilitate input processing and enrich learners' L2 knowledge (Pellicer-Sánchez et al., 2018). In the case of audiovisual input, research has shown that the use of captions and imagery compensates for learners' knowledge gaps to reach appropriate levels of comprehension (Pujadas & Muñoz, 2022; Durbahn et al., 2020, 2022). Indeed, the investigation conducted by Pujadas and Muñoz (2022) with university

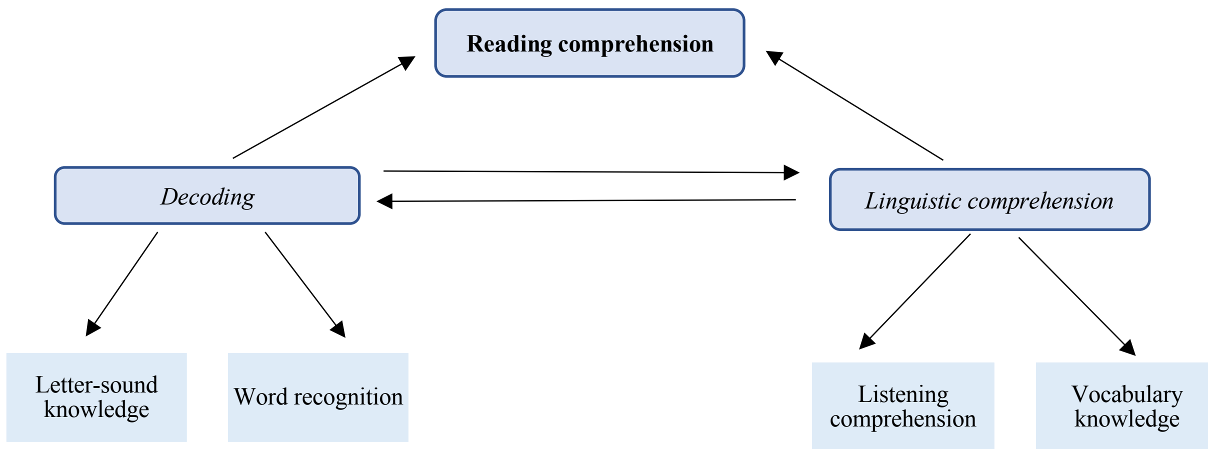
EFL learners in Spain indicated that only at around C1 level of proficiency (CEFR), viewers may attain 80% of comprehension without the support of captions. Thus, the use of onscreen text support appears to be crucial for lower proficiency learners if the goal is comprehension and L2 learning.

With respect to the reading of onscreen text, the empirical data obtained from eye-tracking investigations has also shed light on the influence of L2 proficiency on the reading process. Although the evidence suggests that the processing of captions may be automatic regardless of learners' proficiency level (Gass et al., 2019), the amount of time devoted to reading appears to depend on L2 proficiency. Specifically, at lower proficiency levels, the processing of onscreen text seems to be more effortful, increasing the amount of time viewers spend on captions/subtitles (Gass et al., 2019; Muñoz, 2017a; Tragant & Pellicer-Sánchez, 2019). This finding is unsurprising considering the complexity of L2 reading and the fact that learners' ability to read texts with ease and high levels of comprehension is mainly accounted by L2-related factors (Alderson et al., 2016; Sparks, 2021), such as L2 vocabulary knowledge (Miralpeix & Muñoz, 2018; Stæhr, 2018). The strong relationship between L2 reading and L2 proficiency has been explained in light of the Simple View of Reading model (Gough & Tunmer, 1986; Hoover & Gough, 1990; Tunmer & Chapman, 2012) (see Figure 3), which posits that reading comprehension is mainly explained by word decoding and oral general language comprehension. As regards word decoding, it is associated to learners' knowledge about the target language sounds and the alphabetic system, and how they map onto each other; whereas oral comprehension refers to the extent to which learners comprehend oral input, moving from words (vocabulary) to text level (listening comprehension). In view of the complexity of L2 reading skills, one may assume that this factor may play a role in the extent to which L2 learners' benefit from captioned-video viewing. The following section (1.3.3) further discusses this possibility.

Taken together, the evidence presented in this section supports the notion that L2 learners' proficiency level and vocabulary knowledge influence the processing of audiovisual input and determine how profitable this activity may be (Gesa & Miralpeix, 2022; Montero Perez, 2020; Vanderplank, 1988).

Figure 3.

The modified Simple View of Reading



Note. (adapted from Tunmer & Chapman, 2012, p. 464).

1.3.3 L1 and L2 reading skills

It is now well established from a variety of studies that, in the case of L2 learners, captions support the processing of audiovisual input (Mayer et al., 2020). However, few researchers have focused their attention on the direct relationship between L2 reading skills and L2 learning from captioned videos (Muñoz et al., 2022), and no one to the best of our knowledge has studied the specific influence of L1 reading skills in this regard. Reading skills may be particularly relevant in the case of young learners due to their developing L1 reading skills (Holmes & Myles, 2019), the effort required to cope with the speed of captions (Muñoz, 2017a), as well as the need to integrate verbal and pictorial information to achieve appropriate levels of comprehension (Pellicer-Sánchez et al., 2020; Sadoski & Paivio, 2013; see section 1.2.1.2).

In foreign language settings, it is clear that learners do not have to learn to read in the L2 from scratch. The literature suggests that learners progressively assimilate and accommodate their linguistic infrastructure to the characteristics of the L2 (Birch & Fulop, 2021; Jiang et al., 2019; Perfetti et al., 2007), which is a process that relies on their L2 proficiency and familiarity with the characteristics of target language (Jiang et al., 2019). At earlier stages, learners' L1 orthography may support and facilitate L2 reading to compensate for L2 knowledge gaps and lack of practice, as long as there is an overlap between both systems (Birch, 2015). Thus, one may expect that at least in the case of young L2 learners, both, L1 and L2 reading skills might play a role in the processing and learning through captions. For instance, in the study with young L2 learners (10-11 years old) by Tragant and colleagues (2019), the development of L2 reading fluency through graded readers (with and without audio support) was significantly predicted by L1 reading fluency (23%). Yet, it is worth mentioning that research has also shown that L1 reading skills may only compensate to a certain extent for learners' knowledge gaps (Yamashita, 2002), given that in different age groups, learners' L2 reading comprehension has been found to be mainly explained by L2-related factors rather than L1-reading skills and the underlying cognitive factors (Alderson et al., 2016; Jeon & Yamashita, 2014). In fact, the investigation by Kormos et al. (2019) with sixth graders indicated that reading-while-listening differed from the reading-only and listening-only conditions in that (lower-level) L1 reading skills did not explain readers' performance significantly, which was associated to the facilitating effects of bimodal verbal input. Therefore, the extent to which L1 and L2 reading skills may influence the reading of onscreen text and foster learning seems uncertain.

With respect to the role of L2 reading skills, the investigation by Muñoz et al. (2022) with university EFL learners from Spain (B2 CEFR level) attempted to fill this knowledge gap by examining the influence of L2 reading efficacy (reading speed and comprehension; see Llanes, 2018) on vocabulary learning through repeated captioned-video viewing. The results indicated that L2 reading efficacy was not a strong predictor of word learning, given that its effects were overshadowed by previous vocabulary knowledge. This outcome was contrasted with the results that showed significant effects of sound recognition (as aptitude component) on word meaning recognition. As Muñoz et al. (2022) hypothesized, L2 reading efficacy might be a significant predictor at lower proficiency levels. Due to the lack of studies in this regard, the following paragraphs will discuss the potential role of L2 reading skills on L2 learning from viewing on the grounds of the literature on (L2) reading.

Skilled reading encompasses lower-level and higher-level reading process, as well underlying cognitive processes (Grabe & Stoller, 2020). From an information-processing perspective, the automatization of lower-level reading skills, such as word decoding is key to attain appropriate levels of comprehension (Nassaji, 2014). Yet, the automatization of lower-level reading skills is only attained with plenty of practice and exposure to print, which is a condition that is barely met in foreign language contexts (Grabe & Stoller, 2020). When lower-level reading processes do not work fluently, learners' cognitive effort increases and text comprehension is hindered (Grabe & Stoller, 2020; Nassaji, 2014). By the same token, Sadoski and Paivio (2013) drew on the Dual-Theoretical model of reading to confirm that in the case of multimodal input, inefficient text decoding affects comprehension by hampering the associational and referential processing between verbal and non-verbal information (i.e. text and imagery). In other words, when learners' cognitive resources are concentrated on lower-level reading skills, the reading process becomes effortful, reducing the odds of learning from the input.

As for the role of lower-level reading skills, it is important to point out that the literature has also shown that their influence is not stable over time. For instance, the longitudinal investigation by Verhoeven and van Leeuwe (2012) with L1 and L2-Dutch learners at primary school level explored the effects of word decoding and listening skills on the development of reading comprehension (from first to sixth grade). The results showed evidence of learners' reliance on both, decoding skills and listening comprehension, which is consistent with the Simple View of Reading model (Gough & Tunmer, 1986; Hoover & Gough, 1990). However, the results also indicated that the influence of word decoding skills decreases over the years, whereas the effects of L2 listening comprehension increases with age. Based on this outcome and the facilitating effects of aural support in reading-while-listening and viewing conditions as regards text decoding (Serrano & Pellicer-Sánchez, 2019) and the integration between verbal and non-verbal modes (Pellicer-Sánchez et al., 2020), it may also be hypothesized that in late primary school students, the effects of reading comprehension may be more associated to linguistic comprehension processes, rather than text decoding. It is also worth mentioning that a number of studies with fifth and sixth graders has indicated that the synergy between aural and written input may not necessarily foster greater levels of comprehension (Serrano & Pellicer-Sánchez, 2019) nor L2 gains (Tragant et al., 2019) in comparison with the reading-only condition under relatively short interventions. Hence, congruent with Muñoz et al.'s (2022) findings, it should not be assumed that learners' reading ability as a whole may predict learners' gains from viewing. This is an important issue that still needs to be carefully examined.

1.3.4 L2 Listening skills

The majority of studies on audiovisual input have focused on viewing comprehension or the development of listening skills (see section 1.2.1.1), rather than studying the role of L2 listening in language learning from viewing, with few exceptions (e.g. Pattemore & Muñoz, 2020; Pujadas & Muñoz, 2019; Suárez & Gesa, 2019). It is widely accepted that L2 viewing without text support may be quite challenging for lower proficiency learners (Pujadas & Muñoz, 2022; Vanderplank, 2019). Therefore, one may assume that L2 proficiency, including L2 listening skills may play a significant role in L2 learning from viewing without captions. Nonetheless, the scant existing evidence indicates that listening skills may also predict the outcomes under the presence of captions, which is a factor that may not only be attributed to the input received through the aural channel but the general comprehension processes involved while viewing.

The investigations by Pattemore and Muñoz (2020), and Pujadas and Muñoz (2019) integrated the score of an L2 listening comprehension test (OPT: Oxford Placement Test; Allan, 2004) as part of a proficiency index to assess the influence of the latter on grammar and vocabulary learning, respectively. Both studies showed significant effects of L2 proficiency on learning, thus, it is reasonable to assume that listening skills, as part of the proficiency index, influenced the extent to which L2 learners benefitted from captioned-video viewing; however, the exact contribution of L2 listening skills was not specified. This question was answered by Suárez and Gesa (2019), whose analyses showed a clearer picture of the role of listening skills in intentional vocabulary learning (word form and meaning recall) through captioned videos. The participants (secondary school and university EFL learners from Spain) were administered the same instrument employed in the investigations aforementioned (OPT; Allan, 2004). In line with Pattemore and Muñoz (2020), and Pujadas and Muñoz (2019), the results indicated that L2 listening significantly predicted both, written-word form and meaning recall, which was associated to learners' proficiency level (Miralpeix & Muñoz, 2018; Stæhr, 2008), and the availability of cognitive resources to identify and learn unknown vocabulary words (Lin & Siyanova-Chanturia, 2015). In the case of written-word form recall, learners' listening skills appeared to be key in the identification of aural word form representations, which had to be linked to their corresponding written forms; while in meaning recall, listening skills were associated to learners' higher comprehension processes, which determine learners' ability to derive the meaning of unknown words (Suárez & Gesa, 2019).

The findings obtained by Suárez and Gesa (2019) give some insight into the potential roles that L2 listening skills may play in L2 learning from captioned-video viewing. To start with, vocabulary learning requires students' capacity to match aural and written representations, especially in the case of non-transparent languages since learners need a greater amount of time and practice to become familiar with the orthographic patterns of the target language (Sun-Alperin & Wang, 2008). Thus, skilled listening may allow learners to take greater advantage of the synergy between audio and text due to their capacity to process and integrate both modalities efficiently. In addition, based on the Simple View of Reading model (Gough & Tunmer, 1986; Hoover & Gough, 1990; Tunmer & Chapman, 2012), listening comprehension is closely connected to reading comprehension (Proctor et al., 2005), which is why learners' capacity to comprehend the videos might be mediated by listening comprehension, regardless of the support of captions. In like manner, learners' ability to figure out the meaning of target language constructions may rely on general comprehension processes, as well as learners' capacity to decode and integrate the information obtained through the verbal and non-verbal channels (Sadoski & Paivio, 2013). All in all, the scant evidence presented in this section suggests that listening skills may play a significant role in L2 learning through captioned videos. Still, it has not yet been explored whether these outcomes might also be obtained with primary school learners.

1.3.5 Cognitive skills

L2 learners' cognitive skills, as other individual differences, have been found to influence learners' performance and outcomes in L2 learning. Still, this variability among participants does not only seem to be encountered in adult language learners, but also young students (Porter, 2017). In this investigation, we assessed the influence of working memory (phonological loop and central executive), and visual processing speed.

1.3.5.1 Working memory

Working memory is defined as "a limited capacity system for the temporary maintenance and processing of information in the support of cognition and action" (Baddeley et al., 2021, p.10). SLA research has been highly influenced by the construct of a multi-componential working memory model (Baddeley, 2015; Baddeley & Hitch, 1974), which has evolved over the years in response to the findings that have emerged from extensive research in the fields of psychology and cognitive science (Baddeley, 2015; Wen & Jackson, 2022). Essentially, working memory comprises three main components: the phonological loop, the visual sketchpad, and the central executive. As for the latter, it is represented as an attentional control system that manages the phonological loop and the visual sketchpad, which work as storage subsystems for verbal and visual information, respectively (Baddeley, 2015). As depicted in Figure 4, the episodic buffer was later incorporated by Baddeley and colleagues as a passive multimodal storage of limited capacity that combines the information registered in the phonological loop and the visual sketchpad (Baddeley, 2015; Baddeley et al., 2011). In the field of SLA, research has primarily examined the role of the phonological loop[1] and the central executive[2] in the L2 learning process (Porter, 2017; Wen, 2015).

With respect to the instruments used to assess working memory capacity, PSTM has been measured by means of tests that simply focus on its verbal storage component, such as the forward digit span test (WISC battery; Wechsler, 2014) and non-word repetition (e.g. Porter, 2017); whereas among the measures employed to tap complex working memory capacity, researchers have reported the use of instruments that assess storage, processing and manipulation of information, such as the backward digit span test (Wechsler, 2014) and the reading span task (Unsworth et al., 2005) (e.g. Kormos & Sáfár, 2008; Suárez et al., 2021). The literature points out that the measures used to assess the capacity of each working memory component should be analyzed separately due to the different cognitive operations elicited through each instrument (Kormos & Sáfár, 2008). This is the case of the administration of the forward and the backward digit span tests with young learners, where the central executive has only been found to play a role in the backward digit span (Alloway et al., 2006; Dehn, 2022; Service & Simard, 2022). This factor explains why a child who

---

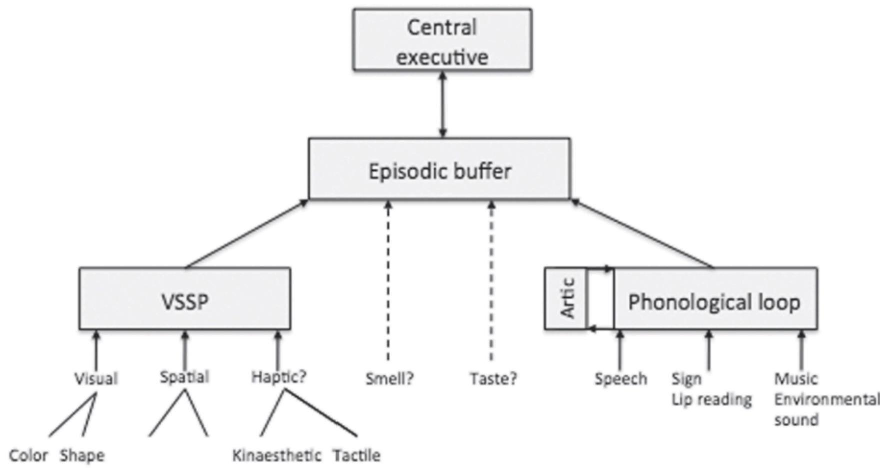
[1] In this dissertation, the terms *phonological loop* and *phonological short-term memory (PSTM)* will be used interchangeably.

[2] In this dissertation, we will consistently use the term complex working memory to address the working memory component that integrates the storage and processing of information

shows an adequate performance at simple working memory tasks may perform poorly at the ones that involve information processing (Dehn, 2022).

Figure 4.

Baddeley et al.'s (2011) multi-component working memory model



Note. (Baddeley et al., 2011, p.1399).

The literature on the role of working memory on L2 learning suggests that the research outcomes are not clear-cut. Yet, certain patterns have been identified, suggesting that the extent to which PSTM and complex working memory influence L2 learning depends on multiple factors, such as learner characteristics (e.g. L2 proficiency, age), and the language aspects under study (vocabulary, grammar, receptive language skills). Wen and Jackson's (2022) review of the literature on the effects of working memory in L2 learning indicates that, overall, working memory is a significant, albeit weak, predictor of L2 learning considering the small effect sizes obtained in the existing meta-analyses (r=0.18-0.16). In addition, when contrasting the effects of complex working memory and PSTM, the results indicate that the former may be a stronger predictor of L2 learning (Wen & Jackson, 2022).

With respect to learner characteristics, the empirical evidence suggests that PSTM has a stronger influence at early L2 learning stages, especially in the case of young L2 learners (Wen & Jackson, 2022; Wright, 2015) in the areas of vocabulary and lexically driven grammar learning (Wright, 2015). As for complex working memory, its effects have mostly been studied with adolescent and adult L2 learners (Wright, 2015), showing evidence of stronger effects at lower proficiency levels when learners are requested to perform cognitively demanding tasks that involve attentional and executive control, such as the noticing of target language constructions, L2 listening, L2 reading, bilingual interpreting, and editing academic writing (see Wen & Jackson, 2022). All in all, the evidence suggests that the influence of PSTM declines as learners' L2 proficiency level increases (Serafini & Sanz, 2016; Wen & Jackson, 2022), given that learners' familiarity with the target language seems to result in their greater reliance on the information stored in long-term memory (e.g. vocabulary knowledge) (Masoura & Gathercole, 2005). As for complex working memory, it seems to be a strong predictor at lower proficiency levels (e.g. Serafini & Sanz, 2016), whereas at later learning stages, complex working memory effects have not been found to be clear-cut (Wen & Jackson, 2022). Indeed, some studies have reported a significant influence of complex working memory in more advanced L2 learners concerning different language aspects, such as L2 vocabulary learning (e.g. Yang et al., 2017) and general proficiency level (e.g. Mitchell et al., 2015).

The findings obtained in Kormos and Sáfár's (2008) investigation with secondary school learners from Hungary (15-16 years old) seem to contradict the tendencies identified in SLA studies in that the group of intermediate learners relied on PSTM to score higher at the proficiency tests, while the beginner learners' performance was found to be influenced by their complex working memory capacity. Specifically, in the case of the beginner learners, complex working memory was shown to correlate significantly with their general proficiency scores, reading, listening, speaking, and the use of English test (vocabulary and grammar). With respect to the intermediate learners, PSTM correlated significantly with their general proficiency scores, the composition task and use of English (vocabulary and grammar). These conflicting results were attributed to the nature of the L2 learning process of each group. The beginner group attended an intensive program that favored explicit instruction to trigger learners' greater progress over time. Therefore, the fact that the beginner learners' outcomes were associated to their complex working memory capacity was explained by the cognitive demands involved in their instruction (e.g. attention) (Kormos & Sáfár, 2008). By comparison, intermediate learners may rely on more implicit learning mechanisms to promote vocabulary growth, which seems to be associated to the verbal storage component of working memory (Kormos & Sáfár, 2008). On the whole, the results obtained by Kormos and Sáfár (2008) appear to be consistent with the literature that suggests that working memory may play a more significant role under explicit L2 learning conditions (Wen & Jackson, 2022). This is exemplified in a study conducted by Li et al. (2019) on the effects of working memory (measured through an operation span test) on L2 learning under different instructional conditions. The participants (eighth graders, 13-15 years old) were taught the English past passive by means of form-focused or meaning-focused instruction. As for the former, it consisted of four possible instructional conditions: 1. Pretask + task, 2. Pretask + within-task feedback + task, 3. Within-task feedback + task, 4. Task + after-task feedback, while meaning-focused instruction was restricted to the completion of a task. The results revealed that working memory only explained the outcomes under the within-task feedback condition (with and without pretask instruction), suggesting that working memory plays a significant role in form-focused instruction under heavier cognitive demands (i.e. completing a task and processing feedback).

Overall, the literature suggests that PSTM and complex working memory have differential effects on the learning/development of diverse language aspects (Grabe, 2009; Wen, 2015, p.50). To start with, PSTM seems to play a significant role in the learning of vocabulary, including formulaic language and target grammatical constructions of different levels of complexity. In L2 comprehension (listening and reading), PSTM has been found to play a key role in word decoding and the storage of phonological information for further consultation; while in L2 production, PSTM has been shown to be predictive of the use of narrative vocabulary at early stages, and grammar accuracy in further stages (Grabe, 2009; Wen, 2015, p.50). With respect to complex working memory capacity, this factor has been found to play a significant role in higher level comprehension processes (listening and reading) by facilitating the processing of syntactic and semantic information, and inhibiting the trivial information that is not required to achieve appropriate levels of comprehension. In L2 interaction, complex working memory supports the noticing of corrective feedback, while in L2 production, it appears to predict learners' performance in terms of language accuracy (Grabe, 2009; Wen, 2015, p.50).

1.3.5.1.1 Working memory effects in primary school learners

As regards the role of complex working memory capacity in primary school students' L2 learning, further research is still required given that the evidence has primarily focused on the effects of PSTM (Wright, 2015). The main reason behind this lack of research may be associated to the complexity of the tasks and children's under-developed cognitive and language skills. In fact, research on working memory development in childhood has shown that children's working memory capacity may not reach adult-like levels before the age of 14, (Gathercole et al., 2004; Wright, 2015). Therefore, the mixed findings or non-significant effects of complex working memory obtained to date may be expected until children's cognitive abilities are fully developed (Wright, 2015). By way of illustration, in a study on predictors of L2-English reading skills, Alderson et al. (2016) administered a backward digit span test to measure the effects of complex working memory capacity in three age groups from Finland: 10, 14 and 17. The results revealed significant, albeit weak effects of complex working memory in the older groups but not the youngest. Similar results were obtained by Pattemore and Serra (2021) with sixth graders from Spain (M=11.8 years old) by also employing the backward digit span test. The analyses yielded a minimal role of complex working memory in L2 reading skills, while learners' executive control (attention and inhibition; Flankers task) was not found to correlate with L2 reading skills significantly. Taken together, these studies suggest that young learners might gradually increase their reliance on working memory capacity or, alternatively, the measure administered in these investigations was unable to detect enough variability among the participants.

Concerning the role of PSTM in primary school students' L2 learning, the existing evidence seems to confirm that PSTM predicts L2 learning at early stages (Wen & Jackson, 2022; Wright, 2015). By way of illustration, French and O'Brien (2008) studied the influence of PSTM, measured through a non-word repetition test, on L2-English grammar learning in sixth graders from Canada (L1 French, beginner level of English, 11-12 years old). Grammar was assessed at pretest and posttest in order to measure learners' gains from an intensive English program (5-month long). The results showed that PSTM explained 27.9% of the variance in grammar gains, while vocabulary knowledge only accounted for 9.5% of the scores. In sum, this study indicated that PSTM was a significant predictor of English grammar learning. In a study with beginner primary school learners of French in England

(M=9.98 years old), Porter (2017) studied the effects of PSTM on the development of L2 proficiency and literacy skills as a result of the implementation of a principled L2 program, which included explicit phonics instruction. Overall, the results indicated that PSTM, measured through a non-word repetition test, played a significant role in the L2 learning process of the primary school students. Nonetheless, its influence seemed to change as a function of the language aspect under study. First of all, the analyses revealed that PSTM had a weak influence on reading comprehension, which was associated to the complexity of the reading test and the potential effects of complex working memory. Likewise, the relationship between PSTM and the reading aloud test was found to be moderate, while learners' scores at the elicited imitation test were found to have strong relationship with PSTM, especially at delayed posttest. Still, the latter result was expected due to the characteristics and demands associated to the elicited imitation test. As regards receptive vocabulary knowledge, the strength of its relationship with PSTM was shown to decrease over time. As Porter (2017) suggested, the presentation of aural and written-word forms throughout the treatment may have facilitated the vocabulary learning process. In other words, the use of bimodal verbal input might compensate for learners' lower PSTM capacity (Porter, 2017). This finding seems to be in agreement with the results reported by Mitchell and Rule (2022) on a study on vocabulary learning with primary school learners of French (third graders). The analyses indicated that both, PSTM (measured by a non-word repetition test) and L1 literacy skills were significant predictors of vocabulary learning, however the contribution of PSTM was only shown to be mediated by L1 literacy skills (non-significant independent influence).

1.3.5.1.2 Working memory and audiovisual input

In classroom settings, learners' attitudes and behavior during their L2 learning process has been found to be affected by their working memory capacity, encouraging researchers to devote their attention to the multiple strategies that teachers may employ to support disadvantaged students (Beal et al., 2019; Gregersen & MacIntyre, 2014). Learners with low working memory capacity may be found to struggle with task completion, which is associated to task complexity, the amount of information learners have to process, and the requirement of following multiple instructions. Therefore, in absence of appropriate support, these learners may simply give up and get off task (Beal et al., 2019). In view of the consequences, Gregersen and MacIntyre (2014) propose different strategies that may lower the cognitive demands in the L2 classroom. First of all, teachers should not only reduce the amount of information that has to be processed and remembered, but also implement materials that are meaningful and familiar to the students. Equally important, the use of multimodal input may enhance comprehension and prevent working memory overload (Gregersen & MacIntyre, 2014, p.73), which is not only congruent with the dual coding theory (Paivio, 1986), and the cognitive theory of multimedia learning (Mayer, 2022), but also the subtitle principle (Mayer et al., 2020) and the investigations that have shown that the use of captions moderates the effects of complex working memory on L2 learning through audiovisual input (e.g. Pattemore & Muñoz, 2020).

In an investigation with university EFL learners, Pattemore and Muñoz (2020) assessed the effects of complex working memory capacity, measured through a reading span test, on the learning of target grammatical constructions under two experimental conditions: captions vs. no captions. The results indicated that complex working memory capacity only mediated L2 learning in the case of the participants that watched the episodes without text support, suggesting that the absence of captions made the viewing and learning process more effortful. In a study on L2 vocabulary learning (meaning recognition and word recognition) through different TV genres, Suárez et al. (2021) studied the influence of different individual differences on the outcomes: vocabulary knowledge, working memory, attention control and inhibition. As in Pattemore and Muñoz (2020), Suárez et al. (2021) administered a reading span test to measure learners' complex working memory capacity. The results indicated that vocabulary knowledge was the strongest predictor of word learning, while the cognitive factors assessed for the purpose of this study played a minimal role in the outcomes, which was attributed to learners' familiarity with viewing. According to Suárez et al., (2021), one of the possible explanations for the non-significant effects of complex working memory on the results may be the association of this factor to explicit language learning conditions, whereas the participants in this investigation were subjected to incidental learning conditions. Alternatively, as in Pattermore and Muñoz (2020), the presence of captions might have decreased the cognitive demands of the viewing task.

The study conducted by Montero Perez (2020) with intermediate learners of French assessed the influence of vocabulary knowledge and working memory capacity (PSTM and complex working memory) on incidental word learning through audiovisual input (no captions). Complex working memory was assessed through an operation span task and a backward digit span test, while PSTM was measured by means of a forward digit span test. The target vocabulary items consisted of a set of pseudowords that were tested in terms of form recognition, as well as meaning recognition and recall. The results revealed significant gains in form and meaning recognition, which were only predicted by learners' vocabulary knowledge and complex working memory capacity at immediate posttest. PSTM was not shown to be a significant predictor of word learning, which was attributed to learners' proficiency level and knowledge of the phonotactics of the language (Montero Perez, 2020). As mentioned earlier, prior research suggests that, at higher proficiency levels, learners rely on the knowledge representations stored in long-term memory rather than on PSTM (Masoura & Gathercole, 2005). As for the significant influence of complex working memory, this was associated to the complexity of the task, given that apart from the absence of onscreen text support, learners needed to allocate enough cognitive resources on the contextual clues to figure out the meaning of the target words.

Similar results were obtained by Teng and Zhang (2021) with university EFL learners of English in China under three intentional vocabulary learning conditions: 1. Definition only, 2. Definition + extra information about the word, 3. Definition + extra information about the word + video. As for the measures, the researchers adapted the vocabulary knowledge scale (VKS) to test word learning (Paribakht & Wesche, 1997), and administered a reading span test and non-word reading test to assess complex working memory and PSTM, respectively. The results indicated that the use of a video (presence of imagery) enhanced vocabulary learning, which is line with the dual coding theory (Paivio, 1986) and the cognitive theory of multimedia learning (Mayer, 2022). The findings of this investigation differed from Montero Perez's (2020) results in that both, complex working memory and PSTM significantly influenced the participants' vocabulary learning gains in receptive and productive vocabulary knowledge at posttest and delayed posttest, regardless of the treatment. Yet, this investigation did not report learners' L2 proficiency level, nor the extent to which working memory predicted word learning in each experimental condition to further explain the outcomes. Even so, the results indicate that learners' effort to commit the information to memory may have involved the functioning of both, the verbal storage component, as well as the attentional control system. Considering the scant evidence collected to date through the use of multimodal input, it seems that the use of bimodal verbal input (audio and text) reduces the cognitive demands under incidental L2 learning conditions (meaning-focused tasks). However, to develop a full picture of the role of working memory (complex working memory and PSTM) in L2 learning through audiovisual input, further research is strongly required.

1.3.5.2 Visual processing speed

To date, little attention has been devoted to the role of visual processing speed in L2 learning from multimodal input. On the whole, processing speed "…refers to how quickly the brain processes information and how efficiently simple cognitive tasks are executed over a sustained period of time." (Dehn, 2022, p. 226), which is exemplified by the analogy of a clerk that has to complete their work as fast and accurately as possible (Beal et al., 2019). Processing speed has been found to influence the functioning of working memory by making the temporal storage of information more or less efficient, which affects the completion of cognitive tasks (Dehn, 2022, p. 227). In practical terms, the students with low processing speed have been found to take longer to complete assignments, such as reading and problem solving, particularly the ones performed under time pressure (Beal et al., 2019). Although low processing speed does not prevent students from achieving their goals, it implies the allocation of additional time and practice (Beal et al., 2019), which sounds problematic in the case of captioned-video viewing considering the time constraints in the processing of captions and images, and the insufficient amount of practice reported by young primary school learners (Marzá & Torralba, 2015).

In children, processing speed may be assessed by means of the specific subtests of the WISC battery (see Dehn, 2022; Wechsler, 2003, 2014) which essentially focus on the measurement of *visual* processing speed due to the visual stimuli used for this purpose. Along with processing speed, the coding subtest employed in this investigation is also considered to be a measure of short-term visual memory, motor/graphomotor processing speed, visual scanning ability, visual discrimination, multitasking and directing sustained attention to task (Flanagan & Alfonso, 2017; Weiss et al., 2019). Based on the scope of this test, it may thus be hypothesized that visual processing speed may play a significant role in the processing and learning from multimodal input due to the importance of learners' ability to integrate verbal and non-verbal input accurately and efficiently (Pellicer-Sánchez, 2022; Sadoski & Paivio, 2013). Indeed, in L1 contexts, the study of young learners' eye movements while processing a multimodal science text indicated that it is learners' capacity to integrate verbal and non-verbal input the one that fosters greater learning and retention (Mason et al., 2013).

1.4 Treatment and word-related factors in audiovisual research

1.4.1 Use of activities

Viewing is generally approached as a leisure activity rather than learning-oriented (Vanderplank, 2015), this is why in classroom contexts, many investigations (e.g. Fievez et al., 2020) have implemented after-viewing activities as a tool to encourage learners to take the viewing experience more seriously (Rodgers & Webb, 2011; Vanderplank, 2016, Webb, 2015). Hence, when the aim of these activities is restricted to get learners' attention, their items simply test viewing comprehension, and learners are not explicitly instructed to focus on unknown items (i.e. meaning-focused activities, henceforth). By contrast, some investigations (e.g. Montero Perez et al., 2015; 2018) have explicitly encouraged learners to commit unknown target language constructions to memory by using pre-teaching activities (e.g. Gesa, 2019; Pujadas & Muñoz, 2019) or anticipating the administration of testing instruments (e.g. Montero Perez et al., 2015) as a way to maximize learning (i.e. construction-focused activities, hereafter). Meaning-focused and construction-focused activities are directly associated to the distinction between incidental and intentional learning conditions, respectively (Hulstijn, 2001, 2013). Yet, these two concepts only consider methodological procedures given that researchers might not rule out learners' self-motivated intention to commit target language constructions to memory (e.g. single words and collocations) (Uchihara et al., 2020).

Perhaps one of the main disadvantages of incidental learning conditions, namely by-product of meaning-focused activities (e.g. reading, listening or viewing) is the slow rate at which learning takes place (Hulstijn, 2003, 2013; Webb, 2020). Although short interventions (e.g. a single viewing session) may lead to statistically significant gains, these may still be relatively low (Webb, 2020). This is not surprising given that in incidental conditions, learners may devote their attention to input comprehension rather than the noticing of unknown target language constructions (Hulstijn, 2013). In foreign language settings, the evidence suggests that younger learners may not be able to pick up words at the same rate as older learners, which is why deliberate learning seems to be particularly beneficial for this age group (Kim & Webb, 2022a). By the same token, grammar learning requires either explicit or extensive treatments to ensure learning (e.g. Pattemore & Muñoz, 2020), seeing that the implementation of relatively short incidental interventions has been found to trigger little progress, especially in the case of primary school learners (d'Ydewalle & Van de Poel, 1999; Llanes & Tragant, 2021; Van Lommel et al., 2006). Yet, it is worth mentioning that in practical terms, the literature does not discourage incidental nor intentional learning activities, in turn, they should be complemented to increase learners' exposure to the target language and reinforce learning (Hulstijn, 2013; Llanes & Tragant, 2021; Webb, 2020).

Concerning the exact contribution of test announcement, Montero Perez et al. (2015) measured the effects of this enhancement technique on vocabulary learning (form recognition and meaning recall) through captioned videos in L2 French (either in full captions or keyword captions conditions). In addition, by exploring learners' eye movements, they assessed the extent to which test announcement influenced the allocation of attentional resources on unknown vocabulary words. The statistical analyses revealed that test announcement was a significant predictor of vocabulary learning regardless of the captioning type (full captions or keyword captions). Yet, the intentional learning condition was only conducive to significantly higher gains in meaning recall, which was attributed to their depth of processing, and the awareness required to learn this more demanding word dimension. Likewise, the eye tracking data indicated that test announcement increased learners' attention on the target word area on the second pass time, confirming learners' intention to commit words to memory. Besides, in the full captions condition, the amount of time spent on the area of interest was found to be associated to learners' outcomes in word recognition. Nonetheless, these outcomes were not replicated in a subsequent study conducted by the same researchers (Montero Perez et al., 2018), which was associated to the announcement of a comprehension task in both experimental conditions (intentional and incidental). Based on learners' answers to a questionnaire, their attention was primarily devoted to viewing comprehension, and even when they reported certain levels of attention on unknown vocabulary words, they did so regardless of the announcement of an upcoming vocabulary test. Apart from the moderating effects of the upcoming comprehension task, other studies have also suggested that the simultaneous allocation of cognitive resources on comprehension and word learning may result in cognitive overload, therefore learners may need to prioritize one of the tasks (Pujadas & Muñoz, 2020).

The lack of differences between the intentional and incidental learning conditions in Montero Perez et al.'s (2018) investigation may also be associated to the concept of transfer-appropriateness (Brandsford et al., 1979; Lightbown, 2008) given that the instructions provided by teachers or researchers seem to be of paramount importance to direct learners' attention to specific language features (Hulstijn, 2013). That is, if learners are explicitly instructed to focus on meaning, they are likely to score higher at this word dimension, rather than recalling other language features. In addition, the mere instruction of an upcoming vocabulary test may lack the levels of effectiveness found in the pre-viewing and after-viewing activities to promote the learning of specific target language constructions (e.g. Gesa, 2019; Pujadas & Muñoz, 2019; Teng, 2022; Van Lommel et al., 2006). In fact, among the diversity of activities that may be implemented in deliberate learning conditions, research has demonstrated that they are not equally effective (Webb et al., 2020), therefore different frameworks have been developed to carefully analyze their features and predict the outcomes (Laufer & Hulstijn, 2001; Nakata & Webb, 2016; Nation & Webb, 2011; Webb & Nation, 2017).

1.4.2 Narrow viewing

When watching movies and TV series, the literature suggests that viewers need to be familiar with the most frequent 3,000 words in English (95%) in order to reach appropriate levels of comprehension (Webb & Rodgers, 2009a, 2009b). However, based on the facilitating effects of imagery, more recent studies have lowered that threshold by indicating that viewers only need to know 80% of the words in a video to succeed in comprehension (Durbahn et al., 2020, 2022). Nonetheless, vocabulary coverage may still be problematic for low proficiency learners given that, apart from their knowledge gaps, they need to process the input under time pressure (Gesa, 2019; Muñoz, 2017a). In view of this challenge, Rodgers and Webb (2011) studied the potential effects of *narrow viewing* to enhance comprehension and learning. This concept emerged from the literature on narrow reading and listening that has shown the benefits of processing topic-related texts (aural or written).

The evidence suggests that *narrow reading* and *narrow listening* reduce the lexical load and enrich learners' background knowledge on the content, resulting in higher levels of comprehension (Cho & Krashen, 1994; Hwang & Nation, 1989; Krashen, 1996, 2004; Schmitt & Carter, 2000). In a study on reading, Schmitt and Carter (2000) indicated that this facilitating effect was also evident for the students, who reported to be aware of the value of narrow reading. By assessing the vocabulary of related and unrelated TV programs, Rodgers and Webb (2011) lent support to the outcomes obtained in reading and listening studies. Specifically, the analyses revealed that the processing of related TV programs reduced the number of lexical families and word types, which is a factor that may facilitate comprehension and learning (Lin & Siyanova-Chanturia, 2015; Webb & Nation, 2017). In addition, the possibility of encountering the target constructions (e.g. words) in multiple episodes increases the odds of learning, on the grounds that, overall, repetition has been shown to play a key role in incidental learning conditions (Madlener, 2015; Peters & Muñoz, 2020; Rodgers & Webb, 2011; Uchihara et al., 2019). Yet, frequency effects are further examined in section 1.4.2 since research has also indicated that this factor may be moderated by multiple variables, such as the presence of imagery, the language of onscreen text and the language aspect to be picked up (Muñoz et al., 2021; Uchihara et al., 2019).

Considering all of this evidence, it seems that watching episodes of a same TV program or a common topic may facilitate the viewing process. Through narrow viewing, learners may not only benefit from the repetition of different target language constructions but also enrich their knowledge about a TV program (e.g. characters and context) to strengthen their top-down comprehension processes (Rodgers & Webb, 2011). This may be particularly helpful for young and low proficiency learners, whose knowledge gaps may prevent them from attaining appropriate levels of comprehension and learning from the input (Rodgers & Webb, 2011). Yet, it is important to bear in mind that narrow viewing may only work as a facilitating tool. Even when some TV programs may be interesting or enjoyable in viewers' native language, their speed and complexity in the L2 do not necessarily make them suitable alternatives for low proficiency L2 learners. Therefore, these audiovisual resources may be adapted or watched with L1 subtitles to increase comprehension and learners' viewing self-efficacy (Muñoz, 2022). Better yet, the selection of audiovisual materials may be supported by the increasing number of studies that have carefully tested the suitability of multiple resources through diverse methodologies, such as corpus-based and eye-tracking studies (e.g. Muñoz, 2017a; Scheffler et al., 2020; Tragant & Pellicer-Sánchez, 2019).

1.4.3 Lag effects

The field of cognitive psychology has shown evidence of the influence of different learning schedules on information retention (e.g. facts, words, sentences, pictures) (Carpenter, 2017; Cepeda et al., 2006). Namely, research on *distributed practice effect* has intended to identify the optimal conditions that maximize learning through the creation of stronger memory traces that prevent quick knowledge decay (Rogers, 2021). These findings have motivated SLA researchers to test the effects of time distribution on L2 learning (immediate learning and/or later retention) by measuring their influence on diverse language aspects (e.g. vocabulary and grammar) and skills (e.g. listening and reading). Investigations on *the spacing effect* have explored the differential effects of *massed* and *spaced* conditions. Massed learning has to do with the consecutive repetition of a target stimulus without interruptions (i.e. no intervening items in the middle) or the learning of target language aspects/skills that are concentrated/practiced in a single session. Conversely, the spaced condition refers to the learning of language aspects/skills, whose repetitions/practice sessions are separated by time lags of varying lengths (e.g. minutes, hours, days and weeks). Overall, the literature suggests that distributed learning (input spacing) leads to better outcomes than massed learning (Carpenter, 2017; Kim & Webb, 2022b; Küpper-Tetzel et al., 2014; Pattemore & Muñoz, 2022b; Rogers, 2017, 2021; Ullman & Lovelett, 2018).

Concerning distributed learning, its further examination seems to be key to improve curriculum design and maximize L2 learning (Lightbown & Spada, 2020; Rogers, 2021). Research to date has examined the differences between different amounts of spacing for two main purposes: the identification of optimal instructional schedules and the retention of specific target language constructions (e.g. vocabulary and grammar constructions) (Rogers, 2021). With respect to the former, the existing studies have compared the outcomes of intensive and extensive courses by focusing on the development of different L2 skills and language aspects. Perhaps, one of the main drawbacks of early studies on curriculum design has to do with the differences between conditions in terms of teaching methodology and the number of hours of instruction (Collins et al., 1999), as well as the lack of delayed posttests (e.g. Serrano & Muñoz, 2007), which seem to be required to observe the actual long-term distributed practice effects (Cepeda et al., 2006, Rogers, 2021).

In the Canadian context, the empirical evidence has shown that, overall, intensive L2 programs lead to better outcomes in the development of L2 skills (e.g. listening, reading comprehension and oral production) in comparison with regular drip-feed instruction (Lightbown & Spada, 2020). In the study by Collins and White (2011) that compared the performance of sixth graders in intensive and extensive conditions (400 hours), the statistical analyses indicated that both groups improved significantly over time. Even when the group instructed under the intensive condition was found to score higher in vocabulary recognition, listening comprehension and written production (text length and the use of verb inflections), the effect sizes and significance levels were shown to be low. However, it is important to acknowledge that, in this study, the distribution of the instructional time in the extensive condition alternated blocks of full days in L2 English and full days in L1 French, which is a factor that might have affected the comparisons between groups given that the treatment might not be considered to be fully extensive. As the literature suggests, some time concentrations may be more effective than others (Lightbown & Spada, 2020), which is why the comparisons between programs and studies should be done with caution. Furthermore, the superiority of intensive programs seems to be less robust when the comparison groups do not differ significantly in relation to the number of hours of instruction (Serrano, 2011). Still, it is interesting to note that the subsequent analyses reported by Collins and White (2012) indicated that within the intensive group, the less proficient participants seemed to benefit from the concentration of the instructional hours. Specifically, in comparison with the more proficient students, they were capable of showing a higher performance as regards the length of their written narratives, as well as a similar performance in some of the measures: dictation, vocabulary recognition and the use of inflections in written narratives. Yet, this was not the case at the listening tests, where the less proficient participants were found to score lower. As Collins and White (2012) explained, the concentration of L2 instruction may have facilitated learning by moderating (to a certain extent) the influence of learners' individual differences. Nonetheless, Collins and White (2012) only reported the analyses obtained by the group subjected to the intensive treatment, which is a factor that constrain the possible interpretations.

In Spain, the investigations conducted by Serrano and Muñoz (2007), and Serrano (2011) with university EFL learners showed evidence of the advantages of intensive L2 instruction (same number of hours in all the conditions), lending support to the findings obtained in the Canadian context (Lightbown & Spada, 2020). To start with, in Serrano and Muñoz's (2007) investigation, the comparisons between extensive, intensive and semi-intensive programs revealed that, on the whole, the extensive treatment led to lower L2 gains at posttest. In fact, only the participants in the intensive and semi-intensive conditions improved significantly in all the measures administered over time (listening, grammar, vocabulary and reading), whereas the participants that attended the extensive program only obtained significant benefits in vocabulary learning. As for Serrano's (2011) study, it compared intermediate and advanced learners' outcomes under an extensive or intensive program. Overall, the results echoed Serrano and Muñoz's (2007) findings in that the intensive program appeared to be more advantageous than the extensive one. Yet, the concentration of the instructional time was found to be beneficial for the intermediate proficiency group but not for the advanced group, given that time distribution was not shown to influence the latter group's results. The analyses indicated that, in comparison with the intermediate extensive group, the intermediate intensive group scored higher as regards listening skills, grammar, vocabulary knowledge, and lexical complexity (in written and oral production).

Altogether, the studies to date suggest that the L2 gains obtained through intensive language programs may be either higher or comparable to extensive courses, but not inferior. Yet, these studies have mainly focused their attention on the development of L2 skills (generalization of learning) but not the acquisition of specific target language constructions, as in the more experimental studies on *lag effects* (Rogers, 2021), where the same sets of target items are instructed in multiple sessions (two or more) to test immediate gains and/or retention. More specifically, these experimental studies examine the relationship between different learning schedules (ISI=intersession intervals) and L2 gains' durability over time (RI=retention interval) (see Figure 5). The evidence that has emerged from the field of cognitive psychology has demonstrated that the increment of the RI is highly dependent on the ISI increase (Cepeda et al., 2008, Suzuki, 2021), which is a finding that has also been confirmed in SLA (Kim & Webb, 2022b). Based on the learning of different contents, Rohrer and Pashler (2007) suggested that the optimal spacing between sessions (ISI) should be approximately between 10% and 30% of the RI; while Cepeda et al.'s (2008) investigation on the learning of trivial facts obtained a ratio between 5-10% for long RIs (a year), and 20-40% for shorter lags (7-35 days). Nevertheless, these ratios have not consistently matched the results obtained in SLA studies (Serfaty and Serrano, 2022a; Suzuki, 2021), therefore these numbers may just work as a reference. Indeed, the literature suggests that we may be far from finding the optimal spacing if the actual complexity of the learning process is not considered (Suzuki, 2021) (see Figure 6).

Figure 5.

Intersession interval (ISI) and retention interval (RI)

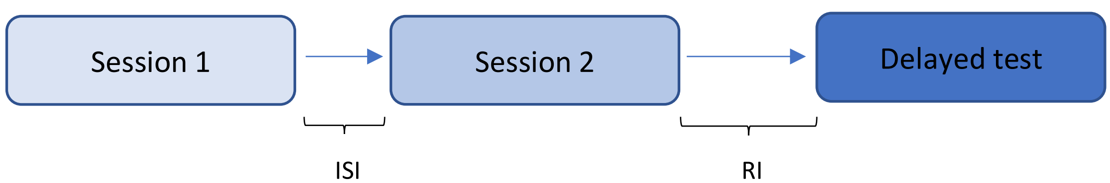
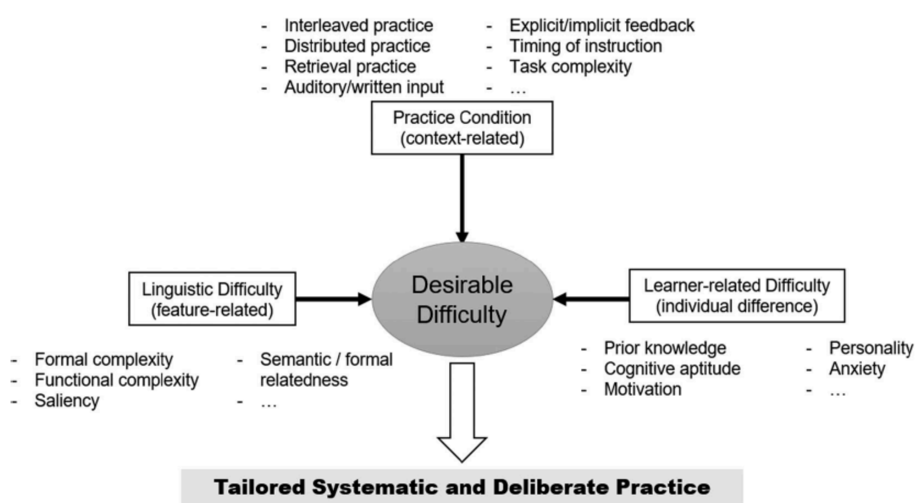


Figure 6.

Suzuki et al.'s proposed theoretical framework for systematic and deliberate L2 practice



Note. (Suzuki et al., 2019, p.715).

It is widely known that L2 learning is influenced by the interplay of multiple factors, such as practice schedules, the intrinsic difficulty of the target language aspects/skills, and learners' prior knowledge, which may potentially explain the mixed findings obtained to date as regards lag effects (Serfaty & Serrano, 2022a). Suzuki et al. (2019) built on the cognitive difficulty framework (Housen & Simoens, 2016) in second language acquisition, and the desirable difficulty framework (Schmidt & Bjork, 1992) from the field of cognitive psychology to elaborate on the multiple factors that should be considered in order to create the optimal conditions required to foster L2 learning and retention through deliberate and systematic L2 practice. Specifically, when some specific information is harder to retrieve/process (desirable difficulty), the resulting effort and depth of processing lead to higher retention (Kasprowicz et al., 2019; Ullman & Lovelett, 2018). By drawing on Housen and Simoens' (2016) cognitive difficulty framework, Suzuki et al. (2019) enlist practice condition (context-related), linguistic difficulty (feature-related) and learner-related difficulty (individual differences) as the three main areas that predict task difficulty and the extent to which a specific learning experience may "…develop knowledge and skills that are durable in the long term and transferrable to a new context" (Suzuki et al., 2019, p.713) (see Figure 6). With regard to lag effects, which is categorized as a context-related factor, Suzuki et al. (2019) posit that shorter lags (e.g. 1-day ISI) may be suitable for the development of complex language skills or the learning of more difficult language aspects, whereas longer lags (e.g. 7-day ISI) may lead to better outcomes in the case of simpler target language skills or aspects.

Serfaty and Serrano's (2022a) study is a good example of the applicability of the desirable difficulty framework on grammar learning since the researchers explored the relationship between lag effects and a set of factors that increase the difficulty of the learning task (see Table 5). Their investigation with high proficiency learners of English from Cambodia (10-18 years old) showed no significant effects of ISI but a significant albeit weak interaction between ISI and RI. In line with the literature, ISI-1 scored higher at RI-7, while ISI-7 outperformed ISI-1 at RI-28 (Rohrer and Pashler, 2007). As the researchers suggested, these outcomes were better explained in light of further analyses, given that they indicated that the ISI-1–RI-7 relationship was more suitable for slower and lower proficiency students, while ISI-7 seemed to be more appropriate for the more proficient participants who may have been able to cope with the desirable difficulty imposed by longer lags in order to promote durable learning gains. In this study, the statistical analyses did not yield significant effects for the interaction between ISI and linguistic difficulty nor age. In other words, the participants were not differently affected by the ISIs, depending on their age or the level of difficulty of the target language constructions. Nonetheless, the results revealed significant interactions between ISI and L2 proficiency, as well as ISI and time on task. As explained earlier, ISI-7 added desirable difficulty for high proficiency and faster learners, while ISI-1 facilitated the learning process for lower-proficiency and slower learners. In another experiment with the same participants, Serfaty and Serrano (2022b) studied the influence of lag effects on intentional vocabulary learning by following the same procedures. In comparison with grammar learning, ISI-7 was found to lead to higher vocabulary gains. Specifically, the greatest difference between ISI-1 and ISI-7 was detected at RI-28 in receptive vocabulary knowledge. As Serfaty and Serrano (2022b) explained, the differential effects of ISI-7 were only evident in the case of vocabulary learning due to the lower complexity of the task. In other words, the longer gaps between practice sessions were required to increase the desirable difficulty needed to foster greater vocabulary retention.

Table 5.

Factors studied in Serfaty and Serrano's (2022a) investigation.

Practice condition	Linguistic difficulty	Learner-related difficulty
• 1-day ISI • 7-day ISI • 7-day RI • 28-day RI	Two grammatical constructions: • Present perfect progressive. • Past perfect conditional in interrogative form.	• Age (children: 10-12 years old; adolescents: 13-18 years old). • L2 proficiency (Low: A1, A2; Medium: B1; High: B2-C1). • Amount of time required to complete the tasks.

Kasprowicz et al. (2019) is perhaps the only study to date that has explored lag effects on the learning of L2 grammar with primary school L2 learners. As they suggest, the overall number of studies on lag effects with young learners is still limited, thus more evidence is highly required. Kasprowicz et al. (2019) studied the learning of verb inflections through digital games with young L1-English learners of L2-French (8-11 years old). To this aim, they examined the differential effects of longer (7-day ISI; 3 sixty-minute training sessions) and shorter (3.5-day ISI; six 30-minute training sessions) time lags on the outcomes. The participants under the long-spaced condition were tested at RI-7 and RI-28, while the ones that completed the treatment with shorter gaps between sessions were tested at RI-3 and RI-28. The results only revealed a minimal advantage for the short-spaced condition, however, the differences between groups were not found to be significant over time. Indeed, the researchers suggested that the ISI-3.5 advantage was associated to their lower pretest score (i.e. greater room for learning) and not necessarily to the treatment conditions. In addition, the overall gains were found to be low for both groups despite their levels of accuracy during the training (over 75%). In a review of the results obtained in this study, Suzuki et al. (2019) hypothesized that the treatment employed in this investigation did not reach the 'desirable' levels of difficulty to promote higher gains and retention. Added to that, the statistical analyses indicated that students' level of accuracy along the treatment and language analytic ability predicted posttest performance significantly, which may explain the higher variability detected among the participants. Hence, an alternative explanation to the comparability between experimental conditions may be attributed to the influence of learners' individual differences, which may have overridden/moderated the lag effects (Kasprowicz et al., 2019).

With respect to the investigations that have tested the influence of lag effects on L2 vocabulary learning with school students (see Table 6), their outcomes seem to point to either a small advantage of shorter lags (ISI-1) or comparable gains between conditions. Yet, in the cases where longer lags are shown to enhance vocabulary retention, the results are still comparable to the scores obtained in shorter-lag conditions (Küpper-Tetzel et al., 2014; Serrano & Huang, 2018). It is important to note that Serrano and Huang (2018) is the only study that measured vocabulary gains in incidental condition. However, given that the students in this investigation had access to a glossary with the target words, the comprehension task may have been easier to complete in comparison with the studies where learners have to use their own strategies to figure out the meaning of unknown words. Therefore, in light of the optimal desirable difficulty proposed by Suzuki et al. (2019), the longer-lags between sessions may have increased the level of difficulty required to foster retention over time (see Table 6). In contrast, Serrano and Huang's (2021) subsequent investigation explicitly promoted intentional vocabulary learning from reading, where learners' aim to commit the target words to memory might have increased the complexity of the task. Therefore, considering the intrinsic difficulty of the task, the shorter lags between episodes may have eased the learning process, which is why ISI-1 led to higher gains than ISI-7 at delayed posttest (see Table 6).

Table 6.

Investigations on the influence of lag effects on L2 vocabulary learning with school students

	Target	Participants	Vocabulary dimension tested	Treatment	Results
Küpper-Tetzel et al. (2014)	Intentional learning of L1 German-L2 English word pairs.	6th graders (11-13 years old) from Germany.	Written-word form recall (cued-recall test – translation).	2 sessions. ISI-0 vs. ISI-1 vs. ISI-10. A specific RI was assigned to half of the participants in each ISI condition: RI-7 RI-35	Delayed posttests: RI-7: ISI-1 > ISI-0 and ISI-10 RI-35: ISI-1 and ISI-10 > ISI-0 The results obtained by ISI-1 and ISI-10 were comparable.
Serrano & Huang (2018)	Incidental vocabulary learning from assisted repeated reading.	Secondary school learners (14-15 years old) from Taiwan.	Receptive form-meaning mapping.	5 sessions, ISI-1vs. ISI-7 ISI-1 – RI-4 ISI-7 – RI-28	Immediate posttest: ISI-1 > ISI-7 Delayed posttest: non-significant differences between conditions. But ISI-7 led to higher retention.
Serrano & Huang (2021)	Intentional vocabulary learning from assisted repeated reading.	Secondary school learners (14-15 years old) from Taiwan.	Receptive form-meaning mapping.	5 sessions, ISI-1vs. ISI-7 ISI-1 – RI-4 ISI-7 – RI-28	Immediate posttest: ISI-1 > ISI-7 Delayed posttest: ISI-1 > ISI-7
Rogers & Cheung (2020a)	Intentional learning of 20 adjectives (descriptions).	Primary school learners (8-9 years old) from Hong-Kong.	Receptive form-meaning mapping (matching word and picture).	Target words learned in 2 sessions. within participants' comparisons. ISI-1 vs. ISI-8 RI-28	Delayed posttest: ISI-1 > ISI-8
Rogers & Cheung (2020b)	Intentional learning of 20 words.	Primary school learners (8-9 years old) from Hong-Kong.	Written-word form recall (crossword).	Target words learned in 2 sessions. within participants' comparisons. ISI-1 vs. ISI-8 RI-28	Delayed posttest: non-significant differences between conditions.

The findings from the studies carried out by Rogers and Cheung (2020a, 2020b) are relevant for this investigation since the data was collected from early primary school students (year 3, 8-9 years old). Rogers and Cheung's (2020a) first investigation showed evidence of the beneficial effects of shorter lags between sessions, in spite of the fact that the delayed posttest was administered at RI-28, which, based on Rohrer and Pashler's (2007) ideal ratio (10-30%), was far from optimal for the short-spacing condition (ISI-1=3.5%). As regards Rogers and Cheung's (2020b) next investigation, the statistical analyses yielded comparable gains between both experimental conditions (ISI-1 vs. ISI-8). More precisely, the non-optimal RI for the ISI-1 condition failed to detect the longer-lag advantage found in some experiments with adult L2-learners (Bird, 2010; Kim & Webb, 2022b; Rogers, 2015). Perhaps, in this study, lag effects were overridden by the complexity of the task since the students were tested in terms of written-word form recall, which implies a higher level of difficulty (González-Fernández & Schmitt, 2020) in a very young group of participants. Alternatively, the literature suggests that longer spacing may be effective for adult L2 learners but not necessarily for young learners since they are still developing their working memory capacity (Kim & Webb, 2022b).

The experimental studies on lag effects cited above tested the learning of target language constructions that are repeatedly encountered in multiple sessions (at least two). For instance, the investigations conducted by Serrano and Huang (2018, 2021) assessed the effects of repeated reading by adapting some graded readers. This is why Greving and Richter's (2021) study with year 7 students from Germany (L1 context) tried to innovate by examining spacing effects through the reading of different expository texts that were only connected in terms of content (e.g. biology and physics texts) given that, in practice, text repetition may not be consistently implemented in the classroom (Greving & Richter, 2021). More precisely, they aimed to determine the extent to which different amounts of spacing between reading tasks influenced the reading process (e.g. depth of processing) and facilitated content learning over time. In the first experiment, they compared the reading of two texts that were read either consecutively (massed condition) or one-week apart (spaced, 7-day ISI), whereas, in the second experiment, the texts in the spaced condition were separated by a shorter ISI (15-minute ISI). In the first experiment, the massed condition was perceived as easier than the spaced condition, which was consonant with the immediate higher learning gains. Yet, at delayed posttest (7-day RI), the scores were shown to be comparable between the two conditions (i.e. 7-day ISI did not lead to higher gains). Therefore, contrary to their expectations, the spaced reading condition was only found to be advantageous as regards knowledge retention. These results were echoed in the second experiment as well.

As Greving and Richter (2021) explained, learners' greater effort to activate previous knowledge to enhance content comprehension in the second text may have triggered greater content retention in the long run, albeit not superior. In addition, in the first experiment, the scores obtained in the long-spaced interval (7-day ISI) were found to be low at immediate posttest. Therefore, in light of the desirable difficulty framework (Schmidt & Bjork, 1992), the researchers indicated that the learning from reading task may have been harder with a long lag between sessions (7-day ISI). Alternatively, Greving and Richter (2021) indicated that the administration of an immediate posttest may have enhanced the outcomes of spaced reading, resulting in non-significant differences between conditions. As for learners' perception of task difficulty, the results tie in with the quantitative analyses since the texts read under a short-spaced condition (15-minute ISI) obtained a similar difficulty rate to the massed condition, which was not the case for the long-spaced interval (7-day ISI), which was perceived as more difficult. In addition, the participants reported a lower feeling of learning and a weaker connection between texts under the 7-day ISI experimental condition. Considering these findings, further research should be undertaken to examine the differences between different time lags (below 7-day ISI). It is also important to acknowledge that these findings were collected in an L1 context, therefore, these results should not be extrapolated to the processing of L2 texts. Nevertheless, it is also true that these findings fall in line with the SLA literature that suggests that, in comparison to adults, long lags between practice sessions may not be advantageous for young school learners (Kim & Webb, 2022b).

Greving and Richter's (2021) attempt to study spacing effects on the learning from unrelated texts is relevant for the present investigation for multiple reasons. Most investigations on extensive viewing use multiple episodes of the same TV series (e.g. Pattemore and Muñoz, 2022a; Pujadas, 2019), which is an action that is labeled as *narrow viewing* (Rodgers & Webb, 2011) (see section 1.4.2). The literature has shown that watching episodes of a same TV series (narrow listening) or processing texts (oral or written) on a related topic (narrow listening/reading) result in lighter lexical load and higher levels of comprehension (Krashen, 1996) which is recommended for lower proficiency learners (Rodgers & Webb, 2011; Schmitt & Carter, 2000). Apart from the lexical load, the facilitating effects of narrow viewing have to do with viewers' cumulative gains in background knowledge about a specific TV program, which facilitates comprehension as a result of top-down processing (Rodgers & Webb, 2011). In addition, the lower cognitive demands allow the viewers to devote greater attention to unknown language constructions (Lin & Siyanova-Chanturia, 2015). Therefore, considering that the use of authentic audiovisual materials does not allow the researchers to manipulate the encountering of the target language constructions along the episodes, it is possible to test the influence of varying time lags between episodes or viewing sessions. Therefore, it remains unknown whether shorter lags between episodes may allow learners to activate previous knowledge with greater ease. That being the case, shorter lags between episodes might potentially facilitate input processing and comprehension, resulting in higher L2 gains (Lin & Siyanova-Chanturia, 2015). Equally important, making a greater effort to retrieve the background knowledge required to enhance comprehension might also trigger desirable difficulty levels to foster a better performance and/or higher levels of retention in the long run (Suzuki et al., 2019).

Concerning distributed practice effects on L2 learning from audiovisual input, very few studies have been conducted on this regard. The investigation carried out by Pattemore and Muñoz (2022b) with university EFL learners from Russia (A2-C1 CEFR level) compared the learning of multiword units from captioned-video viewing (five episodes) under three experimental conditions that differed in terms of intersession interval: ISI-7 (1 episode a week), ISI-1 (one episode a day), and ISI-0 (5 episodes in a session). Although the results showed significant improvement from pretest to posttest in all the experimental groups, the long-spaced condition (ISI-7) was found to lead to higher gains in comparison with the short-spaced (ISI-1) and the massed condition (ISI-0). Yet, the participants were only tested immediately after the treatment, therefore, it is uncertain whether this advantage was kept over time. The analyses also indicated that, as expected, the extent to which learners benefitted from the treatment was influenced by vocabulary knowledge (see section 1.3.2). Nevertheless, this factor was not shown to interact with viewing time distribution. In line with Greving and Richter's (2021) investigation, the participants in the massed condition showed a higher feeling of learning in comparison with the long-spaced condition. Still, the influence of learners' perception of their learning gains was not found to be statistically significant. Therefore, learners' perceptions did not match the quantitative results (i.e. gains).

In a study of repeated captioned-video viewing, Muñoz et al. (2022) compared university EFL students' vocabulary learning under two experimental conditions: watching the same episode twice in a single session (ISI-0) or with one-week interval (ISI-7). To measure vocabulary learning (meaning recall and recognition), the participants (B2 CEFR level) were tested at pretest, immediate posttest and delayed posttest (4.5-week RI, ISI-7=21%). In comparison with a control group that only took the tests, both experimental groups obtained greater gains at immediate posttest. Nevertheless, only the spaced viewing group (ISI-7) was found to score significantly higher than the control group at immediate posttest (meaning recognition and recall). The difference between the massed viewing and the control group only reached marginal statistical significance in meaning recognition. However, considering that the comparisons between spacing conditions were not shown to be significant, the slight immediate advantage obtained by the spaced viewing group was not found to be robust. Concerning the delayed posttest, the results did not show differences between conditions in terms of retention. Still, the researchers also suggested that the participants that watched the two episodes in a row might have searched for the meaning of the unknown words after the immediate posttest given that their scores were found to be higher at delayed posttest in terms of meaning recognition. As a consequence, the long-term spacing effects may be considered to be uncertain. On the whole, the limited evidence on distributed practice effects on L2 learning from captioned-video viewing seem to point to the immediate advantage of longer lags between episodes. However, further evidence is strongly needed to determine whether this tendency is confirmed. It is also important to note that these studies were conducted with university studies of at least A2 proficiency level. Therefore, as mentioned in the aforementioned studies on lag effects, the results may point to the opposite direction with young school learners.

1.4.4 Vocabulary learning: Word and context-related factors

Research on audiovisual input has shown great variability in learners' vocabulary gains (Montero Perez, 2022), which might be attributed to multiple variables. Previous sections have already explained how learner- and treatment-related factors may influence the outcomes. Therefore, based on the premise that words differ as regards their learning burden (Barclay, 2021), this section explicitly focuses on the role of word- and context-related factors in vocabulary learning. In particular, based on Peter's (2020) classification, we studied the influence of word properties (regularity, length, and concreteness) and the use of words in context (frequency of occurrence) (see Table 7).

Table 7.

Word- and context-related factors

Word-related factors	Context-related factor
Regularity *(more or less consistent with L1 patterns)* Word length *(shorter vs. longer words)* Concreteness *(more or less concrete)*	Frequency of occurrence *(repetitions in the input)*

Note. Based on Peters (2020).

1.4.4.1 Word regularity

Word regularity refers to the degree of consistency between phoneme-grapheme correspondences and how aural- and written-word form representations map on to each other (Hamada & Koda, 2008). However, considering the characteristics of the languages involved in this study (L1 Spanish-L2 English), in this investigation regularity will specifically refer to the extent to which the English target words follow the regular (transparent) orthographic patterns of Spanish (i.e. L2 words may be more or less consistent with L1 patterns). This definition was built on the basis that Spanish and English differ in terms of orthographic transparency given that the Spanish orthographic system mainly consists of one-to-one grapheme-phoneme correspondences, with very few exceptions (Hamada, 2021). On the whole, the literature suggests that language transparency facilitates the development of L1 literacy skills (e.g. spelling) (Papp, 2020; Sun-Alperin & Wang, 2008), whereas in the case of non-transparent orthographies, the learning process seems to be more effortful and slower, particularly as regards vowels and diphthongs (Caravolas, 2004; Papp, 2020; Sun-Alperin & Wang, 2008). This is the case of English, whose many-to-many sound-symbol correspondences make it a non-transparent/opaque language. To exemplify, in English the phoneme /i:/ may be written as -ee (green), -ea (beans) and i (she), while the /eɪ/ diphthong may be represented as -ay (play), -a_e (bake), -ai (pain), or even -eigh (eight).

Although SLA research has indicated that the degree of regularity in phoneme-grapheme correspondences may inherently affect the building of phonological and orthographic representations in the L2 (Hamada & Koda, 2008; Ijalba & Obler, 2015), the evidence has also demonstrated that the extent to which word regularity influences L2 learning is moderated by L1 characteristics and L2 proficiency (Birch, 2015; Figueredo, 2006; Ijalba & Obler, 2015; Muñoz, 2017b). Specifically, research has shown that L1 orthographic patterns influence the development of L2 literacy skills (e.g. Sun-Alperin and Wang, 2008), especially at early stages when learners' linguistic infrastructure is still assimilating and accommodating to the patterns of the L2 (Birch & Fulop, 2021; Perfetti et al., 2007). To exemplify, at low-proficiency levels, L2 learners may assimilate L2 sounds to their L1 inventory (Flege & MacKay, 2004), which is why beginner L1 Spanish learners of English may categorize the English /æ/ and /ʌ/ sounds together as the Spanish /a/. As a result, they may well spell the word /mæp/ as 'map', but misspell the word /kʌt/ (cut) as 'cat' by following the regular patterns of the Spanish orthographic system.

The study by Sun-Alperin and Wang (2008) provided further evidence of the influence of L1-Spanish orthographic patterns on EFL learning. They compared L1 Spanish and L1 English primary school students' spelling performance in English vowels to elucidate whether the L1 Spanish groups' errors were influenced by their orthographic system. The results indicated that the L1 Spanish learners made significantly more mistakes in the items that activated their L1 orthographic patterns. These findings were somehow echoed by Ijalba and Obler (2015), whose study attempted to examine the learning of pseudo words in L1-Spanish and L1-English adult learners. The analyses indicated that for Spanish speakers, it was harder to recall orthographically opaque words, whereas English speakers seemed to be influenced by the use of one-to-many decoding strategies, fostering the learning of both, opaque and transparent words. Overall, L1 speakers of transparent languages may be aware of their struggle. In the study by Muñoz (2017b) with L1-Spanish learners of English (sixth and sixth graders), learners' responses clearly indicated that language transparency was one of the main sources of difficulty in their EFL learning process. However, the results obtained in each year level also suggested that learners' struggle with phoneme-grapheme correspondence decreased with age and proficiency. Hence, additional practice and support might be required to eventually overcome this barrier.

In view of the complexity of the English orthographic system, the evidence suggests that the presentation of target vocabulary words through bimodal verbal input, that is aural and written representations, may reduce the learning burden (e.g. Barclay, 2021). Yet, based on the findings obtained by Krepel et al. (2020) with L1-Dutch learners of English (year 6), the facilitating effects of bimodal verbal input may be more evident in the case of the words that follow consistent orthographic patterns, given that they lead to stronger knowledge representations (Krepel et al., 2020). More specifically, when examining the contribution of print, Krepel et al.'s (2020) investigation revealed that, overall, the use of audio and print facilitated the learning of word spelling and reading; whereas, in a forward translation task, aural-word form recall was found to be enhanced by word regularity. In light of the aforementioned studies, this outcome may not be surprising on the grounds that learners' L1 is considered as a semi-transparent language (Krepel et al., 2020). On the whole, despite the methodological differences between Barclay (2021) and Krepel et al.'s (2020) investigations, both experiments concur on the facilitative effects of learners' exposure to audio and print. Still, Krepel et al.'s (2021) findings suggest that it is hard to override the influence of word regularity, which may remain as a strong predictor of vocabulary learning at early stages.

Considering all of this evidence, it seems that the higher learning burden of irregularly spelled words may be accounted by two main factors: learners' poor knowledge of L2 sound-to-spelling correspondences, and L1 influence. This is why, at early stages, L2 learners might need to rely on their capacity to memorize word forms to reach greater levels of accuracy (Birch, 2015). Thus, in view of FL learners' limited exposure to print to learn orthographic patterns incidentally, more explicit instruction and greater exposure may be required to make significant progress over time (Marian et al., 2021; Muñoz, 2017b; Pérez Cañado, 2006; Porter, 2020).

1.4.4.2 Word length

To date, there is a relatively small body of literature that is concerned with word length (Barclay & Pellicer-Sánchez, 2022; Peters, 2020). Among these studies, word length has been operationalized as the number of syllables (e.g. Puimège & Peters, 2019a), letters (e.g. Barclay & Pellicer-Sánchez, 2022; Ellis & Beaton, 1993a) or phonemes (e.g. Willis & Ohashi, 2012) that make up a word. Considering the characteristics of the English orthographic system and the complexity of syllable formation (e.g. caught vs. cat) (Sun-Alperin & Wang, 2008), it sounds reasonable to hypothesize that the number of letters may increase the learning burden for L1-Spanish learners of English (see section 1.4.4.1). Overall, the existing evidence has consistently indicated that longer words are more difficult to learn (e.g. Ellis & Beaton, 1993a; Barclay & Pellicer-Sánchez, 2022) since they need additional time to be processed and recognized (Grabe, 2009), and they are harder to store in the PSTM (Birch, 2015).

In a study on word learning strategies, Ellis and Beaton (1993a) found a negative relationship between word length (number of letters) and word learning. A higher number of letters was found to lead to more mistakes in word form recall, which was connected to the complexity of longer words, which are harder to encode. Likewise, Barclay and Pellicer-Sánchez (2022) studied the learning burden added by word length in intentional learning conditions by using a flashcard software. In this study, learning burden was operationalized as the number of times the participants (B2 learners of English) required to encounter the words to achieve learning (written-word form recognition and recall). At the immediate posttest, the results revealed significant effects for word length, that is longer words increased the learning burden. On the whole, these findings confirm that word length is a significant predictor of word learning when students intentionally try to commit the target items to memory. Interestingly, the investigation by Puimège and Peters (2019b) also showed significant effects for word length in incidental learning conditions; however, their results pointed to the opposite direction, suggesting that longer word forms were easier to recall. In view of input modality (non-captioned video), the researchers explained that in incidental learning conditions, longer aural word forms may be more easily noticed in the stream of speech (Puimège & Peters, 2019b; Peters, 2020).

Together, these studies indicate that word length is a significant predictor of vocabulary learning. However, the actual influence of this factor on the outcomes (shorter vs. longer words) might depend on the learning conditions (incidental vs. intentional) and input modality, since in some contexts, word length might be key to get learners' attention (Puimège & Peters, 2019b; Peters, 2020; Schmidt, 2001).

1.4.4.3 Word concreteness

Words are considered to be concrete when they are easy to imagine, while abstract words are not (Peters, 2020, p. 130). Higher concreteness, which is associated to higher imageability (Peters, 2020), has consistently been found to promote greater learning and recall in both, intentional and incidental learning conditions (E.g. De Groot & Keijzer, 2000; Ellis & Beaton, 1993b; Puimège & Peters, 2019b). With respect to incidental vocabulary learning, concrete words have been found to be more salient, which explains why they may be more easily noticed and picked up (Crossley et al., 2016).

The literature suggests that there is a confound between concreteness and part of speech (Peters, 2020). Ellis and Beaton's (1993b) study on vocabulary learning through the keyword strategy found that nouns were easier to learn than verbs, which was attributed to word imageability. By contrast, in the study conducted by Barclay and Pellicer-Sánchez (2022) on vocabulary learning through flashcards (written-word form recognition and recall), the analyses indicated that part of speech was not a significant predictor of word learning. This result was associated to the fact that the researchers had controlled for word concreteness (Barclay & Pellicer-Sánchez, 2022). Therefore, the advantage of nouns over other parts of speech found in the literature might be accounted by word concreteness rather than part of speech (Peters, 2020).

Overall, the empirical evidence has demonstrated that concreteness influences the learning of different word dimensions (Crossley et al., 2016; Nation, 2020; Peters, 2020). For instance, the investigation by Puimège and Peters (2019a) with Flemish young learners (10-12 years old) indicated that concreteness fostered the learning of receptive vocabulary knowledge at the level of meaning recognition. Likewise, in the experiment by De Groot and Keijzer (2000), word concreteness enhanced the learning of receptive and productive vocabulary knowledge, indicating that retention was also influenced by this factor. It is important to note that word concreteness has not only been found to facilitate the learning of word meanings (e.g. Puimège & Peters, 2019a; Van Zeeland & Schmitt, 2013), but also word forms (e.g. Puimège & Peters, 2019b; Van Zeeland & Schmitt, 2013). This facilitative effect may be interpreted in light of the Dual Coding theory (Paivio, 1986) (see section 1.1.1), where the strength of the referential connections between the verbal and the non-verbal systems that foster learning and recall is mediated by concreteness (Clark & Paivio, 1991).

More precisely, learners' capacity to evoke images and/or visualizations depends on their degree of concreteness (Clark & Paivio, 1991).

The results aforementioned may be closely connected to the findings obtained by studies on vocabulary learning through audiovisual input, since more concrete words tend to be graphically represented onscreen (Peters, 2020). As Mitchell and Rule (2022) point out, "…multimodality could enhance encoding and lead to deeper memory traces which are longer lasting and more easily retrieved" (p. 40). This explains why the existing evidence has consistently pointed to the significant effects of imagery on vocabulary learning (Fievez et al., 2020; Neuman & Koskinen, 1992; Peters, 2019; Pujadas, 2019; Rodgers, 2020). For instance, the investigation conducted by Peters (2019) with secondary school students from Flanders indicated that the presence of imagery tripled the odds of picking up the target words at the level of form recognition and meaning recall. However, in the study by Pujadas (2019) with secondary school students from Spain, imagery was only found to be a strong predictor at the level of meaning recall but not word-form recall. Thus, the extent to which imagery influences vocabulary learning at a level other than meaning may still need further exploration.

1.4.4.4 Frequency of encounters

Much of the literature on vocabulary learning has paid particular attention to frequency effects. On the whole, the evidence suggests that in incidental learning conditions, the number of encounters increases the likelihood of picking up words (Uchihara et al., 2019). Initial research on frequency effects concentrated its efforts on the identification of a threshold that indicated the number of repetitions that are conducive to vocabulary learning (Uchihara et al., 2019). Although, research conclusively showed that repetitions are key to foster learning (e.g. Horst et al., 1998; Waring & Takaki, 2003), the resulting thresholds turned out to be inconsistent. Roughly speaking, eight to ten repetitions may be required to learn words incidentally through reading (Peters, 2020). Yet, the number of encounters might vary depending on the word dimensions under study (Peters, 2020; van Zeeland & Schmitt, 2013), and the intrinsic difficulty of the target items, such as length, regularity and concreteness (Barclay & Pellicer-Sánchez, 2022). In addition, the conflicting results between studies may be attributed to the multiple external factors that interact with word repetitions, such as learners' age, proficiency, and input modality, which is why current research is shifting its attention to study frequency effects in interaction with other variables (Uchihara et al., 2019). Indeed, usage-based theorists concur on that frequency effects cannot be detached from other factors, since it is only one of the many variables that play a role in second language acquisition (Ellis & Wulff, 2015).

The meta-analysis performed by Uchihara et al. (2019) shed some light on this regard. To start with, the results showed that word repetitions have a medium effect on vocabulary learning (r=34). Then, the further analyses indicated that frequency effects are moderated by multiple factors (see Uchihara et al., 2019 for a full account). As for learner-related variables, word repetition plays a more important role at early proficiency stages and when the target words are completely unknown. Yet, the results also indicated that frequency effects are larger with older learners (e.g. university students) in comparison with young learners (primary and secondary school). More precisely, primary school learners may not be sensitive to frequency effects due to their developing cognitive and literacy skills, which make the processing of input more effortful (Uchihara et al., 2019; see section 1.3.1). Concerning methodological and treatment-related factors, frequency effects are more prominent when words are concentrated in massed conditions (see section 1.4.3). In addition, the effects of repetition were found to be higher when learners were forewarned about an upcoming vocabulary test, which may be associated to the explicit allocation of higher attentional resources on unknown words, increasing learners' sensitivity to frequency effects (Montero Perez et al., 2015; see section 1.4.1). With respect to input mode, frequency appears to play a role in all modalities, reading (e.g. Vidal, 2011), listening (e.g. van Zeeland & Schmitt, 2013; Vidal, 2011), reading-while-listening (Brown et al., 2008), and viewing (e.g. Fievez et al., 2020); however, the results suggested that the effect size is higher in reading (r=.41), and listening (r=.39), in comparison with reading-while-listening (r=.28) and viewing (r=.22). Therefore, as can be observed in these results, frequency effects are more robust in written mode (Peters, 2020). These differences suggest that frequency effects may be overridden by multimodality on the grounds that the processing of bimodal verbal input and/or imagery may make the words more salient in the input and enhance learning (Uchihara et al., 2019). Hence, other word-related factors may be stronger predictors of vocabulary learning in multimodal conditions.

An increasing number of studies on viewing has confirmed that, overall, frequency plays a significant role in vocabulary learning in this modality (E.g. Fievez et al., 2020; Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013; Teng, 2019a). Some of these studies have contributed to the existing literature by giving interesting insights into the relationship between frequency of occurrence and some factors associated to the viewing experience, such as number of episodes (Fievez et al., 2020) and the language of onscreen text (Muñoz et al., 2021). In the study by Fievez et al. (2020) with secondary school students, the researchers assessed the influence of word frequency on vocabulary learning from L1/L2-subtitled-viewing (word recognition, meaning recall and meaning recognition; 15 episodes). The analyses indicated that frequency of occurrence played a limited role in the results, which was associated to the fact that most of the target words appeared in multiple episodes and were not concentrated in a single one as in some previous studies (e.g. Peters & Webb, 2018). This result is consistent with that of Webb and Chang (2015), whose analyses yielded non-significant frequency effects under an extensive 13-week reading-while-listening treatment.

In the viewing study by Muñoz et al. (2021) with secondary school learners of English (elementary proficiency level), word frequency affected the outcomes differently depending on the language of onscreen text (L1 subtitles or L2 captions). Overall, frequency effects were found to be significant at the level of both, word form and meaning recall; however, the running of further analyses indicated that in the L1 subtitles condition, frequency played a significant role in meaning recall, while in the L2 captions condition, frequency effects were prominent in written-word form recall. As Muñoz et al. (2021) explained, the learners exposed to L1 subtitles had direct access to word translations, therefore a higher number of encounters with this information promoted learning at the level of meaning recall. Likewise, in L2 captions condition, the target word forms were available on the onscreen text, thus their number of repetitions was key to foster form recall. The weak frequency effects obtained in this study are in agreement with Fievez et al.'s (2020) findings, which may be attributed to the length of the treatment and participants' age (secondary school students in both studies). In the second sub-study reported by Muñoz et al. (2020) with university learners of English, frequency effects were stronger for the learning of grammar constructions (in comparison with the vocabulary study), in particular in the L2 captions condition (captions vs. non-captions). These outcomes were not only associated to the participants' age but to the need of higher repetitions under more difficult learning conditions. These results match those observed in studies on deliberate vocabulary learning, where the words that are intrinsically more difficult to learn need a higher number of trials (repetitions) (Barclay & Pellicer-Sánchez, 2022).

Together, these studies suggest that the contribution of frequency effects to vocabulary learning is mediated by multiple factors. While a higher number of repetitions may be required under more difficult learning conditions, frequency effects seem to be weaker in the processing of multimodal input. Likewise, seeing that younger learners may be less sensitive to frequency effects, they might need the explicit instruction of focusing on vocabulary to increase the influence of this factor. However, the evidence is still limited to raise hypotheses in this regard, which is why this seems to be a fruitful area for research.

II. Research design and methodology

2.1 Introduction and research questions

This investigation aimed to contribute to the existing literature by exploring the use of captioned videos with primary school learners of English as a foreign language. As summarized in Figure 7, this study examined learners' gains as regards L2 written-word form recall, written-word form and meaning recognition, and the development of receptive language skills (English reading efficacy, Spanish reading efficacy and English listening skills). In addition, it investigated the role played by a series of factors on the outcomes, which concerned treatment (viewing distribution and after-viewing activity type), word (regularity, word length, concreteness and frequency of occurrence), and learner-related characteristics (cognitive abilities, year level, L2 vocabulary knowledge, L2 listening skills, L1 and L2 segmentation, L1 and L2 reading efficacy, and L1 reading habits and attitude towards reading).

The research questions that guided this study were as follows:

1) To what extent does viewing distribution (i.e. shorter vs. longer lags) influence young L2 learners' gains from captioned video viewing?

2) In comparison with meaning-focused activities, what are the effects of construction-focused after-viewing activities on L2 learning through captioned-video viewing?

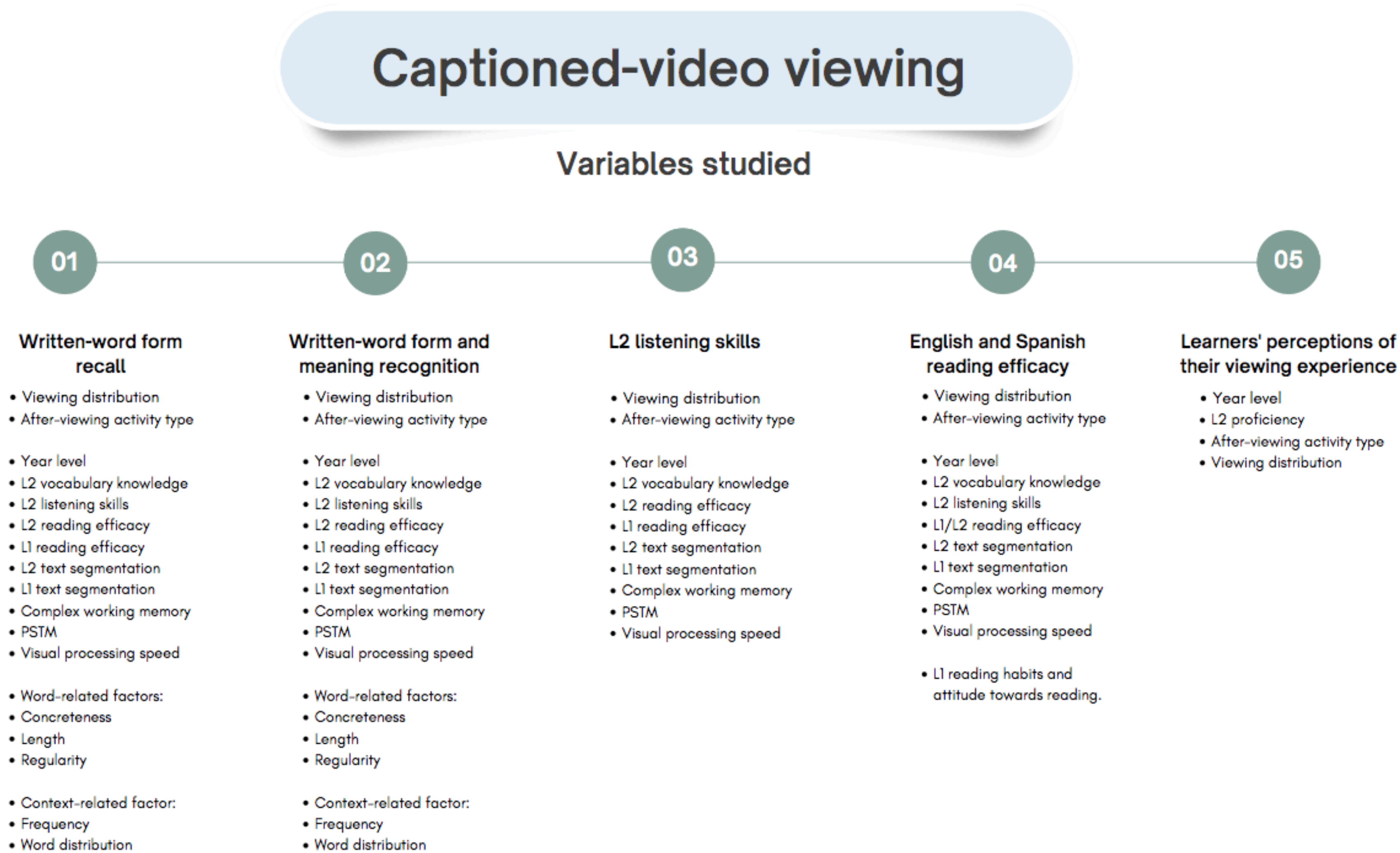
3) To what extent do learner characteristics influence young L2 learners' gains from captioned-video viewing? (i.e. age, vocabulary knowledge, cognitive abilities [phonological short-term memory, complex working memory, and visual processing speed], L1 and L2 reading skills [reading efficacy and text segmentation], L2 listening skills, and L1 reading habits and attitude towards reading).

4) To what extent do context and word-related factors (frequency of occurrence, regularity, word-length, and concreteness) influence vocabulary learning?

5) What are students' perceptions of the viewing experience? How do their answers contribute to the interpretation of the quantitative findings?

Figure 7.

Variables studied in this investigation



2.2 Participants and context

2.2.1 Experimental groups

This investigation was carried out with a convenience sample of 136 L1-Spanish primary school learners in year 4 (aged 9-10; N=71) and year 5 (aged 10-11; N=49) from two private schools in Chile (medium-high socioeconomic status). School 1 and school 2 were located in the sixth (O'higgins) and the tenth (Los Lagos) regions of the country, respectively. The key requirements to be included in the analyses were as follows: watching 100% of the episodes, doing at least a set of pre and posttests (e.g. written-word form recall) to assess their progress over time, and not being part of the group of students with special educational needs since they received additional support to complete the activities. As a result, 16 participants were excluded from the analyses. The remaining 120 participants (59 male and 61 female) belonged to six intact classes, which were randomly assigned to a treatment condition as regards viewing distribution (i.e. the number of episodes the participants watched a week; see Table 8).

Table 8.

Group characteristics

School	Year level	Viewing distribution	Group	Number of participants
1	Year 4	3 episodes a week	3-fourth	16
1	Year 4	4 episodes a week	4-fourth	15
1	Year 4	2 episodes a week	2-fourth	16
2	Year 4	1 episode a week	1-fourth	24
1	Year 5	2 episodes a week	2-fifth	24
1	Year 5	4 episodes a week	4-fifth	25

It is worth mentioning that according to a relatively recent report of the British Council (2015),

> Chile has the fourth largest proportion of children enrolled in private schools in the world, and private enrolment is three times higher than the OECD average: only 37 percent of 15-year-olds attend publicly funded, publicly run schools compared to an OECD average of 82 percent (as cited in Enever, 2018, p. 144).

Therefore, in this context, attending a private or semi-private school does not necessarily imply that learners attend bilingual/immersion programs, nor that they attain high proficiency levels, especially when schools are located outside the three metropolitan centers: Santiago, Concepción and Valparaíso (Enever, 2018). This information partly explains the characteristics described in the paragraphs below.

With respect to EFL teaching and learning, the participants from school 1 had been formally instructed since preschool (6 hours a week), whereas the students from school 2 started in second grade (5 hours a week). Nevertheless, differences concerning starting age or number of hours of instruction did not result in significant differences in terms of L2 proficiency between the two schools (Pre-A1 and A2 according to the CEFR) (section 3).

As for out-of-school contact with the target language, the questionnaire administered to the families from school 1[3] confirmed that in more than 50% of the participants, English exposure was rather limited or non-existent (see Table 9). To start with, of the 88 participants that answered the questionnaire, only 31 reported watching videos with subtitles in Spanish for an average of 11 hours a month, while 21 students reported watching captioned videos (*M*= 9.66 hours a month) and only 17 indicated watching videos in English without textual support (*M*= 8.9 hours a month). In total, 39 out of 88 participants reported watching at least one of the three types of videos in English (*M*= 18 hours a month, *Minimum*= 1, *Maximum*= 80).

Even when the difference concerning total viewing time (number of hours a month) appeared to benefit fifth graders (see Table 9), a Mann-Whitney U test indicated that the comparison did not reach statistical significance (U=216.5, z=.746, p= .461, r= .11). Therefore, the descriptive statistics just reflected the great variability among participants (see the standard deviations). Likewise, the difference between gender groups (see Table 10) was not statistically significant (U=197.000, z=.197, p= .857, r= .03). As for gaming, 29 out of 88 students reported playing either individual or multiplayer videogames (*M*= 15,37 hours a month, *SD*=11.96). In line with studies on out-of-school exposure (e.g. De Wilde et al., 2019), frequencies were found to be even lower as regards reading in English (physical or online resources) (*N*=11, *M*=4.81 hours a month, *SD*=3.57) and English instruction outside school (*N*=5, *M*=4.2 hours a month, *SD*=2.28). Conversely, listening to music in English

[3] Only some of the instruments were administered at school 2.

was found to be the most popular activity among students, where around 85% of the participants claimed doing it every month (*N*=78, *M*=14.06 hours a month, *SD*=12.72).

Table 9.

Monthly viewing time per year level

		Year 4			Year 5			Total		
		N	*M*	*(SD)*	N	*M*	(SD)	N	*M*	(SD)
Spanish subtitles	No	30	0	0	27	0	0	57	0	0
	Yes	15	8.53	(7.66)	16	14.68	(18.52)	31	11	(14.45)
Captions	No	35	0	0	32	0	0	67	0	0
	Yes	10	7.5	(4.81)	11	11.63	(16.03)	21	9.66	(11.97)
Videos without textual support	No	37	0	0	34	0	0	71	0	0
	Yes	8	6.75	(4.20)	9	10.81	(9.03)	17	8.9	(7.27)
Total viewing time	No	26	0	0	23	0	0	49	0	0
	Yes	19	13.52	(11.45)	20	23.01	(24.36)	39	18	(19.54)

Table 10.

Monthly viewing time per gender group

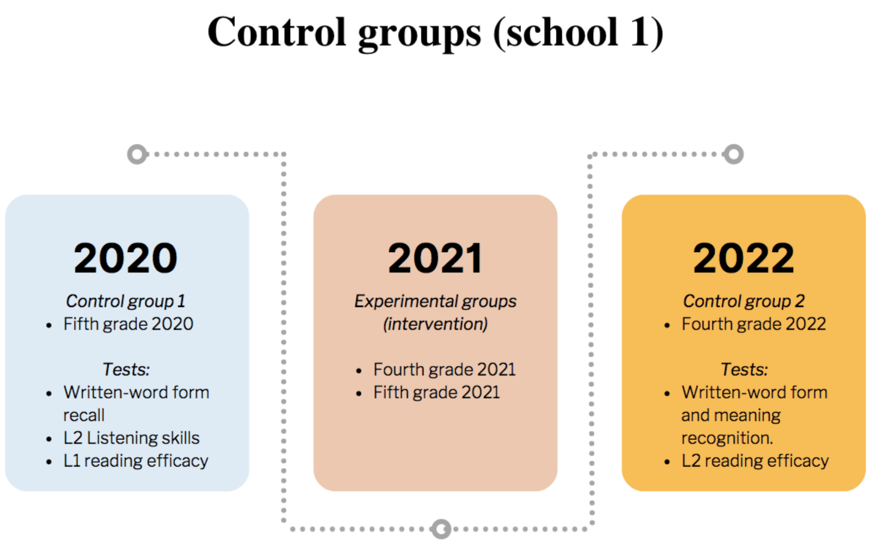
		Male			Female		
		N	*M*	*(SD)*	*N*	*M*	*(SD)*
Spanish subtitles	No	27	0	0	30	0	0
	Yes	17	13.47	(18.79)	14	9.57	(6.17)
Captions	No	33	0	0	34	0	0
	Yes	11	6.81	(5.11)	10	12.80	(16.39)
Videos without textual support	No	38	0	0	30	0	0
	Yes	6	10.66	(9.7)	11	7.93	(5.88)
Total viewing time	No	24	0	0	25	0	0
	Yes	20	18.40	(20.21)	19	18.38	(19.36)

2.2.2 Control groups

Two groups of students from school 1 were recruited as control groups in order to determine whether the language gains in the experimental groups could be attributed to the treatment, test effects or cognitive maturation. To ensure that the students in the control groups matched the experimental groups in age, the administration of the instruments took place at different points in time (see Figure 8). In addition, considering that the control groups did not go through any kind of treatment and minimal learning is expected from mere test administration, each control group was assigned a specific set of tests (see Figure 8).

Figure 8.

Control groups



2.2.2.1 Control group 1

A small group of fifth graders (*N*= 16 students, 9 male and 7 female) from school 1 was recruited as control group six months before the actual experiment. After informing the school families, these participants volunteered to participate in additional English sessions (outside their school schedule), where they were administered some of the tests (see Figure 8) and completed unrelated practical activities (e.g. vocabulary games). Five weeks separated pretest from posttest administration. At the end of the process, they were awarded the maximum score in terms of participation in their English course and also received special presents to thank them for their time and willingness to do the activities. These incentives ensured that the control group was formed by students from different proficiency and academic levels, making it comparable to the characteristics of the experimental groups.

Due to the pandemic, most of the academic year 2020 in Chile took place online, therefore, in order to facilitate the administration of the instruments in online format, the students were randomly assigned to three small groups (5-6 students each). They had to keep their cameras on during the whole session, and specific time restrictions were set to prevent them from using additional resources (e.g. online dictionaries) to complete the activities. Given that these students were tested before the experimental groups, they were additionally administered an online questionnaire[4] after the posttest to rate their experience and the instruments. Specifically, they had to explain their opinions and decide which test was the easiest and the most difficult one. This information was useful to anticipate potential flaws in the actual experiment.

2.2.2.1 Control group 2

A small group of fourth graders (*N*= 17 students, 9 male and 8 female) from school 1 participated as control group a year after the actual experiment (see Figure 8). In contrast with control group 1, the activities took place onsite by following exactly the same procedures as with the experimental groups. The students that volunteered to complete the activities (and whose parents consented their participation), where taken to the school library during class hours. In addition, the average grades obtained in the English class were taken into consideration in order to resemble the variability in L2 proficiency observed in the experimental groups.

2.3 Audiovisual materials and after-viewing activities

2.3.1 Audiovisual materials

Charlie and Lola (Carrington & Child, 2005-2008) is an animated cartoon based on Lauren Child's picture books. It was selected for the purpose of this study for being age and content appropriate for the target participants. An important characteristic of this animated cartoon is that each episode lasts 10 minutes, which is the amount of time suggested for young L2 learners. Previous research has suggested that after ten minutes, young L2 learners may not be able to cope with the cognitive demands of the viewing task, in particular the ones that are less skilled readers (Marzá & Torralba, 2015; Zabalbeascoa et al., 2015). Thus, after ten minutes, learner-viewers might simply disconnect from this activity (Marzá & Torralba, 2015; Zabalbeascoa et al., 2015). Along with timing, the characteristics of this

[4] https://forms.gle/dKMCexu1CrftXLue6

animated cartoon tie in with the selection criteria suggested by Donaghy (2019) to ensure comprehension. Specifically, the dialogues have a high degree of visual support and there is close connection between verbal input and actions (see Figure 9). In addition, only one character speaks at a time, and the episodes do not contain complex storylines. Finally, speech is clearly enunciated and there are no interfering factors such as loud music.

Figure 9.

Visual support in Charlie and Lola

In this investigation, learners' familiarity with the TV series was not an important factor that had to be controlled for. The British TV series was aired in Latin-American countries on Discovery Kids, a pay-per-view TV channel, between 2006 and 2009, when only the eldest participants of this study had been born (2009). Given that at the time this investigation was carried out Charlie and Lola was not available on any video streaming platform in Chile, the possibility that students were familiar with its content or the episodes in its original version was highly unlikely. Furthermore, the episodes available on YouTube are automatically blocked in Chilean territory due to its copyright (BBC).

As for its linguistic suitability, the analyses carried out with a VocabProfiler on Lextutor (BNC and COCA) (Cobb, 2019) showed that most of the episodes reached 90% coverage at K1 (see Table 11). Based on the threshold proposed by the only studies on vocabulary coverage with audiovisual input (80%; Durbahn et al., 2020, 2022), these results appear to indicate that Charlie and Lola is appropriate for low proficiency learners. In addition, the eye-tracking study conducted by Tragant and Pellicer-Sánchez (2019) may be used as further support. By studying fifth graders' eye movements while watching an episode of Charlie and Lola, Tragant and Pellicer-Sánchez (2019) concluded that even when learners spent longer time processing written input, they were fully capable of integrating text and images. Thus, again, the outcomes of this investigation suggest that Charlie and Lola is appropriate for this age group.

We additionally pilot tested a sample episode of Charlie and Lola to further confirm its suitability in the target context. More precisely, this pilot study was conducted with two groups of fourth graders from school 1 (*N*=40) in 2019 (onsite pilot testing group, see Appendix 1). We did not only explore learners' capacity to process the input (viewing self-efficacy) but also their perceptions on the materials (e.g. levels of enjoyment). On a Likert scale from 0-6, 90% of the participants indicated that the experience was highly enjoyable (4-6). As for episode comprehension, 80% of the participants found the audiovisual material comprehensible (4-6). Nevertheless, learners' responses suggested that following L2 captions was a possible but challenging task. Specifically, 52.5% of the respondents reported doing the task with ease (4-6), while 37.5% found it challenging (2-3) and 10% admitted that reading captions was a struggle (0-1).

Overall, the results indicated that the few students from the onsite pilot testing group that found the experience less enjoyable were also part of the group of students that struggled with written input processing and comprehension. This means to say that most low achievers enjoyed the experience regardless of the greater cognitive effort required to make as a result of their poorer language skills. This is evident in learners' answers to the open question given that 72.5% of the participants were open to watch more episodes of Charlie and Lola in class, whereas the rest of the respondents explained that the main reason why they were unsure of repeating the experience was associated to their concern about the consequences of skipping regular English lessons. Specifically, some students seemed to be unaware of the educational value of captioned-video viewing with respect to foreign language learning. By way of illustration, one of the participants commented "I like watching videos but we have to attend our English class." Likewise, among the participants who were willing to watch more episodes, only 31% emphasized the EFL learning potential of captioned videos, while the rest of the answers primarily focused on their levels of enjoyment and/or mentioned how funny the video and the characters were. By contrast, in a similar pilot study with secondary school learners from the same school, the participants' answers mainly addressed the potential benefits of watching captioned videos in the English class (Avello, 2020). Hence, it seems reasonable to assume that younger primary school learners need help to become more aware of the learning value of audiovisual input.

Along with students' participation, two EFL teachers from school 1 (year 4) were in charge of observing the pilot session and filling a questionnaire (see Appendix 1) to provide feedback to the researcher. On the Likert scale (1-6), both teachers reported learners' high levels of comprehension (5/6) and their ability to cope with the speed of captions (5/6). Even when both teachers believed that the students enjoyed the episode, one group (6/6) seemed to enjoy it more than the other (4/6). Yet, there was no clear explanation for this difference between groups. They believed that Charlie and Lola was appropriate for the age group (6/6) and hypothesized that captions may have supported comprehension (6/6). Learners' behavior was found to be outstanding (6/6) and they indicated that they would definitely use the animated TV series in class if they were given the materials. Once again, the teachers also confirmed that Charlie and Lola was appropriate for the target context.

2.3.1.1 Analysis and selection of the episodes

As mentioned above, the episodes were analyzed with a VocabProfiler on Lextutor (Cobb, 2019) in order to calculate their coverage (BNC and COCA), number of tokens, word types, type-token ratio and lexical density (see Table 11). To this aim, the scripts were downloaded from Subsaga.com (n.d.), which is an unofficial source of information. Therefore, the scripts were carefully checked while viewing each episode to correct their minor mistakes. In addition, by following Rodgers' (2013) procedures, the analysis was facilitated by erasing marginal words (e.g. interjections) and categorizing proper names (e.g. Lola) as K1.

Initially, 20 episodes from the full DVD collection (Carrington & Child, 2005-2008) were carefully watched and analyzed to assess their content and select potential target language constructions. Then, further analyses of the scripts were performed with AntConc 3.5.8 (Anthony, 2019) to more accurately identify potential unknown target language constructions whose repetitions (over three) were either concentrated in a single episode or distributed in multiple episodes. These procedures were also influenced by the extensive experience of this researcher teaching English to primary school learners in Chile. Finally, the yearly plans and course materials (coursebook and digital platform) implemented from first to fifth grade in school 1 were carefully studied to identify the target language constructions students encounter as part of the school program. Thus, apart from being comprehensible, the selected episodes had to fulfill the requirements to foster learning.

The resulting 15 episodes were pilot tested online between April and May 2020 with a small group of six primary school students from different schools in Chile (one-to-one), whose ages ranged from 9 to 12 years old (online pilot testing group, see Appendix 2). Their perceptions together with the adaptation of the vocabulary tests helped us narrow down the selection, which finally consisted of 11 episodes (see Table 11).

Table 11.

Vocabulary profile of each episode

Episode	Coverage	Tokens and Types	Type-token ratio	Lexical density
E1*: I will not ever eat a tomato.	K1: 82% K2: 91%	Tokens: 544 Types: 184	0.34	0.51
E2: I do not ever want my wobbly tooth to fall out.	K1: 92.4% K2: 94.9%	Tokens: 1.165 Types: 265	0.23	0.48
E3: Say cheese.	K1: 90.1% K2: 94.9%	Tokens: 1044 Types: 266	0.25	0.60
E4: You won't like this present as much as I do.	K1: 94.7% K2: 96.6%	Tokens: 1005 Types: 257	0.26	0.51
E5: I am just not keen on spiders.	K1: 91% K2: 94.6%	Tokens: 1.126 Types: 300	0.27	0.53
E6: There is only one sun, and that is me.	K1: 92.1% K2: 93.9%	Tokens: 1.025 Types: 246	0.24	0.48
E7: We do promise honestly, we can look after your dog.	K1: 95.8% K2: 97.6%	Tokens: 968 Types: 246	0.25	0.61
E8: Boo! Made you jump.	K1: 93.8%	Tokens: 777	0.31	0.55

	K2: 97.1%	Types: 240		
E9: I'm far too extremely busy.	K1: 90.7% K2: 94.2%	Tokens: 1.130 Types: 296	0.26	0.57
E10: I must take absolutely everything.	K1: 91.1% K2: 93.9%	Tokens: 952 Types: 285	0.30	0.51
E11: I will be especially, very careful.	K1: 90.6% K2: 93.8%	Tokens: 1.084 Types: 272	0.25	0.50
Total tokens: 10.820				
Total time: 110 minutes				

E= Episode

2.3.2 After-viewing activities

Viewing tends to be perceived as an effortless and leisure-oriented task (Vanderplank, 2016). Therefore, the use of activities may contribute to the learning process by encouraging viewers to make a greater mental effort to enhance comprehension (Montero Perez et al., 2018; Vanderplank, 2016). As explained in section 2.3.1, primary school students do not necessarily see pedagogical value in the use of videos. The onsite pilot study with primary school learners suggested that videos may be considered as fillers rather than learning activities. By the same token, in the online pilot study, half of the participants believed that the viewing experience was not conducive to learning unless it was complemented with activities. On the whole, learners' perceptions on the use of audiovisual materials might have been highly influenced by their experience as foreign language learners under more traditional methodologies, where the use of audiovisual materials is either limited or non-existent.

In view of the findings aforementioned, the sets of activities planned for the purpose of this study were implemented *after* viewing each episode (see Figure 10) for a number of reasons. First of all, the activities would somehow keep the lesson structure and methodology learners were used to. In addition, their implementation would raise awareness of the fact that the episodes were not fillers and required attention to complete the activities successfully. Finally, depending on the instructions provided to each subgroup (i.e. focus on meaning or focus on meaning and constructions (i.e. words and phrases), it was possible to compare learners' outcomes in incidental and intentional learning conditions (see Figure 11; see section 1.4.1).

Figure 10.

Viewing session structure

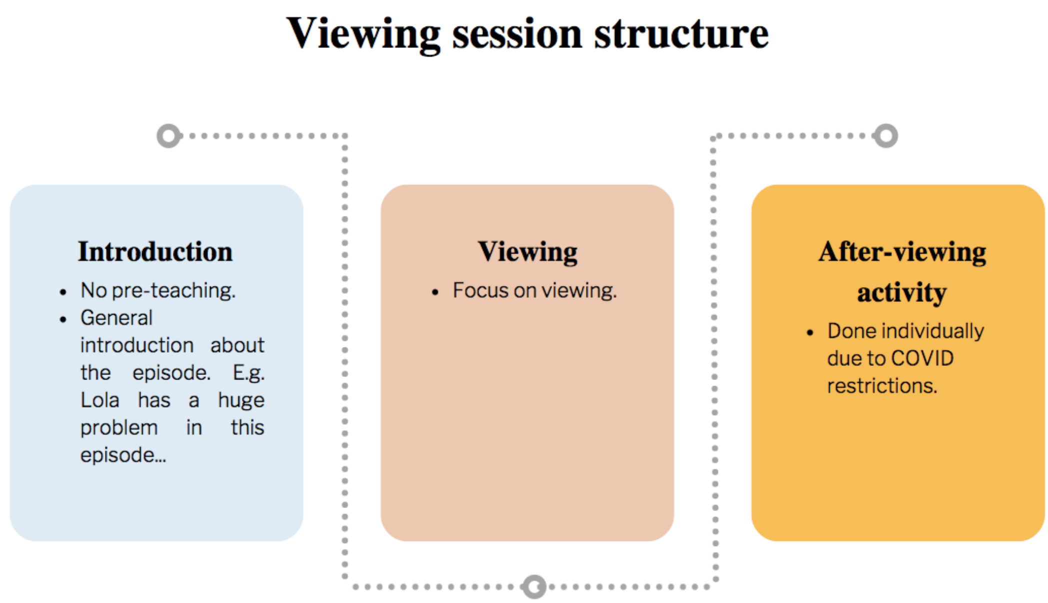
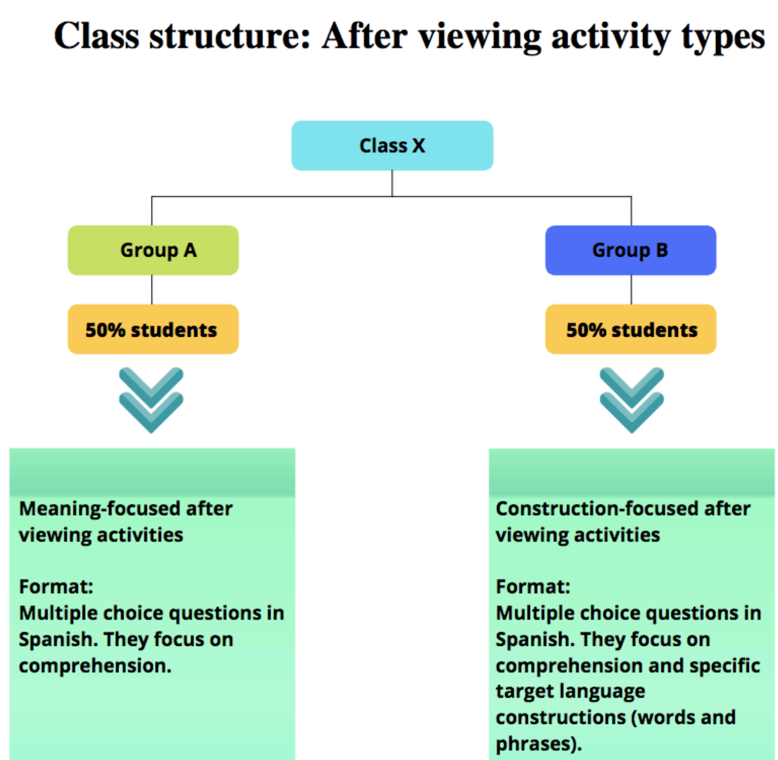


Figure 11.

Class structure: After viewing activity types



As illustrated in Figure 11, the participants in each class were asked to complete either a meaning- or construction-focused activity individually after watching the episodes. Specifically, the students learned that they belonged to a specific sub-group (A or B) but they were not provided further information about it. The list (poster size) with the group members was permanently pasted on their classroom board to avoid misunderstandings. Given that in different courses it was a common practice to be assigned a test form (A or B) to prevent cheating, the participants did not ask further questions about this organization. Exceptionally, at the beginning of the first viewing session, the students were given the sheets before watching the video to scan the questions and somehow direct their attention as a function of the activity assigned. Specifically, the participants were given 3 minutes to go over the questions, and then were asked to turn over the sheets to devote their attention to the viewing task.

The meaning-focused activities aimed to encourage learners to comprehend the main ideas from each episode. Therefore, these activities never directed learners' attention to the target language constructions during the investigation. Instead, they sought to foster incidental construction learning (e.g. vocabulary), which would occur as a by-product of the viewing experience (Hulstijn, 2003). As for the construction-focused activities, they promoted intentional L2 learning. This type of activity worked as a sort of test announcement (see section 1.4.1) since learners' attention was not directed to specific target language constructions before viewing. Yet, in the case of the constructions that appeared in multiple episodes[5], they were tested after their first encounter(s) to increase the possibility of being noticed in subsequent encounters. Hence, some additional analyses were run in order to determine whether this factor played some role in the outcomes (see sections 4.7 and 5.6). According to Nation & Webb's (2011) Technique Feature Analysis, the construction-focused activity was relatively effective (see Appendix 3).

2.3.2.1 Meaning-focused activities

The meaning-focused activities consisted of five multiple choice questions (literal and inferential) (see Appendices 4-14). They were fully written in learners' mother tongue (L1 Spanish) to assess comprehension and not learners' L2 knowledge. They included four possible

[5] The target language constructions were either encountered in a single episode (massed) or in multiple episodes (spaced). This is a factor that cannot be controlled when using authentic audiovisual input.

answers and the 'I don't know' alternative to prevent students from guessing the possible answer. The questions could only be answered correctly if the participants had watched the episode and comprehended its content. The activity was short since the schedule at both schools was restricted by the pandemic and the EFL teachers needed to have enough time to implement their plans.

2.3.2.2 Construction-focused activities

The construction-focused activities consisted of two comprehension questions and four construction-focused questions. Only the key target language constructions were written in English, while the questions were formulated in students' mother tongue. The construction-focused questions could only be answered if students had figured out their meaning and/or structure while watching the episodes. The key target language constructions were either part of the question statement or the alternatives (e.g. Lola repeats that *'she is busy'* several times along the episode. What's the meaning of *'busy'*?) (see Appendices 4-14).

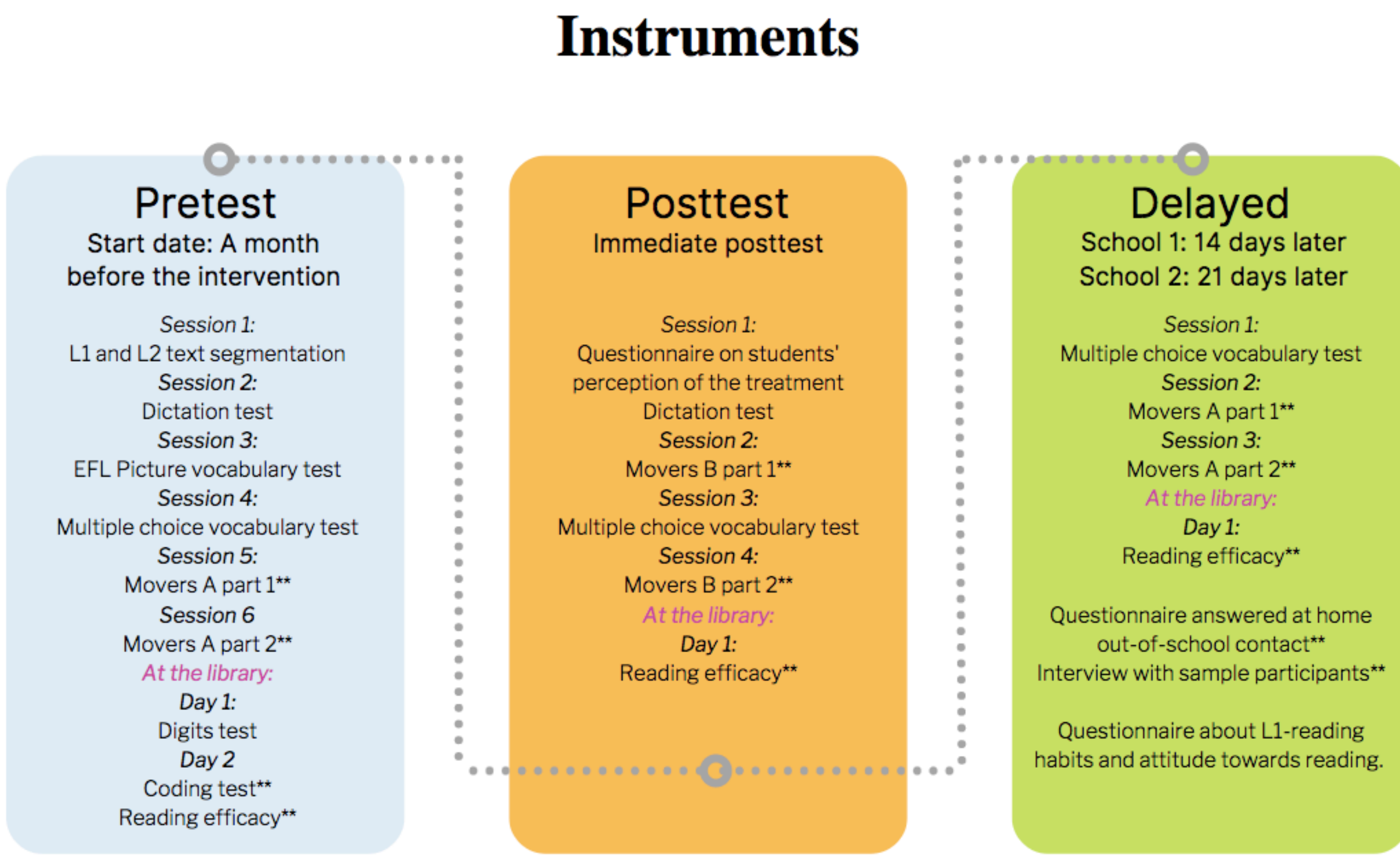
2.4 Instruments

Due to the pandemic, the instruments were carefully designed to be administered in pen-and-paper and online format (Google forms). Although this study was carried out onsite with the experimental groups at both schools (i.e. pen-and-paper instruments), the students that were put in quarantine had to complete some activities online through the official platform used by each school (Zoom and Microsoft Teams) individually or in small groups (4 students maximum, depending on the requirements). Group size was strictly reduced to control for any possible factor that may affect learners' outcomes. The participants were asked to keep their camera on during the whole session (see Appendix 39). Figure 12 summarizes the instruments administered over time.

It is worth mentioning that at school 2, the investigation was conducted by their EFL teachers under the supervision of this researcher, requiring additional meetings and explanations. Specifically, while a teacher was implementing one of the sessions onsite, the activities were observed through Zoom or WhatsApp. Thus, the number of instruments administered in school 2 was reduced to alleviate teachers' workload (see Figure 12).

Figure 12.

Summary of the instruments administered over time



**Only administered in school 1.

2.4.1 EFL picture vocabulary test

This test measured general vocabulary knowledge at the level of meaning recognition in young EFL learners. Its format is based on the adaptation made by Puimège and Peters (2019a) of the Picture Vocabulary Size test (PVST) created by Anthony and Nation (2017) to assess young learners (L1 and L2 English speakers). The test by Puimège and Peters (2019a) had been administered to the online pilot-testing group and the analyses detected a floor effect, which was corroborated by the head of the English department from school 1. In addition, the test included a considerable number of conflicting items due to their cognateness (see Muñoz, 2020a) or borrowing to L1 Chilean Spanish. Therefore, we only kept its format and designed a test with a reduced number of items since the students from the online pilot-testing group indicated that the test was longer than expected.

The picture EFL vocabulary test designed for the purpose of this study consists of 50 target words, equally subdivided into K1 and K2 frequency bands according to the analysis performed on Lextutor (Cobb, 2019). These target words (see Appendix 15) were selected from the A2 key for schools vocabulary list developed by Cambridge English Assessment (2020) (Lanes et al., 2019) which drew on different sources of information to ensure their appropriateness. First of all, they determined the CEFR level for each item according to the English Vocabulary Profile. Also, they calculated the frequencies for each item as found in the Cambridge Learner Corpus and Spoken British National Corpus 2014.

As in Puimège and Peter's (2019a) test, each target word was uttered in isolation and then in a non-defining sentence that only suggested the corresponding part of speech. As shown in Figure 12, the verbal input was presented in written and aural form. The prompts were read by an English native speaker. Out of four pictures, the students had to select the one that represented the meaning of the target word (A, B, C or D). Also, they were given the chance to select the 'I don't know' option to prevent them from guessing the correct answer. Each sentence was repeated only once and students had only ten seconds to think and select the correct alternative. We created a video with all the questions in order to keep the same testing conditions in all the groups[6].

The pictures were taken from Pixabay (n.d.) and the selection process was supported by two EFL teachers and an expert researcher. The resulting test was first pilot tested with a native

[6] https://drive.google.com/file/d/1444Xd2C_W7Gkbjo_ZVIshhilduZ-pD6t/view?usp=sharing

English speaker who was expected to score 100%. Then, this instrument was administered online to 6 groups of EFL learners from Chile (see Appendix 16) (*N*=188) to rule out the possibility of a floor effect. As expected, a Wilcoxon Signed Rank test revealed that learners scored significantly higher at K1 than K2 words (*M*=16.56, *SD*=6.27 vs. *M*=14.15, *SD*=6.22; z= -8.863, p<.01, r=.64). The Cronbach alpha coefficients obtained with the pilot groups were .908 for K1 words, .898 for K2 words and .898 for the whole test. Likewise, in the experimental groups, the Cronbach alpha coefficients obtained were .866 for K1 words, .814 for K2 words, and .913 for the whole test. As pointed out by Pallant (2016), the Cronbach alpha coefficient should always be above .7, thus the values obtained in this investigation were high. This test was also found to correlate significantly with the cognitive and L2-related instruments administered in this investigation (see Table 12). Along with these analyses, at the end of the pilot testing sessions, the EFL teachers filled in a questionnaire to provide feedback to the researcher (see Appendix 17). In short, the teachers believed that the test and the target words were appropriate for the context. The pictures addressed as confusing were replaced by clearer options.

When the test was administered online, students could hear the audio and see the pictures on their form (see Figure 13). As for the pen-and-paper format, a projector and speakers were employed to show the video, while the students had to record their responses on their answer sheet (see Appendix 18). The total amount of time required to administer this test was 15 minutes, approximately. After question 25, students had one minute to stretch and then focus on the test again. Each correct answer was awarded one point (maximum score=50).

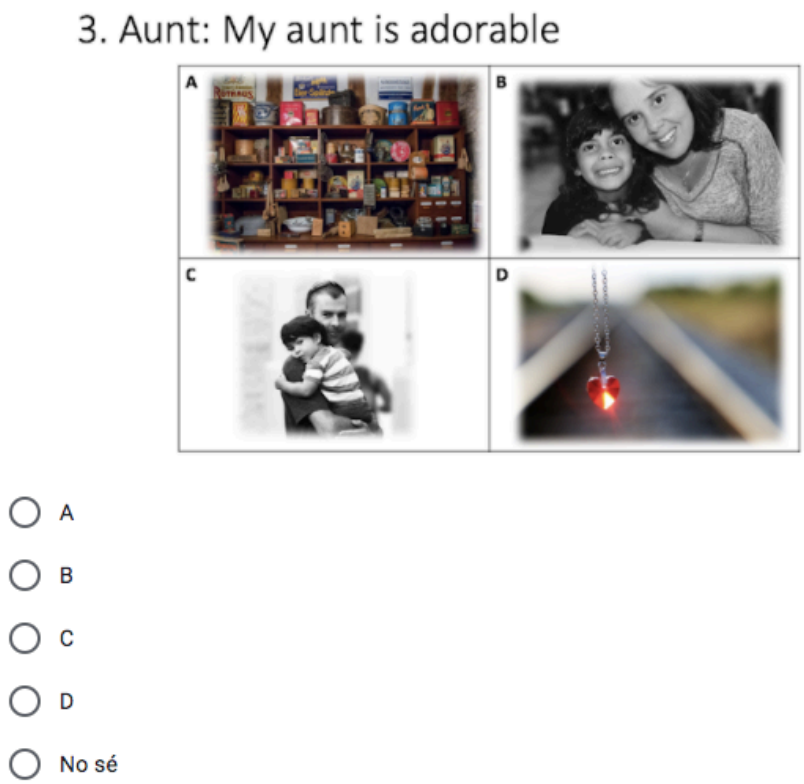
Table 12.

Pearson correlations between the EFL picture vocabulary test with the cognitive and L2-related instruments

	PSTM	WM	Processing speed	English text segmentation	English reading efficacy	Listening skills
EFL picture vocabulary test	r=.290 p=.006 R^2= .08	r= .215 p=.044 R^2= .04	r= .272 p=.010 R^2= .07	r=.592 p<.001 R^2= .35	r=.620 p<.001 R^2= .38	r=.673 p<.001 R^2= .45

Figure 13.

Sample item of the Picture EFL vocabulary test (online format)

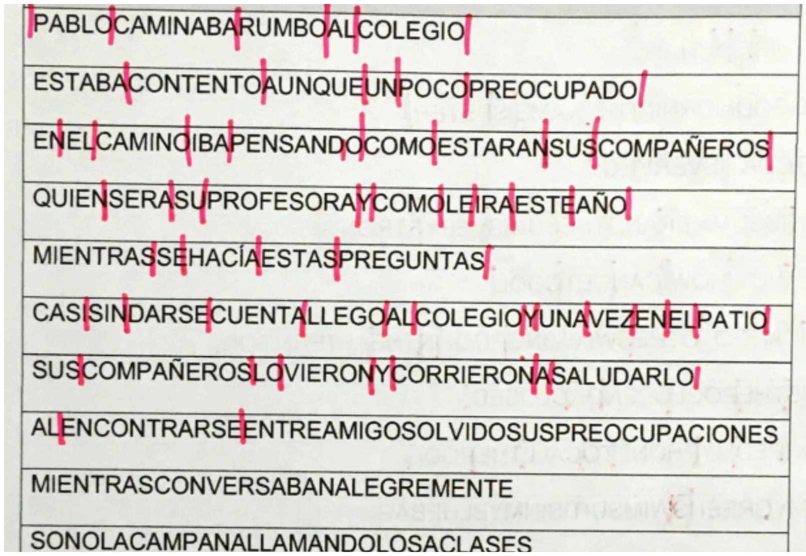


2.4.2 Text segmentation

This test attempted to measure silent reading fluency (Toscano-Fuentes & Julián de Vega, 2018) in L1-Spanish and L2-English. This type of instrument has been shown to encompass the assessment of lower-level reading skills such as phonological decoding and visual word recognition (Torres-Díaz et al., 2020), as well as the role of vocabulary and grammatical knowledge (Alderson et al., 2015) to achieve text comprehension. In other words, this test does not neglect reading comprehension given that it is difficult to detach lower from higher reading processes. Still, this measure provides further evidence concerning lower-level reading processes (see Figure 14).

Figure 14.

Text segmentation test in Spanish (sample)



PABLOCAMINABARUMBOALCOLEGIO
ESTABACONTENTOAUNQUEUNPOCOPREOCUPADO
ENELCAMINOIBAPENSANDOCOMOESTARANSUSCOMPAÑEROS
QUIENSERASUPROFESORAYCOMOLEIRAESTEAÑO
MIENTRASSEHACÍAESTASPREGUNTAS
CASISINDARSECUENTALLEGOALCOLEGIOYUNAVEZENELPATIO
SUSCOMPAÑEROSLOVIERONYCORRIERONASALUDARLO
ALENCONTRARSEENTREAMIGOSOLVIDOSUSPREOCUPACIONES
MIENTRASCONVERSABANALEGREMENTE
SONOLACAMPANALLAMANDOLOSACLASES

Due to the testing procedures, this test was only administered in pen-and-paper format (see Appendices 19 and 20). Learners had one minute to draw vertical lines between the words from a text that lacked spaces and punctuation marks. They had to do it in order, so they were not allowed to skip lines to improve their performance. The use of erasers was forbidden in all the tests that had to be completed under time pressure. Therefore, in case they made a mistake, they could indicate it with a cross and then draw the line on the right place. The texts were adapted from the EGRA test (Early grade reading assessment) used by Fernández Corbacho (2016) to ensure that they were suitable for the age group and the target context. The resulting test was pilot tested with students from school 1 (five students from year 6 and three classes in year 3) in order to anticipate its appropriateness and improve the administration process (e.g. instructions). The words that were accurately segmented were awarded one point. The maximum score was 72 in Spanish and 74 in English.

The Pearson correlations in Table 13 show evidence of the relationship between text segmentation in English and Spanish with other language-related measures. Considering the factors that influence reading comprehension (see sections 1.3.2 and 1.3.3), it is unsurprising that the relationships between most of these measures are strong (R^2 >.25; Larson-Hall, 2010, p.162).

Table 13.

Pearson correlations between English and Spanish text segmentation and the rest of the language-related measures

	English text segmentation	Spanish text segmentation
Spanish text segmentation	r= .668, p<.001, R^2= .44	1
EFL picture vocabulary test	r= .592, p<.001, R^2= .35	r= .378, p<.001, R^2= .14
Spanish reading efficacy	r= .531, p<.001, R^2= .28	r= .515, p<.001, R^2= .26
English reading efficacy	r= .508, p<.001, R^2= .25	r= .309, p=.012, R^2= .09
L2 listening skills	r= .683, p<.001, R^2= .46	r= .451, p<.001, R^2= .20

2.4.3 Questionnaires

2.4.3.1 Questionnaire on L1-reading habits and attitude towards reading

This questionnaire was administered to measure the influence of learners' L1 reading habits and attitude towards reading on the English and Spanish reading efficacy scores over time (pretest, posttest and delayed posttest). To this aim, we adapted the questionnaire developed and administered by Granena et al. (2015). The main reason why the questions and the alternatives were adapted had to do with contextual factors (educational and cultural issues). Even when Chile is one of the leading countries in South America as regards reading skills development and instruction, the evidence has shown that there is plenty of room for improvement. Besides, reading is seen as an obligation, so this activity has not been found to be an important part of Chilean's leisure time (Mineduc, 2015; Rivas, 2019).

The questionnaire consisted of 8 questions (2 Likert-scale and 6 multiple-choice questions) that were assessed by an expert professor and an EFL teacher prior to their administration (pen-and-paper format). The instrument (see Appendix 21) was only administered to the experimental groups from school 1 and the Cronbach alpha coefficient obtained was .791, which is above the minimum value expected (Pallant, 2016). The scoring criterion and main findings are reported in Appendix 21. This instrument correlated significantly with the test on Spanish text segmentation (r=.237, p=.026, R^2=.05). As for the administration procedures, the questions were read out loud to provide further explanations and guarantee that the participants comprehended the questions.

2.4.3.2 Questionnaire on out-of-school contact

This questionnaire, based on Muñoz's (2020b) instrument, aimed to explore learners' contact with the target language outside the classroom. The modifications made to the original questionnaire attempted to match contextual and learner characteristics (e.g. age). In addition, the alternatives allowed the researcher to estimate the number of hours each student was exposed to each source of input monthly. In view of children's inaccuracy to calculate the amount of time they spend doing an activity, the questionnaire was sent to their parents through the school's official channels, so it could be answered in conjunction (parent and child). The instrument was administered in online format since the respondents were directed to specific questions according to their answers[7].

A preliminary version of this questionnaire was pilot tested with parents from school 1 (convenience sample) and school 2 (years 4, 5 and 6) a year before the actual experiment. Therefore, the experimental groups were not included in the piloting. Some families from school 2 figured out the aim of the experiment from the questionnaire and showed their willingness to use videos with written support with their children to increase their exposure to the target language (comments section). As a result, we decided to administer this questionnaire at the end of the experiment. It is important to point out that this questionnaire was only given to the families from the experimental groups in school 1. As explained earlier, only some of the instruments were administered at school 2 to alleviate the teachers' workload.

2.4.3.3. Questionnaire on students' perceptions of the treatment

This questionnaire was administered at the end of the treatment (before the posttests) to explore learners' perceptions of the treatment and their performance. It was given before the posttests to prevent the participants from providing responses based on test outcomes. Its original version was administered to the online pilot testing group and it was modified according to their responses and the identification of unclear questions. It was available in pen-and-paper and online versions[8] (see Appendix 22). Of the eight questions, three had a multiple-choice format and eight prompts had to be rated with a Likert scale.

[7] https://forms.gle/zQ1oae6oHmFqmVxT8

[8] https://forms.gle/81sZwaJ2857tE2nP8

The questions were read out loud by the researcher to provide further explanations. Then, the participants were given additional time to think and check their answers. While the students were answering the questionnaire, the researcher was fully available to clarify the questions that emerged.

2.4.3.4. Interview on students' perceptions of the treatment

To further explore students' answers in the questionnaire, three sample groups of students (*n*=18) from school 1 were taken to the school library in order to be interviewed. As shown in Table 14, the participants differed as regards year level, gender, vocabulary knowledge (as a measure of proficiency) and activity type. Their teachers of English supported the selection process. The conditions appeared to be appropriate to encourage the participants to give their opinions since they were familiar with the researcher leading process, and students knew each other since preschool. In the case of fifth graders, the students seemed to feel more comfortable when interacting with students of their same gender, which is why they were interviewed separately (group 2: female; group 3: male). The questionnaire described in section 2.4.3.3 was used as a *mediational tool* since learners were already familiar with the instrument (A. Pinter, personal communication, June 24, 2022). The questions used in the questionnaire triggered learners' discussions and co-construction of answers. Thus, the interview process focused on learners' discussions and opinions, since learners provided rich information about their viewing experience.

Table 14.

Characteristics of the students that participated in the group interviews

Participant	Gender	Year level	Interview group	Vocabulary level	Activity type
TA	Female	Fourth grade	1	High	Meaning-focused
CV	Female	Fourth grade	1	Low	Construction-focused
JJ	Male	Fourth grade	1	Intermediate	Construction-focused
JA	Male	Fourth grade	1	Low	Construction-focused
TE	Female	Fourth grade	1	Intermediate	Meaning-focused
RB	Female	Fourth grade	1	Low	Meaning-focused
MJ	Female	Fifth grade	2	Low	Construction-focused
AU	Female	Fifth grade	2	Intermediate	Construction-focused
AG	Female	Fifth grade	2	High	Meaning-focused
MG	Female	Fifth grade	2	High	Meaning-focused
JG	Female	Fifth grade	2	High	Meaning-focused

RT	Female	Fifth grade	2	Low	Construction-focused
JM	Male	Fifth grade	3	Low	Meaning-focused
JP	Male	Fifth grade	3	Intermediate	Construction-focused
RA	Male	Fifth grade	3	High	Meaning-focused
NM	Male	Fifth grade	3	High	Construction-focused
JC	Male	Fifth grade	3	High	Construction-focused
MC	Male	Fifth grade	3	Intermediate	Meaning-focused

2.4.4 Digits test

The digits subtest from the WISC-V battery measures working memory capacity (Wechsler Intelligence Scales for Children) (Wechsler, 2014). In the current investigation, we administered the first two sections: forward and backward digit span. Forward digit span measures auditory rehearsal and working memory's brief storage capacity (phonological short-term memory) (Weiss et al., 2019), whereas backward digit span taps complex working memory, and involves transformation and manipulation of information (Weiss et al., 2019; see section 1.3.5.1).

The tests were administered at the school library (one-to-one) in order to ensure that the students could fully focus on the tasks. It was administered in students' L1 and the numbers had to be orally uttered by the researcher at a consistent tone and speed (one number per second) to avoid chunking or interference. Two practice items were included at the beginning to clarify doubts and ensure that the students followed the instructions. The test finished when the students got two incorrect answers from the same item (see Appendix 23). As for the forward digit span test, the number sequences went in increasing order and they only needed to be repeated by the participants. At the backward digit span test, students were also asked to repeat the increasing number sequences but in reversed order. Self-corrections were allowed as long as the participants clearly stated which was the correct response. The maximum raw score per section was 18, and the time required to administer this test was five minutes maximum.

This test was pilot tested with control group 1 in December 2020 in order to improve its implementation (e.g. instructions) and assess its appropriateness in the target context. The experience confirmed that the test was suitable for the target participants, and indicated that the instrument successfully discriminated between students with lower and higher PSTM/working memory capacity.

2.4.5 Coding test

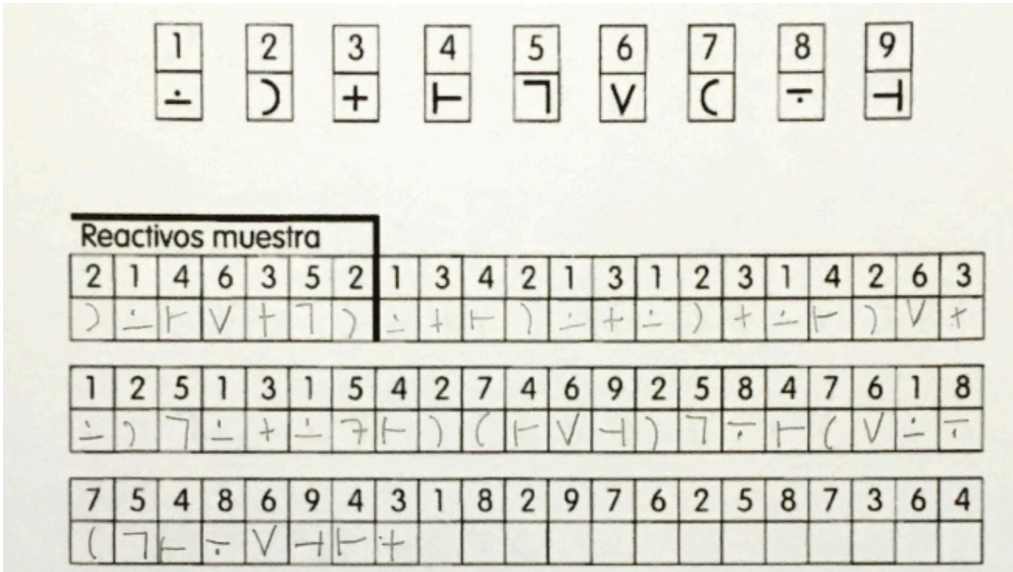
Coding, a subtest from the WISC-IV battery (Wechsler Intelligence Scales for Children) (Wechsler, 2003), essentially focuses on the measurement of *visual* processing speed due to the visual stimuli used for this purpose. Along with processing speed, this instrument is also considered to be a measure of short-term visual memory, motor/graphomotor processing speed, visual scanning ability, visual discrimination, multitasking and directing sustained attention to task (Flanagan & Alfonso, 2017; Weiss et al., 2019). Based on the scope of this instrument, it could be hypothesized that visual processing speed might play a significant role in the processing and learning from multimodal input due to the importance of learners' ability to integrate verbal and non-verbal input under time pressure (Pellicer-Sánchez, 2022; Sadoski & Paivio, 2013).

This test was administered at the library in both schools since students needed to be fully focused on the task. Prior to its administration, students were asked to do some practice items to ensure that the instructions were followed. The participants had two minutes to copy the symbols displayed on top of the page according to the numbers indicated in each square (see Figure 15). They were told to do it as fast as possible, so erasers were not allowed and all possible mistakes had to be amended with their pencil. The symbols had to be drawn in order, therefore the participants could not skip lines nor spaces. The tests that did not fulfill this requirement were removed from the data. Each correct answer was awarded one point, even when the symbols were not perfectly drawn (see Appendix 24).

It is important to acknowledge that the main drawback of this instrument has to do with the fact that children rely on their graphomotor skills when drawing the codes. Therefore, considering that in some cases a lower score might be the reflection of a child's poor graphomotor skills rather than his/her slow processing speed, this test was administered in small groups (2-3 students) to observe individual performance and rule out this possibility.

Figure 15.

Sample coding test



2.4.6 Listening skills

Two sample Movers tests (paper A and paper B) (Cambridge Assessment English, 2018) were administered in order to assess learners' listening skills over time (pretest, posttest and delayed posttest) (see Figure 16). Paper A (see Appendix 25) was pilot-tested onsite with two groups of fifth-graders from school 1 in November 2019 (1.5 year before the experiment) in order to determine whether this type of instrument was suitable for the context. The amount of time required to test the five sections was 45 minutes (approximately), and even when the test successfully discriminated between low and high achievers, the instrument appeared to be cognitively demanding. Specifically, after completing the second section, some of the participants from the pilot groups were found to struggle to stay focused on task. As a result, we decided to implement the first four sections in two different sessions.

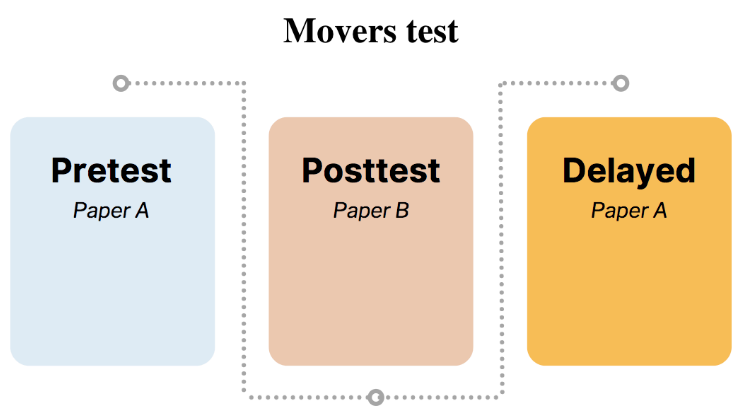
Correct answers were awarded one point, so each section was worth five points and the maximum score was twenty points. The test was administered onsite in pen-and paper format (see Appendices 25 and 26). An online version[9] was also developed in order to assess the control group, as well as the students that were unable to go to school (Covid-19).

[9] Movers A part 1 https://forms.gle/8CBegBKENDJ7CBnQ9
Movers A part 2 https://forms.gle/kULFfgDnZyNmtGjn6

The Cronbach alpha coefficient obtained for test A was .722 with the pilot group. As regards the experimental groups, the Cronbach alpha values obtained for paper A were .655 at pretest, and .794 at delayed posttest, while the coefficient obtained for paper B at posttest was .682. Considering that the reliability analyses performed with a small number of items lead to low values (Pallant, 2016), and this test generated four main scores, the results suggest that both instruments have acceptable internal consistency (Pallant, 2016).

Figure 16.

Movers test administration



2.4.7 Reading efficacy in English and Spanish

Reading efficacy (see Llanes, 2018) is a measure that integrates learners' reading speed (words read per minute=WPM) and comprehension, given that both, lower- and higher-level reading processes have been shown to be equally important to discriminate high- and low-achievers (Grabe, 2009). This measure differs from the text segmentation test in that the emphasis is on comprehension and learners' performance as regards lower-level reading processes is gathered from reading speed. Yet, the experimental groups' outcomes of both instruments were highly correlated (see Table 15). It is important to note that silent reading was chosen over oral reading (reading aloud) since the former seems to be closer to the processing of captions. Although it is true that oral reading provides further evidence of the reading process, it remains unclear whether it can reliably assess comprehension, this is why

Movers B part 1 https://forms.gle/Pk42QL4u7veVf7jv9

Movers B part 2 https://forms.gle/maeqemipgEGvtVCi9 it has been addressed as mere “barking at print” (Alderson et al., 2015, p.71). As Sadoski and Paivio (2013) highlight, word recognition differs from lexical access in that L2 readers may well be able to decode a word whose meaning is unknown.

Table 15.

Correlations between the reading measures

	ER efficacy	SR efficacy
English text segmentation	r= .508, p<.001, R^2= .25	r= .531, p<.001, R^2= .28
Spanish text segmentation	r= .309, p=.003, R^2= .09	r= .515, p<.001, R^2= .26
	ER speed (WPM)	SR speed (WPM)
English text segmentation	r= .431, p<.001, R^2= .18	r= .618, p<.001, R^2= .38
Spanish text segmentation	r= .324, p=.002, R^2= .010	r= .631, p<.001, R^2= .39

As for Spanish reading efficacy (SR efficacy henceforth) (see Appendices 27-29), the fiction texts (A, B and C) were adapted from the supplementary materials (Santillana, 2014) developed to train Chilean fourth graders for the national standardized test on reading skills (SIMCE= Sistema de Medición de la Calidad de la Educación). To ensure text comparability, the texts were analyzed on Renaissance (n.d.) to obtain their ATOS readability formula (text complexity) (see Table 16). The comprehension questions were adapted to ensure that the answers to the comprehension questions could only be extracted from the text. The final poll of items consisted of six multiple-choice questions that focused either on textually explicit/literal (4 questions) or textually implicit information (2 questions). A textually implicit question requires the connection of different statements within the text, and does not consider L2 readers’ background knowledge (see Alderson, 2000, pp. 87-88). Each test item had a single correct answer, three distractors, and the ‘I don’t know’ option to avoid the ‘threat of guessing by test takers’ (Jeon & Yamashita, 2021, p. 270).

Table 16.

Spanish reading efficacy. Text characteristics

Text	ATOS level (readability formula)	Word count	Average word length	Average sentence length
Text A	8.6	200	4.5	20
Text B	8.1	200	4.5	20
Text C	8.8	200	4.5	18.2

With respect to English reading efficacy (ER efficacy henceforth) (see Appendices 30-32), the non-fiction texts (A, B and C) were adapted from Pre-A1 starters sample papers (Cambridge English Assessment, 2018) to match the characteristics of the target participants. The tool used to assess text readability (Web FX, n.d.) indicated that all the texts were easy to read (see Table 17). The five comprehension questions had a multiple-choice format that followed the same structure used in the SR efficacy tests. Of the five questions, four focused on explicit information and only one tested textually implicit information.

Table 17.

English reading efficacy. Text characteristics

Text	Flesch Kincaid Reading Ease	Flesch Kincaid Grade Level	Number of sentences	Number of words	Average Nº of words per sentence	Average syllables per word
Text A	91.9	2.9	11	108	9.82	1.24
Text B	92.7	2.8	11	108	9.82	1.23
Text C	93.5	2.7	11	108	9.82	1.22

The evidence obtained from these measures was reading speed (WPM= *[nº of words in the text/number of seconds used to read the whole passage]*60*) and comprehension, separately. Each comprehension question was assigned one point. Then, the raw comprehension score was used to calculate the percentage of comprehension (number of correct answers*100/Nº of questions). Finally, the formula used to calculate reading efficacy (as in Llanes, 2018) was ([WPM x % comprehension]/ 100). In view of the fact that reading speed (WPM) varied among the participants, there was no maximum score for this test.

The procedures followed to pilot test and administer the instruments are summarized in Figure 19. Texts A and B (in English and Spanish) were pilot tested with the online pilot testing group to examine the quality of the questions and their alternatives, and to detect all possible flaws (e.g. ceiling or floor effect). They were administered individually by sharing the screen with the participants. The resulting tests were made available in pen-and-paper and online format (Google forms[10]) according to needs of each group (see Figure 19). Texts A and B (in both languages) were given at pretest and posttest in counterbalanced order to

[10]SR efficacy test A (part 1) https://forms.gle/7bXXUyrWZrJvso389 ; SR efficacy test B (part 1) https://forms.gle/Xhaa4ZCok5V76mqh7 ER efficacy test A (part 1) https://forms.gle/MSA3E5sSb9yvEK1Z9 ; ER efficacy test B (part 1) https://forms.gle/GbRie6MjdcdJk5mRA

guarantee that learners' progress or performance over time was not associated to a specific text but the treatment (see Figure 19). In online format, the reading efficacy tests (A and B) were subdivided into two forms (one for reading and one for questions) to accurately measure reading speed and prevent the participants from going back to the text to answer the comprehension questions. Texts C (in English and Spanish) were pilot tested with sixth graders from school 1 to examine the quality of the test items, and rule out ceiling or floor effects. The resulting tests (text C in both languages) were implemented at delayed posttest with the experimental groups (see Figure 19).

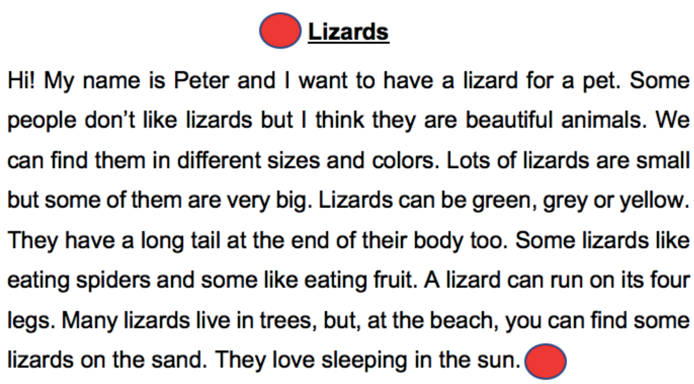
The administration of the reading efficacy tests took place in small groups to track reading speed more accurately (by using a stopwatch per child) (see Figure 17), and also guarantee that the instructions were followed. First of all, the students were instructed to read the texts at their own pace to achieve comprehension. This instruction was key to fulfill the main aim of this instrument; therefore, it was repeated several times. In addition, it was emphasized that the text had to be read only once. Having listened to the instructions, the students were explicitly told to start reading. As shown in Figure 18, the texts contained red circles, which signaled the beginning and the end of the reading process. Specifically, the learners were asked to raise their hand when reaching the second circle. In onsite format, the sheet with the text had to be turned over to answer the set of multiple-choice questions on a different piece of paper, whereas in online format, a link directed the participants to the questions.

Figure 17.

Administration of the reading efficacy tests

Figure 18.

The red circles that signaled the beginning and the end of the reading process

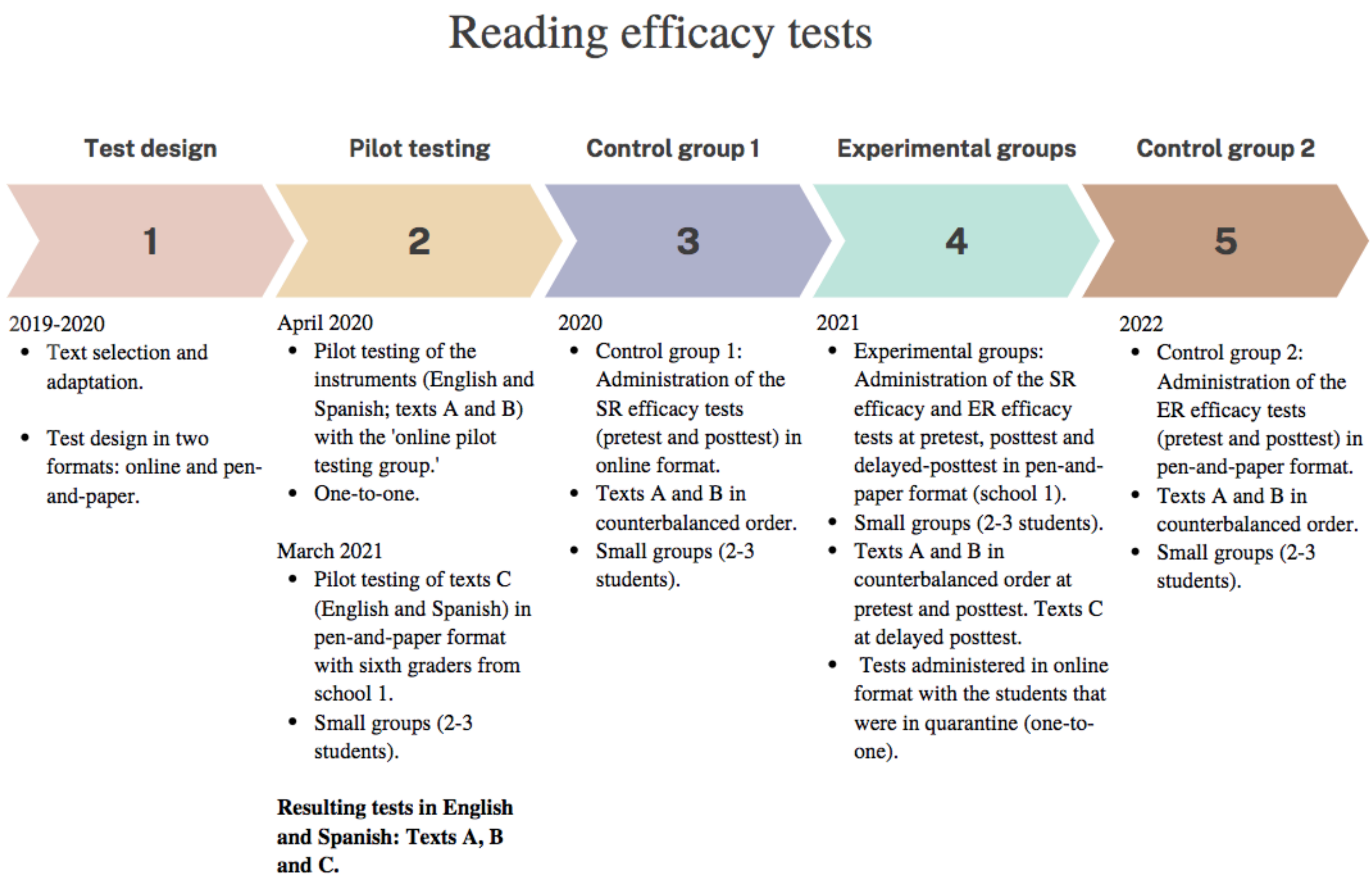


Lizards

Hi! My name is Peter and I want to have a lizard for a pet. Some people don't like lizards but I think they are beautiful animals. We can find them in different sizes and colors. Lots of lizards are small but some of them are very big. Lizards can be green, grey or yellow. They have a long tail at the end of their body too. Some lizards like eating spiders and some like eating fruit. A lizard can run on its four legs. Many lizards live in trees, but, at the beach, you can find some lizards on the sand. They love sleeping in the sun.

Figure 19.

Procedures followed to pilot test and administer the reading efficacy tests



2.4.8 Vocabulary tests

The target words consisted of 36 target items whose repetitions (3-20) were either concentrated in a single episode (18 words) or distributed in multiple (2-3) episodes (18 words) (see Table 18). Of these 36 target words, 24 were nouns and 12 were adjectives. These words were not cognates and they were not explicitly included in the yearly plans from both schools nor in the course materials from school 1[11]. These words were also judged as unknown by the teachers of English from both schools. On average, 4 new target words appeared in each episode.

The target words were classified in terms of regularity (more or less consistent with L1 patterns) with the aim of assessing the influence of this factor in the outcomes (see section 1.4.4.1). Although most of the target words included orthographic patterns that are not found in L1-Spanish (e.g. double consonant clusters such as bb and tt), regularity referred to the extent to which the correspondence between phonemes and graphemes (spelling) better resembled the transparent orthographic patterns found in L1-Spanish. These words did not have to include vowel digraphs (e.g. s**au**sage) or diphthongs (e.g. pill**ow**). This classification specifically focused on vowels and diphthongs since they seem to be the main source of difficulty at early stages (Sun-Alperin & Wang, 2008). To guarantee an accurate classification, two Spanish speakers (adults) that had never been instructed in English were asked to read the target words out loud to identify the target items whose vowel sound-symbol correspondence were more consistent with L1 patterns. Therefore, the words labelled as 'less consistent with L1 patterns' were expected to be decoded inaccurately (e.g. /mud/ instead of /mʌd/ for mud).

Prior to the intervention, the vocabulary tests were built in pen-and-paper and online format (Google forms) according to the groups' needs. With the online pilot testing group, the students were asked to keep their camera and microphone on while completing the tests in order to hear their comments (think aloud) and identify possible flaws. The initial selection of target items consisted of 63 words, however, the pilot administration indicated that test length was inappropriate for primary school learners. As a result, the number of target words was reduced to 36. The amount of time required to complete the dictation test was 14 minutes, and approximately 10 minutes for the multiple-choice test.

[11] We only had access to the materials from school 1.

Multiple actions were taken to reduce the risk of test effects. To start with, the administration of the dictation preceded the multiple-choice test. In addition, these instruments were given on different days, including an additional session in the middle, where a different test was applied (see Figure 11). The administration of the vocabulary tests took place approximately 25 days before the beginning of the treatment. Finally, in the multiple-choice test, the target words were available within a total of 144 alternatives, therefore, the high number of words read in a limited amount of time might reduce the possibility of learning target word forms from this instrument.

Table 18.

Target vocabulary words

Word	Nº of Episodes	Repetitions	TV corpus (Cbeebies[a])	Concreteness ratings[b]	Part of speech	Regularity[c]	Nº of letters
Bandage	1	5	4.09	4.85	Noun	1	7
Busy	1	9	5.58	2.41	Adjective	1	4
Cabbage	2	3	4.67	4.75	Noun	1	7
Careful	3	6	5.42	1.86	Adjective	1	7
Clever	2	7	5.25	1.79	Adjective	2	6
Costume	1	3	3.61	4.57	Noun	1	7
Creaky	1	5	3.61	4.07	Adjective	1	6
Drop	1	6	4.89	4.21	Noun	2	4
Fairy	2	18	4.82	4.11	Noun	1	5
Fluffy	1	19	4.84	3.86	Adjective	1	6
Forest	2	3	4.75	4.76	Noun	2	6
Hairy	1	7	4.71	4.48	Adjective	1	5
Handbag	1	7	3.87	4.93	Noun	2	7
Kitten	2	3	4.11	4.86	Noun	2	6
Lead	1	3	4.85	4.21	Noun	1	4
Leaf	1	20	5.02	5	Noun	1	4
Mermaid	3	6	4.03	4.5	Noun	1	7
Mud	2	3	5.05	4.86	Noun	1	3
Pea	2	8	4.45	4.9	Noun	1	3
Pillow	1	3	4.3	5	Noun	1	6
Pleased	2	6	4.84	2.37	Adjective	1	7
Puddle	2	9	4.6	4.67	Noun	1	6
Sausage	2	4	4.55	4.88	Noun	1	7
Shell	1	6	4.93	4.8	Noun	1	5
Slipper	1	3	3.91	4.86	Noun	2	7
Sticky	2	4	5.27	3.59	Adjective	2	6

Stripy	1	3	4.43	4.72	Adjective	1	6
Suitcase	1	4	4.66	4.97	Noun	1	8
Track	2	7	4.94	4.31	Noun	2	5
Trolley	1	4	4.47	4.73	Noun	1	7
Useful	2	4	4.84	2.14	Adjective	1	6
Wand	1	5	4.49	4.73	Noun	1	4
Web	1	8	4.4	4.37	Noun	2	3
Wide	3	7	4.9	3.06	Adjective	1	4
Wing	2	3	4.42	4.86	Noun	2	4
Wobbly	2	18	4.84	3.15	Adjective	1	6

[a] Van Heuven, W., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190.
Values 1-3 are low-frequency words, 4-7 are high-frequency words.

[b] Brysbaert, M., Warriner, A. & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Method, 46*(3), 904–911.

[c] 1= Less consistent with L1 patterns/ 2= More consistent with L1 patterns.

2.4.8.1 Dictation

The dictation test intended to measure written-word form recall, which was expected to be learnt through learners' simultaneous processing of aural and written input (captions). Each target word was read in a sentence, so after listening to each sentence twice, students were asked to fill in the blanks (see Figure 20). These sentences were recorded by a native English speaker to create a video[12] that integrated all the prompts and ensured that all the groups did the test under the same conditions (e.g. time and number of repetitions). The answer sheet was available in pen-and-paper (see Appendix 33) and online[13] format. When the test was administered online (e.g. control group 1), time restrictions did not allow the participants to share or find the answers on the internet.

Two different scales were created to score learners' responses, a dichotomous scale (full knowledge scale, FKS), where correct answers were awarded one point, and a partial knowledge scale (PKS) that considered learners' improvement over time by awarding two points for correct answers and one point for partial knowledge (see section 4.1 for a detailed description). The Cronbach alpha coefficients (internal consistency) obtained for this instrument were high (.957 at pretest and .959 at posttest). This test was only administered at

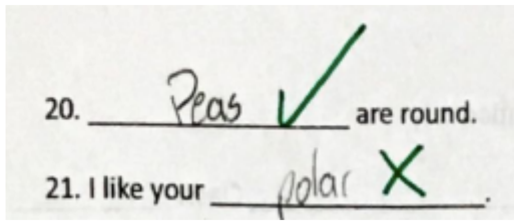
[12] https://drive.google.com/file/d/1e4NLissbWSEaM2jRSB9IIWZqgKEG0IXs/view?usp=sharing

[13] https://forms.gle/a7JGTzNuqp6djfKr5

pre and posttest since it was shown to be highly demanding. The students expected to answer all the questions and many of them claimed to be unsure of their answers. To keep students' willingness to participate in all the activities, we only administered the multiple-choice test below at delayed-posttest.

Figure 20.

Dictation test: Sample items



2.4.8.2 Multiple-choice test

This test assessed written-word form and meaning recognition. As shown in Figure 21, the target words were removed from each sentence and students were given a translation in L1-Spanish to figure out which alternative provided the missing target word. To prevent students from guessing the correct answer, they could select the 'I don't know' option. Since they had to read the questions on their own, 15 seconds were estimated per item, thus the time limit to take this test was 10 minutes. This test was available in pen-and-paper (see Appendix 34) and online[14] format (Google Forms), and in case of online test administration, the time constraints ruled out the possibility that the students' found the answers on the internet (e.g. using a dictionary).

This test was first administered to the online pilot-testing group on a one-to-one basis. While observing students' strategies to answer the questions, we realized that the most proficient participants first discarded the words they already knew to figure out the correct answers before selecting the 'I don't know' option. Thus, this aspect was improved by including English nonwords from the ARC Nonword Database (Rastle et al., 2002), as well as words that also appeared in the episodes as distractors. This test was administered to the control group 2 at pretest and posttest, and to the experimental groups at pretest, posttest and delayed-posttest.

[14] https://forms.gle/XmKSMFZDw7tf5XgR6

Figure 21.

Multiple-choice vocabulary test: Sample items

21. Technology is ______________________ (útil).

a) Smart
b) Useful
c) Fruft
d) Charming
e) No sé

22. Some men are ______________________ (peludos).

a) Hairy
b) veilful
c) Scary
d) Creepy
e) No sé

This test was scored dichotomously, that is correct answers were awarded one point and incorrect answers received zero points. The 'I don't know' option was considered as a wrong answer (0 points). The Cronbach alpha coefficients (internal consistency) obtained for this instrument were high: .897 at pretest, .905 at posttest, and .888 at delayed-posttest.

2.5 Procedures and treatment conditions

The investigation adopted a pretest-posttest-delayed posttest design to measure learners' L2 gains over time. Prior to the implementation, a letter was sent to the families of both schools to inform them about the investigation and get their consent (see Appendix 35). Even though the specific objectives of the investigation were not revealed, the researcher's e-mail address was indicated in case they required further information to authorize their child to participate in the activities. The message was sent through the different channels of information employed by the schools (web page, e-mails and Instagram) and the class delegates (WhatsApp groups). Additionally, parents were informed about the investigation once again in the first parents' meeting of the year. Along with parents' consent, learners' decision was also considered. On the whole, two students from school 2 were unwilling to participate in the activities.

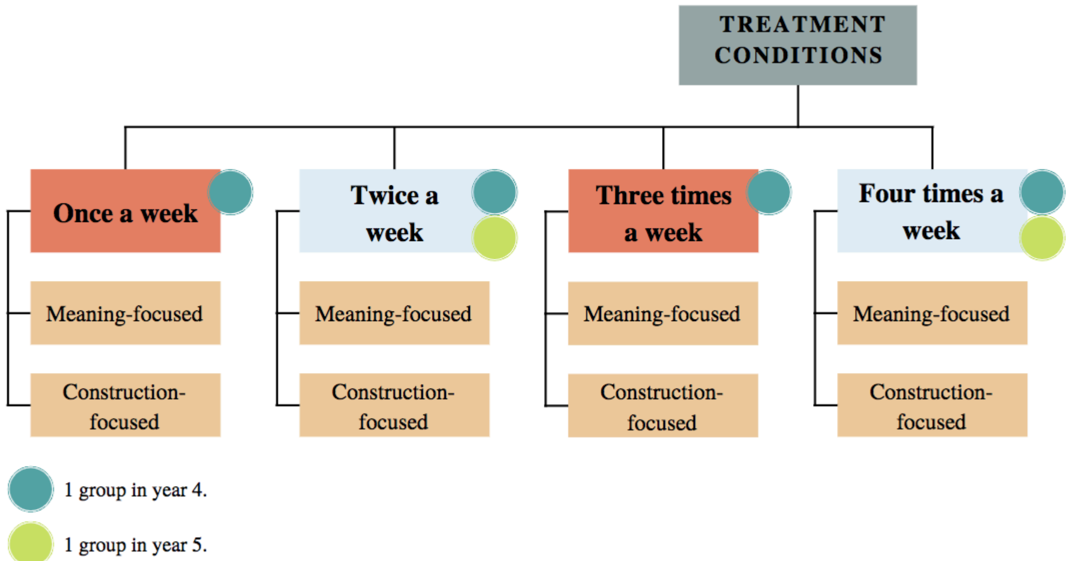
As mentioned earlier, data collection with the control groups was carried out before and after the main experiment (see section 2.2.2). The testing times were separated by five weeks, where the participants attended their regular English sessions. As for the experimental groups, both schools followed the same structure except for the number of episodes the students had to watch a week (viewing distribution), and the specific instruments administered at the three testing times (see section 2.2.1 and Figure 11). As explained earlier, the EFL teachers from school 2 played a more active role in the investigation under the supervision of this researcher (through Zoom or WhatsApp). Therefore, they participated in additional meetings and tasks to ensure the correct implementation of the treatment and the tests, which increased their workload. As a result, only some of the instruments were applied at school 2.

As noted above, this study was conducted onsite at both schools. However, due to the pandemic, most of the instruments had to be computerized and different actions had to be taken to ensure their reliability in case they had to be administered at any point of the investigation. The exception to the onsite implementation were the students that had to stay in quarantine for a few days. They were contacted through the school's official video conferencing application (Zoom or Microsoft teams) to complete the activities online (camera on). With respect to the treatment, the screen was shared with the students when a small group was online (1-3 students). With bigger groups (4-6 students), the episodes were uploaded to EdPuzzle (n.d.) since this platform allows researchers to supervise learners' actions. More precisely, on EdPuzzle: (1) Students cannot skip parts of the clip, (2) if students minimize the window or open a different tab, the video automatically stops, (3) the video owner has access to each student's progress bar. Additionally, learners' performance at the after-viewing comprehension questions (on Google Forms) was also checked as the episodes had to be fully watched to select the correct answers or a feasible distractor.

Concerning the treatment conditions, the groups differed as regards viewing distribution, and within each group, the participants were randomly assigned to a type of after-viewing activity (meaning-focused or construction-focused) (see Figure 22).

Figure 22.

Summary of the treatment conditions



2.5.1 Viewing distribution

The classes were randomly assigned to a viewing distribution group, which differed as regards the distance between the episodes (shorter vs. longer lags; see section 1.4.3), which is addressed as intersession interval (ISI) (see Table 19). The administration of the of the delayed-posttests took place either 14 or 21 days after the posttest, depending on their ISI and school's availability. Overall, intersession intervals were between 12 and 25% of the retention intervals, which are between the optimal ranges estimated by Rohrer & Pashler (2007) to assess retention (10-30%).

Table 19.

Viewing distribution

Class	School	Viewing distribution	Average ISI*	Delayed-posttest administration	ISI % from the retention interval
1-fourth	2	Once a week	7 days	21 days after posttest	23,3%
2-fourth	1	Twice a week	3,5 days	14 days after posttest	25%
3-fourth	1	Three days a week	2,33 days	14 days after posttest	16%
4-fourth	1	Four days a week	1,75 days	14 days after posttest	12%
2-fifth	1	Twice a week	3,5 days	14 days after posttest	25%
4-fifth	1	Four days a week	1,75 days	14 days after posttest	12%

*Intersession interval

2.6 Analyses

2.6.1 Quantitative analyses

Data analyses were conducted in SPSS v.25. A series of ANOVAs and T-tests were calculated in order to determine whether the groups were comparable as regards cognitive and language-related factors. Also, Pearson's correlations were performed to study the relationships between variables. The variables that were not normally distributed (Kolmogorov-Smirnov/Shapiro Wilk= $p<.05$) were square root (SQRT) transformed to conduct the analyses that required normally distributed continuous variables. In case these transformations did not lead to normal distribution, non-parametric analyses were selected.

A series of GLMMs (Generalized linear mixed models; linear models and binary logistic regressions), and multiple linear regressions were run to assess the influence of treatment, cognitive, and language-related factors on the outcomes over time. Thus, separate models were calculated for each group of variables (e.g. L2-related related factors). These analyses included Satterthwaite approximation and robust covariances, which are suggested for small sample groups and unbalanced data. The visual binning tool in SPSS was used to transform the continuous variables into categorical (equal percentiles) when their relationship with the target variable was not linear or the interactions between independent variables required graphs or further analyses to facilitate their interpretation. Prior to the calculation of

GLMMs and multiple linear regressions, we assessed normality of distribution and collinearity between variables (Tolerance > .3; VIF < 3.33). As regards GLMMs, model fit was estimated through AIC (Akaike Information Criterion). Therefore, the best fitting models were the ones that obtained lower AIC values. The GLMMs consisted of a compound-symmetry structure with student identification as subjects, and time as repeated measure. When studying vocabulary learning at item level, the models also included 'word' as repeated measure in order to more accurately assess learners' improvement over time.

As for viewing distribution, this variable was studied from two different perspectives. First of all, by using the variable class, which integrated viewing distribution and year level. In addition, by contrasting the groups from both year levels that watched either two or four episodes a week. In the latter analyses, 1-fourth and 3-fourth were excluded since they did not have a counterpart in year 5.

2.6.2 Qualitative analyses

All the data from the interviews was transcribed and translated. The resulting files were checked by a teacher from school 1 to guarantee that the translations conveyed exactly the same ideas. The data was coded inductively using NVivo 12 by following a series of recursive stages so as to ensure a more accurate processing of the qualitative data and their interpretations (Braun & Clarke, 2013). The coding process started with the familiarization with the data collected in the three interviews, which was coded and recoded based on their relevance to the research questions: What are students' perceptions of the viewing experience? How do their answers contribute to the interpretation of the quantitative findings?

In order to ensure a good thematic analysis, the following criteria drove the process (Braun & Clarke, 2013, p.287):

-Data have been analyzed – interpreted, made sense of – rather than just paraphrased or described.

-Analysis and data match each other – the extracts illustrate the analytic claims.

-Analysis tells a convincing and well-organized story about the data and topic.

III. Introduction to results and preliminary analyses

The results of this investigation are reported in seven main sections (see Table 20). Section 3 focuses on preliminary analyses (i.e. groups comparability and correlations between the measures on learners' characteristics). Then, sections 4 and 5 are devoted to vocabulary learning, more specifically *written-word form recall*, and *written-word form and meaning recognition*, respectively. Section 6 focuses on the development of L2 listening skills, and section 7 examines the development of English and Spanish reading efficacy. Finally, sections 8 and 9 focus on learners' perceptions of the viewing experience by presenting the outcomes of the questionnaire and the group interviews, respectively. Given that only some of the instruments were administered at school 2, Table 20 includes a summary of the groups included in each section.

Table 20.

Overview of the results sections

Section	Title	Groups included in the analyses	Discussion
III	Preliminary analyses.	**School 1:** 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth, control group 1, control group 2. **School 2:** 1-fourth.	✗
IV	Written-word form recall.	**School 1:** 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth, control group 1. **School 2:** 1-fourth.	✓
V	Written-word form and meaning recognition	**School 1:** 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth, control group 2. **School 2:** 1-fourth.	✓
VI	L2 listening skills	**School 1:** 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth, control group 1.	✓
VII	ER efficacy and SR efficacy	**School 1:** 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth, control group 2 (ER efficacy) and control group 1 (SR efficacy).	✓
VIII	Learners' perception of the treatment.	**School 1:** 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth. **School 2:** 1-fourth.	✓
IX	Interview on learners' perception of the treatment.	Three sample groups from **school 1**.	✓

3.1 Preliminary analyses

To start with, a series of ANOVAs and T-tests were run in order to determine whether the groups were comparable as regards cognitive and language-related factors at pretest (see descriptive statistics in Tables 21 and 22). Considering that the participants from school 2 only completed part of the instruments, Table 21 specifies which variables were tested in each group. Also, Pearson's correlations were performed to study the relationships between variables.

The following variables were square root (SQRT) transformed to improve the results obtained in the normality tests:

- EFL picture vocabulary test
- ER efficacy (pretest)
- SR efficacy (pretest)
- WM
- PSTM
- English text segmentation
- Spanish text segmentation

3.1.1 Comparisons between groups in terms of cognitive factors

3.1.1.1 Comparisons between year levels. The difference between year levels was not shown to be significant with respect to phonological short-term memory (t (107)= .686, p= .494, r= .06). However, fifth graders were found to score significantly higher than fourth graders in visual processing speed (t (91)= 4.291, p< .001, r= .40), and their difference in working memory approached statistical significance (t (107)= 1.935, p= .056, r= .18). Specifically, the values indicate that their difference was more evident in visual processing speed (see Tables 22 and 23).

3.1.1.2 Comparisons between classes. The differences between classes did not reach statistical significance for neither phonological short-term memory (F (5)= .476, p= .793, η^2 = .032) nor working memory (F (5)= 2.159, p= .064, η^2 = .095). However, an ANOVA test showed that the classes differed significantly in terms of visual processing speed (F (4)= 4.525, p=.002, η^2 = .171). Specifically, the Tukey post-hoc comparisons indicated that 2-fourth scored significantly lower than 2-fifth (p=.029) and 4-fifth (p=.031).

Table 21.

Descriptive statistics: Cognitive and language-related factors per group

	Class															
	1-fourth		2-fourth		3-fourth		4-fourth		2-fifth		4-fifth		CG1-fifth*		CG2-fourth*	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
PSTM	.	.	8.6	(1.7)	8.6	(1.0)	8.36	(1.4)	8.6	(1.6)	9.2	(1.8)	8.6	(2.7)	.	.
WM	.	.	8.56	(2.1)	8.6	(2.1)	8.29	(1.5)	8.6	(1.5)	9.1	(1.9)	10.1	(1.3)	.	.
Visual processing speed	.	.	35.4	(7.2)	36.7	(5.3)	36.1	(9.8)	42.6	(8.0)	42.6	(6.9)	.	.	.	.
Vocabulary knowledge	18.4	(10.2)	15.68	(7.0)	13.4	(6.35)	14.28	(5.4)	20.47	(12.6)	20.5	(8.9)	.	.	.	.
Segmentation in Spanish	26.9	(10.9)	24.3	(10.4)	22.6	(8.8)	25.3	(7.9)	36.1	(13.6)	37.3	(17.4)	.	.	.	.
Segmentation in English	23.2	(10.5)	21.2	(9.3)	18.8	(10.7)	21.1	(9.2)	37.0	(17.7)	42.2	(18.9)	.	.	.	.
SR efficacy	.	.	88.5	(54.6)	71.1	(23.8)	77.7	(36.0)	84.9	(37.3)	115.8	(58.9)	80.9	(27.2)	.	.
ER efficacy	.	.	55.4	(38.7)	36.3	(26.7)	42.9	(20.5)	66.6	(35.6)	79.4	(33.4)	.	.	50.05	(31.2)
L1 reading habits and attitude…	.	.	11.94	(5.48)	11.64	(5.54)	14.33	(5.24)	13.12	(4.17)	11.80	(5.76)	.	.	.	.
Listening skills	.	.	9.1	(3.4)	7.7	(3.6)	8.6	(3.1)	11.8	(5.0)	10.8	(4.0)	11.6	(3.1)	.	.

CG: Control group

Table 22.

Descriptive statistics: Cognitive and language-related factors per year level

	Year level			
	Year 4		*Year 5*	
	M	*(SD)*	*M*	*(SD)*
PSTM	8.52	(1.39)	8.81	(1.99)
WM	8.52	(1.95)	9.17	(1.7)
Visual processing speed	36.04	(7.38)	42.62	(7.38)
Vocabulary knowledge	15.82	(8.01)	20.5	(10.81)
Segmentation in Spanish	25.01	(9.68)	36.72	(15.46)
Segmentation in English	21.31	(9.94)	39.56	(18.26)
SR efficacy	79.15	(40.17)	94.85	(46.33)
ER efficacy	46.36	(30.43)	72.72	(34.81)
L1 reading habits and attitude…	12.64	(5.43)	12.45	(5.04)
Movers	8.46	(3.35)	11.38	(4.16)

Table 23.

Summary: Between-groups comparisons in terms of cognitive factors

Factor	Statistically sig. differences between year levels	Statistically sig. differences between classes
PSTM	-	-
Working memory	Year 4 < year 5*	-
Processing speed and attention	Year 4 < year 5	2-fourth < 2-fifth and 4-fifth

*Only approached statistical significance

3.1.2 Comparisons between groups in terms of language-related factors

3.1.2.1 Comparisons between year levels. The results demonstrated that year-5 scored significantly higher in all the measures except for L1 reading habits and attitudes towards reading (*t* (92)= .181, *p*= .857, *r*= .01): vocabulary knowledge (*t* (110)= 2.624, *p*= .010, *r*= .24), segmentation in Spanish (*t* (112)= 4.749, *p*< .001, *r*= .41), segmentation in English (*t* (111)= 6.010, *p*< .001, *r*= .49), listening skills (*t* (105)= 3.895, *p*< .001, *r*= .35), and ER efficacy (*t* (107)= 4.328, *p*< .001, *r*= .38). Yet, the difference between year levels in SR efficacy only approached significance (*t* (106)= 1.923, *p*= .057, *r*= .18).

3.1.2.2 Comparisons between classes. The between-groups comparisons in terms of vocabulary knowledge was only shown to approach significance (*F* (5)= 2.113, *p*= .069, η^2 = .091). Nevertheless, the classes differed significantly in terms of listening skills (*F* (5)=

3.361, p= .008, η^2 = .143), segmentation in English (F (5)= 7.587, p< .001, η^2 = .262) and segmentation in Spanish (F (5)= 4.702, p= .001, η^2 = .179). Specifically, the Tukey post-hoc comparisons indicated that only 2-fifth significantly outscored 3-fourth in terms of listening skills (p= .017). In relation to segmentation in English, the Tukey post-hoc comparisons showed that 2-fifth differed significantly from 3-fourth (p= .002), 4-fourth (p= .039), 2-fourth (p= .037), and 1-fourth (p= .044), while 4-fifth was also found to score significantly higher than all year-4 groups (p< .004). Looking at the post-hoc comparisons concerning segmentation in Spanish, we found that 2-fifth outperformed 3-fourth (p= .015), whereas 4-fifth was found to score significantly higher than 3-fourth (p= .011), 4-fourth, and 2-fourth (p= .040). The higher score obtained by 2-fifth in comparison with 2-fourth approached statistical significance (p= .053).

When comparing the groups in terms of reading efficacy, a set of ANOVAs revealed that their differences approached significance in Spanish (F (5)= 2.179, p= .062, η^2 = .097) and English (F (5)= 4.996, p<.001, η^2 = .195). The Tukey post hoc tests indicated that the lower score obtained by 3-fourth in comparison with 4-fifth in SR efficacy only approached statistical significance (p= .055), while in English, 2-fifth outscored 3-fourth (p= .024) significantly; and 4-fifth scored significantly higher than 3-fourth (p< .001), 4-fourth (p= .017) and CG2-fourth (p= .051). As for L1 reading habits and attitudes towards reading, the differences were not found to be significant (F (4)= .797, p= .530, η^2 = .035), which was already anticipated by the descriptive statistics (see Table 21). The results are summarized in Table 24.

Table 24.

Summary: Between-groups comparisons in terms of language-related factors

Factor	Statistically sig. differences between year levels	Statistically sig. differences between classes
Vocabulary knowledge	Year 5 > Year 4	4-fifth > 3-fourth
Segmentation in English	Year 5 > Year 4	2-fifth > 1,2,3,4-fourth 4-fifth > 1,2,3,4-fourth
Segmentation in Spanish	Year 5 > Year 4	2-fifth > 3-fourth, 2-fourth* 4-fifth > 2,3,4-fourth
Reading efficacy in English	Year 5 > Year 4	2-fifth > 3-fourth 4-fifth > 3,4-fourth, CG2-fourth
Reading efficacy in Spanish	Year 5 > Year 4*	4-fifth > 3-fourth

L1 reading habits and attitude…	-	-
Listening skills	Year 5 > Year 4	2-fifth > 3-fourth

*Approached statistical significance

In short, the between-groups comparisons confirm that the classes within each year level were linguistically comparable. However, as displayed in Table 21, 3-fourth students' appeared to have a slightly lower proficiency level (in L1 and L2), which is why this group emerged in all the comparisons that reached statistical significance. Likewise, 1-fourth's (school 2) scores were shown to be slightly higher in most of the measures this group completed, approaching fifth graders in vocabulary knowledge.

3.1.3 Relationships between cognitive and language-related factors

As displayed in Table 25, the cognitive factors were shown to be either weakly or non-significantly associated, confirming that these three tests measured three different constructs. In contrast to working memory, phonological short-term memory and visual processing speed were the cognitive factors that had a small but more significant relationship with language-related factors (see Table 25). With regard to language-related factors, all of them were found to correlate significantly, except for L1 reading habits and attitudes towards reading. Considering that an R^2 value of over 25% accounts for a large effect size (Larson-Hall, 2010), the shared variance between language-related factors ranged from small (e.g. 5% for Spanish segmentation, and L1 readings habits and attitude towards reading) to large (e.g. 46% for English segmentation and listening skills). It is interesting to note that English text segmentation turned out to have a strong relationship with the majority of the language-related factors.

Table 25.

Relationships between cognitive and language-related factors

		EFL PVT	Movers listening skills	SR efficacy	ER efficacy	Spanish segmentation	English segmentation	L1 reading habits and attitude…	PSTM	WM	Visual processing speed
EFL picture vocabulary test	Pearson Correlation	1	**.673** (R^2=.45)**	.467** (R^2=.21)	**.620** (R^2=.38)**	.378** (R^2=.14)	**.592** (R^2=.35)**	.105 (R^2=.01)	.290** (R^2=.08)	.215* (R^2=.04)	.272* (R^2=.07)
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.334	.006	.044	.010
	N	112	87	88	88	109	109	86	88	88	88
Movers listening skills	Pearson Correlation	**.673** (R^2=.45)**	1	**.512** (R^2=.26)**	.481** (R^2=.23)	.451** (R^2=.20)	**.683** (R^2=.46)**	.192 (R^2=.03)	.224* (R^2=.05)	.261** (R^2=.06)	.290** (R^2=.08)
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.072	.021	.007	.006
	N	87	107	106	90	87	86	89	106	106	90
SR efficacy	Pearson Correlation	.467** (R^2=.21)	**.512** (R^2=.26)**	1	**.523** (R^2=.27)**	**.515** (R^2=.28)**	**.531** (R^2=.28)**	.078 (R^2=.006)	.379** (R^2=.14)	.156 (R^2=.02)	.235* (R^2=.05)
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.465	.000	.107	.024
	N	88	106	108	92	89	88	90	108	108	92
ER efficacy	Pearson Correlation	**.620** (R^2=.38)**	.481** (R^2=.23)	**.523** (R^2=.27)**	1	.309** (R^2=.09)	**.508** (R^2=.25)**	.011 (R^2=.01)	.235* (R^2=.05)	.227* (R^2=.05)	.043 (R^2=.001)
	Sig. (2-tailed)	.000	.000	.000		.003	.000	.921	.024	.029	.684
	N	88	90	92	92	89	88	90	92	92	92
Spanish segmentation	Pearson Correlation	.378** (R^2=.14)	.451** (R^2=.20)	**.515** (R^2=.28)**	.309** (R^2=.09)	1	**.668** (R^2=.44)**	.237* (R^2=.05)	.391** (R^2=.15)	.225* (R^2=.05)	.327** (R^2=.10)
	Sig. (2-tailed)	.000	.000	.000	.003		.000	.026	.000	.033	.002
	N	109	87	89	89	114	113	88	90	90	90
English segmentation	Pearson Correlation	**.592** (R^2=.35)**	**.683** (R^2=.46)**	**.531** (R^2=.28)**	**.508** (R^2=.25)**	**.668** (R^2=.44)**	1	.170 (R^2=.02)	.405** (R^2=.16)	.283** (R^2=.08)	.361** (R^2=.13)
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.116	.000	.007	.001
	N	109	86	88	88	113	113	87	89	89	89
L1 reading habits and attitude…	Pearson Correlation	.105 (R^2=.01)	.192 (R^2=.03)	.078 (R^2=.006)	.011 (R^2=.01)	.237* (R^2=.05)	.170 (R^2=.02)	1	.078 (R^2=.006)	.109 (R^2=.01)	-.047 (R^2=.002)
	Sig. (2-tailed)	.334	.072	.465	.921	.026	.116		.463	.306	.661
	N	86	89	90	90	88	87	94	91	91	91
PSTM	Pearson Correlation	.290** (R^2=.08)	.224* (R^2=.05)	.379** (R^2=.14)	.235* (R^2=.05)	.391** (R^2=.15)	.405** (R^2=.16)	.078 (R^2=.006)	1	.208* (R^2=.04)	.230* (R^2=.05)
	Sig. (2-tailed)	.006	.021	.000	.024	.000	.000	.463		.030	.026
	N	88	106	108	92	90	89	91	109	109	93
WM	Pearson Correlation	.215* (R^2=.04)	.261** (R^2=.06)	.156 (R^2=.02)	.227* (R^2=.05)	.225* (R^2=.05)	.283** (R^2=.08)	.109 (R^2=.01)	.208* (R^2=.04)	1	-.043 (R^2=.001)
	Sig. (2-tailed)	.044	.007	.107	.029	.033	.007	.306	.030		.685
	N	88	106	108	92	90	89	91	109	109	93

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

IV. Written-word form recall

This section focuses on vocabulary learning from captioned-video viewing at the level of written-word form recall (also addressed as WWFR). Specifically, it attempts to answer the following research questions:

1) To what extent does viewing distribution (i.e. shorter vs. longer lags) influence young L2 learners' gains from captioned video viewing?

2) In comparison with meaning-focused activities, what are the effects of construction-focused after-viewing activities on L2 learning through captioned-video viewing?

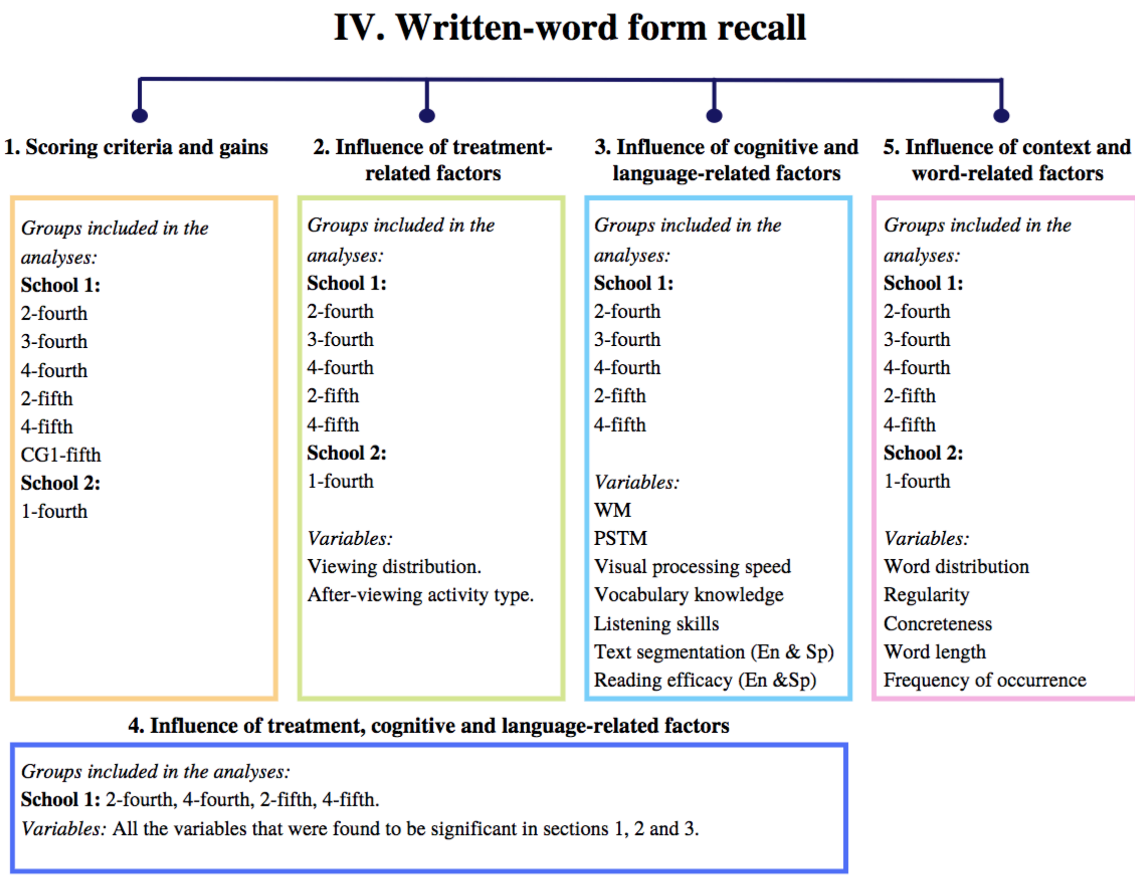
3) To what extent do learner characteristics influence young L2 learners' gains from captioned-video viewing? (i.e. age, vocabulary knowledge, cognitive abilities [phonological short-term memory, complex working memory, and visual processing speed], L1 and L2 reading skills [reading efficacy and text segmentation] and L2 listening skills).

4) To what extent do context and word-related factors (frequency of occurrence, regularity, word-length, and concreteness) influence vocabulary learning?

The overview of this section is displayed in Figure 23. As explained earlier, written-word form recall, measured by means of a dictation, was only tested at pretest and posttest.

Figure 23.

Section 4 overview



4.1 Scoring criteria

The evidence has demonstrated that written-word form recall and orthographic patterns are challenging aspects to acquire in terms of vocabulary knowledge, especially in the case of young learners, whose L1 literacy skills are still developing (Birch and Fulop, 2021). This was certainly true in the case of the participants of this study. The data collected through the dictation tests showed evidence of learners' poor (or nonexistent) knowledge of phoneme-grapheme correspondence and orthographic patterns in English. Learners' attempts to write unknown words appeared to have been influenced by their L1 phoneme-grapheme correspondence rules, which are mostly one-to-one (Sun-Aperin & Wang, 2008). The fact that some students wrote, for instance, 'flafi' (instead of fluffy) or 'cabich' (in place of cabbage) is in line with the literature on the area, which notes that typological differences concerning phonological processing and orthographic depth influence their performance and learning process in the L2 (Papp, 2020). Specifically, L2 learners of English have been found to struggle with reading and spelling due to their many-to-many sound-letter correspondences (Kormos, 2017), together with the use of double consonants, diphthongs and clusters that are not part of the inventory in their L1 (Papp, 2020).

Accordingly, it is expected that recalling written-word form representations of a language that has deep orthography takes time. Therefore, in this investigation, learners' accuracy in terms of written-word form recall was assessed through two different scales. The first one (full knowledge scale, FKS henceforth) was dichotomous and assigned one point to each 100% accurate response, whereas the second one (partial knowledge scale, PKS henceforth) was more sensitive to learners' improvement over time. Specifically, two points were given to each 100% accurate response while one point was awarded to those responses that approached the target form as a result of previous encounters (pretest) or the treatment (posttest). To this aim, a list of possible answers was specified by adapting the criteria used by Gesa (2019), which was, in turn, adapted from Muñoz (2006) (see Appendix 36).

The correlations between the two scales were strong and significantly high for pretest (rho= .965, p< .01, R^2= .93) and posttest (rho= .979, p< .01, R^2= .95), therefore, the majority of the analyses were performed by using the FKS at item level. Given that the PKS scores were not normally distributed, this variable was square root (SQRT) transformed to perform the analyses that required a normally distributed continuous variable.

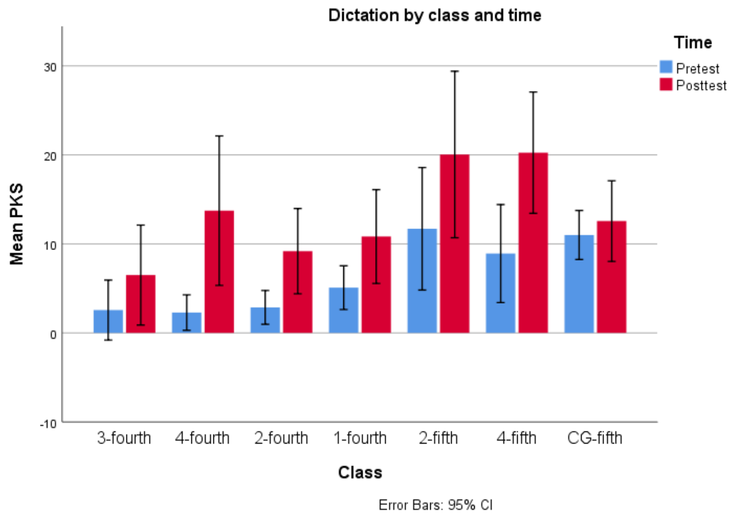
4.2 Written-word form recall: preliminary analyses

First of all, a set of between-groups comparisons were run with the aim of establishing whether the groups were comparable in terms of written-word form recall at pretest (PKS and FKS). As for the PKS scores (see Figure 24), a T-test showed that year-5 students outperformed fourth graders (t (129)= 5.440, p<.001, r=.42). Also, a One-Way ANOVA revealed that the differences between groups as regards pretest results were statistically significant (F (6)= 6.38, p<.001, η^2 = .236). The Tukey pairwise comparisons indicated that 3-fourth differed significantly from 2-fifth (p=.002), 4-fifth (p=.044) and CG1-fifth (p<.001), while 4-fourth knew significantly fewer words than 2-fifth (p=.013) and CG1-fifth (p=.002). As for 2-fourth, this group scored significantly lower than 2-fifth (p=.041) and

CG1-fifth (p=.007), whereas 1-fourth's scores did not differ significantly from fifth graders' outcomes.

Figure 24.

WWFR: Pretest and posttest scores (PKS)



With respect to the FKS scores (see Table 26), a Kruskal-Wallis test indicated that the differences between groups were also significant (H (6)=31.208, p< .001, η^2= .021). Specifically, the pairwise comparisons showed that 3-fourth scored significantly lower than 4-fifth (p=.012), 2-fifth (p=.003) and CG1-fifth (p<.001); while 4-fourth scored significantly lower than 4-fifth (p=.022), 2-fifth (p=.007) and CG1-fifth (p<.001); and 2-fourth was significantly outscored by 2-fifth (p=.041) and CG1-fifth (p<.001). In the case of 1-fourth, this group was only outperformed significantly by CG1-fifth (p=.003). A Mann-Whitney U test confirmed that year-5 students knew significantly more words than year-4 participants before the treatment (U=3089.5, z=4487, p< .001, r= .38). As summarized in Table 27, fifth graders scored significantly higher at pretest, and the between-groups comparisons indicated that within each year level, the classes were comparable with respect to previous knowledge.

Table 26.

WWFR: Descriptive statistics (FKS and PKS scores)

			Pretest		Posttest		Posttest gains		
			Mean	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*	*N*
FKS	Class	1-fourth	2.2	(2.6)	6.0	(6.3)	3.9	(4.3)	24
		2-fourth	1.3	(1.5)	3.6	(3.4)	2.5	(2.6)	16
		3-fourth	1.3	(3.0)	2.7	(4.8)	1.4	(2.0)	16
		4-fourth	.9	(1.5)	5.5	(5.6)	4.7	(4.8)	15
		2-fifth	4.9	(7.2)	8.8	(10.8)	4.1	(5.0)	24
		4-fifth	4.1	(6.2)	9.0	(7.8)	4.6	(4.9)	25
		CG1-fifth	4.7	(2.2)	5.6	(3.8)	.9	(2.4)	16
	Year level	Year 4	1.5	(2.3)	4.6	(5.3)	3.2	(3.8)	71
		Year 5*	4.5	(6.6)	8.9	(9.3)	4.3	(4.9)	65
PKS	Class	1-fourth	5.1	(5.8)	10.8	(12.2)	6	(8.1)	24
		2-fourth	2.9	(3.4)	9.2	(9)	6.3	(7.1)	16
		3-fourth	2.6	(6.3)	6.5	(10.5)	3.9	(4.7)	16
		4-fourth	2.3	(3.5)	13.7	(15.2)	11.9	(13.6)	15
		2-fifth	11.7	(15.9)	20	(22.2)	8.8	(9.3)	24
		4-fifth	8.9	(12.7)	20.2	(16.1)	10.6	(9.9)	25
		CG1-fifth	11	(5.2)	12.6	(8.5)	1.6	(4.8)	16
	Year level	Year 4	3.4	(5.1)	10.1	(11.9)	6.8	(8.9)	71
		Year 5*	10.3	(14.3)	20.1	(19.2)	4.3	(9.6)	65

*Without considering the control group

Table 27.

Summary: Between-groups comparisons in terms of PKS and FKS pretest scores

Factor	Statistically sig. differences between year levels	Statistically sig. differences between classes
PKS pretest scores	Year 5 > Year 4	3-fourth < 2-fifth, 4-fifth, CG1-fifth 4-fourth < 2-fifth, CG1-fifth 2-fourth < 2-fifth, CG1-fifth
FKS pretest scores	Year 5 > Year 4	3-fourth < 2-fifth, 4-fifth, CG1-fifth 4-fourth < 2-fifth, 4-fifth, CG1-fifth 2-fourth < 2-fifth, CG1-fifth 1-fourth < CG1-fifth

In addition, Spearman rho correlations were run in order to explore the relationships between learners' outcomes at pretest and posttest, and the factors assessed for the purpose of this study (cognitive and language-related factors) (see Table 28). The outcomes obtained when using FKS and PKS scale were highly similar, this is why only FKS's correlations are reported. The results revealed stronger relationships between learners' outcomes as regards written-word form recall and language-related factors, especially L2 vocabulary knowledge, segmentation in English, ER efficacy and listening skills.

Table 28.

Correlations between written-word form recall (pretest and posttest) and factors (cognitive and language-related)

		FKS pretest score	FKS posttest score
FKS posttest score	Correlation coeff.	**,775** (R^2 = .58)**	1,000
	Sig. (2-tailed)	,000	.
	N	113	134
PSTM	Correlation coeff.	,299**(R^2 = .08)	,365**(R^2 = .13)
	Sig. (2-tailed)	,004	,000
	N	90	92
WM	Correlation coeff.	,326**(R^2 = .10)	,283**(R^2 = .08)
	Sig. (2-tailed)	,002	,000
	N	90	92
Visual processing speed	Correlation coeff.	,128 (R^2 = .01)	,167 (R^2 = .02)
	Sig. (2-tailed)	,228	,111
	N	90	92
Vocabulary knowledge	Correlation coeff.	**,612**(R^2 = .37)**	**,639**(R^2 = .40)**
	Sig. (2-tailed)	,000	,000
	N	110	110
Segmentation in Spanish	Correlation coeff.	,460**(R^2 = .21)	**,511**(R^2 = .26)**
	Sig. (2-tailed)	,000	,000
	N	111	112
Segmentation in English	Correlation coeff.	**,559**(R^2 = .31)**	**,677**(R^2 = .45)**
	Sig. (2-tailed)	,000	,000
	N	110	111
SR efficacy	Correlation coeff.	**,510**(R^2 = .26)**	,479**(R^2 = .22)
	Sig. (2-tailed)	,000	,000
	N	89	91

ER efficacy	Correlation coeff.	**,598******(R^2 = .35)**	**,557******(R^2 = .31)**
	Sig. (2-tailed)	,000	,000
	N	89	91
Listening skills	Correlation coeff.	**,680******(R^2 = .46)**	**,690******(R^2 = .47)**
	Sig. (2-tailed)	,000	,000
	N	90	90

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

4.3 Written-word form recall: Progress over time

In order to compare the trajectories of both year levels, we ran a compound symmetry structure GLMM (binary logistic regression) with student identification as subjects, as well as time and word as repeated measures. Binary logistic regressions require a dependent measure that includes dichotomous values (e.g. Accuracy: 0= incorrect, 1= correct), as well as independent variables to determine the extent to which they predict learners' performance (Eddington, 2015). Thus, FKS values at item level were entered into this model as target variable, while time, year level and their interaction were included as factors. In this model, the scores obtained by the control group were not included. The results yielded significant effects for year level (F (1,91)= 13.847, p< .001), time (F (1,3126)= 124.216, p< .001), and a significant interaction between year level and time (F (1,3126)= 4.275, p= .039) (see Table 29). As shown in Table 30 and Figure 25, both year levels improved significantly from pretest to posttest. However, fifth graders scored significantly higher at both testing times (p<.05) and obtained greater gains from the treatment.

Table 29.

WWFR: The influence of time and year level

Source	F	df1	df2	Sig.
Corrected Model	49.135	3	342	.000
Level	13.847	1	91	.000
Time	124.216	1	3126	.000
Level * Time	4.275	1	3126	.039

Probability distribution: Binomial
Link function: Logit
a. Target: Binary scores

Table 30.

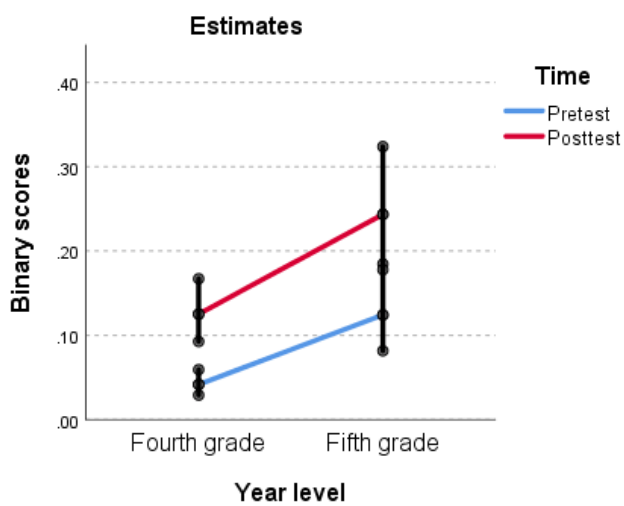
WWFR: Time comparisons per year level over time

Year level	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Fourth grade	Pretest - Posttest	-.084	.014	-5.816	523	.000	-.112	-.055
Fifth grade	Pretest - Posttest	-.119	.019	-6.129	441	.000	-.158	-.081

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 25.

WWFR: The trajectory of each year level over time



Then, separate analyses were calculated to assess the performance of each class over time. To start with, a Generalized Linear Mixed Model (GLMM; linear model) with repeated measures (time) compound-symmetry structure was run. The first model was calculated with PKS scores as the target variable, which included the two total scores obtained by each participant (pretest and posttest). The fixed effects included in the analysis were class, time, and their interaction (see Table 31). The results showed significant main effects for class (F (6,137)= 5.135, p< .001) and time (F (1,73)= 172.267, p< .001). Their interaction was also found to be significant (F (6,109)= 7.131, p< .001). Specifically, the Bonferroni adjusted results revealed that all the groups improved significantly from pretest to posttest (p< .001), except for the control group (p= .775) (see Table 32).

Table 31.

WWFR over time under each treatment condition (PKS)

Source	F	df1	df2	Sig.
Corrected Model	22,975	13	179	,000
Class	5,135	6	137	,000
Time	172,267	1	73	,000
Class* Time	7,131	6	109	,000

Probability distribution: Normal
Link function: Identity
a. Target: PKS scores

Table 32.

WWFR: Time comparisons per group over time (PKS)

Class	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-fourth	Pretest - Posttest	-1,220	,217	-5,632	251	,000	-1,647	-,793
4-fourth	Pretest - Posttest	-2,191	,409	-5,357	28	,000	-3,029	-1,352
2-fourth	Pretest - Posttest	-1,444	,198	-7,283	251	,000	-1,834	-1,053
1-fourth	Pretest - Posttest	-1,032	,225	-4,581	112	,000	-1,479	-,586
2-fifth	Pretest - Posttest	-1,106	,197	-5,625	187	,000	-1,493	-,718
4-fifth	Pretest - Posttest	-1,585	,225	-7,056	125	,000	-2,029	-1,140
CG1-fifth	Pretest - Posttest	-,057	,199	-,286	251	,775	-,448	,334

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

After that, in order to control for previous knowledge and confirm the outcomes above, we ran a compound symmetry structure GLMM (binary logistic regression) with student identification as subjects, as well as time and word as repeated measures. The dichotomous values (FKS at item level) were entered into the model as target variable, while time, class and their interaction were included as factors. As shown in Table 33, results confirmed a significant main effect for time (F (1,1353)= 532.119, p< .001) and class (F (6,131)= 74.613, p< .001), and a significant interaction between class and time (F (6,1217)= 101.448, p< .001). The Bonferroni pairwise comparisons demonstrated once more that all the classes improved from pretest to posttest (p<.001), except for the control group (p=.104).

Table 33.

WWFR over time by class (FKS at item level)

Source	F	df1	df2	Sig.
Corrected Model	618,407	13	340	,000
Time	532,119	1	1353	,000
Class	74,613	6	131	,000
Time * Class	101,448	6	1217	,000

Probability distribution: Binomial
Link function: Logit
a. Target: FKS scores by items

In addition, a Kruskal-Wallis test was run to compare posttest absolute gains (=PKS posttest score – PKS pretest score) between classes (H (6)=17.432, p= .008, η^2= .101). The results indicated that the control group's gains were significantly lower than 2-fourth (p=.041), 2-fifth (p=.006), 4-fourth (p=.002), and 4-fifth's (p<.001) outcomes. Also, 4-fifth, scored significantly higher than 3-fourth (p=.025), while 4-fourth outperformed 3-fourth significantly (p=.047) (see Figure 26).

Taken together, the results indicated that even when both year levels improved significantly over time, fifth graders obtained higher gains from the treatment. In addition, when assessing each group's trajectory from pretest to posttest, the results yielded significant gains in all the experimental groups, suggesting that the groups that watched four episodes a week benefitted more from the treatment (see Table 34). However, it is also important to note that learners' absolute and relative gains seemed to be limited (see Table 35), and there was high variability among participants in all the groups. The relative gains (Horst et al., 1998) were calculated by using the following formula: [Nº of words learnt / (Nº of words tested – Nº of words known at pretest)] x 100.

Figure 26.

WWFR: Groups' improvement over time

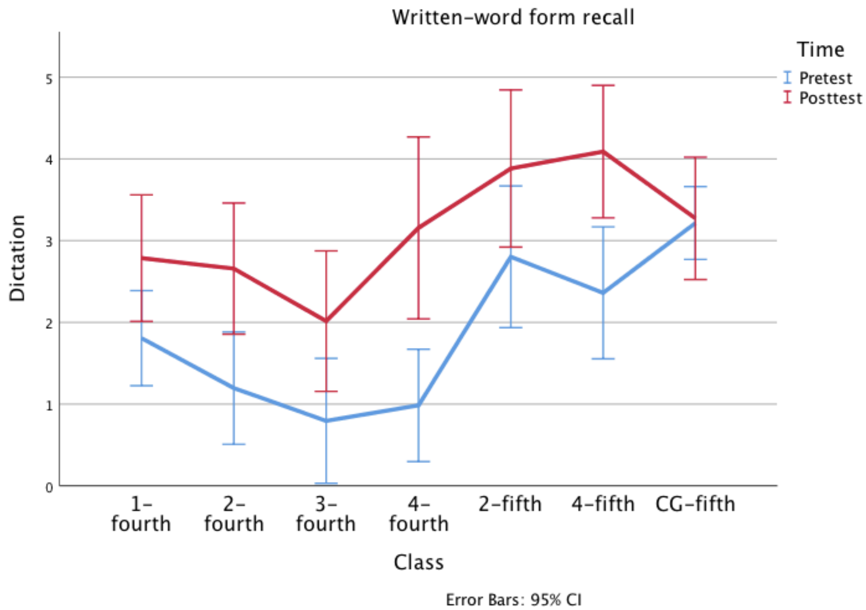


Table 34.

Summary of findings: Written-word form recall over time

Analysis	Outcome
Significant improvement from pretest to posttest (p<.05).	Both year levels. However, year 4 < year 5. All the experimental groups. 1-fourth, 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth.
Between-groups comparisons in terms of absolute gains (posttest-pretest)	CG < 2-fourth, 4-fourth, 2-fifth, 4-fifth 4-fifth > 3-fourth 4-fourth > 3-fourth

Table 35.

WWFR: Absolute and relative gains

	Minimum	Maximum	Mean (SD)
Absolute (N)	-2,00	23,00	3,75 (4,61)
Absolute gains (%)	-5,56	63,89	10,42 (12,81)
Relative gains (%)	-7,69	85,71	12,96 (17,97)

4.4 Written-word form recall: The influence of treatment-related factors

4.4.1 After-viewing activity type

In order to measure the influence of after-viewing activity type (see descriptive statistics in Table 36), a compound symmetry structure GLMM (binary logistic regression) was performed with student identification as subjects, as well as time and word as repeated measures. The dichotomous values (FKS at item level) were entered into the model as target variable, while activity type, year level, time and a triple interaction between these variables were included as factors. As shown in Table 37, the results showed non-significant effects for activity type (F (1, 81)= .682, p= .411), but statistically significant effects for time (F (1, 3560)= 135.238, p< .001), year level (F (1, 81)= 13.867, p< .001), as well as a significant interaction between activity type, year level and time (F (4, 393)= 2.403, p= .049). The Bonferroni pairwise comparisons (see Table 38) showed that in both year levels the use of construction-focused activities led to higher gains.

Table 36.

WWFR: Descriptive statistics per activity type

	FKS pretest		FKS posttest	
Activity type	M	(SD)	M	(SD)
Meaning-focused	3.19	(5.68)	6.60	(8.04)
Construction-focused	2.26	(3.73)	6.35	(7.68)

Table 37.

WWFR: The influence of activity type on learners' outcomes

Source	F	df1	df2	Sig.
Corrected Model	23.510	7	250	.000
Activity type	.682	1	81	.411
Level	13.867	1	81	.000
Time	135.238	1	3560	.000
Activity type * Level * Time	2.403	4	393	.049

Probability distribution: Binomial
Link function: Logit
a. Target: Binary scores

Table 38.

WWFR: Time pairwise contrasts by year level and activity type

Activity type	Year level	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Meaning-focused	Year 4	Pretest - Posttest	-.076	.016	-4.663	1295	.000	-.108	-.044
	Year 5	Pretest - Posttest	-.106	.025	-4.225	384	.000	-.156	-.057
Construction-focused	Year 4	Pretest - Posttest	-.091	.024	-3.833	262	.000	-.137	-.044
	Year 5	Pretest - Posttest	-.132	.029	-4.539	332	.000	-.189	-.075

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Then, a new model was built to further explore the influence of activity type in each year level. This time, vocabulary knowledge (EFL PVT) was included as a covariate. Again, a compound symmetry structure GLMM (binary logistic regression) was performed with student identification as subjects, as well as time and word as repeated measures. The dichotomous values (FKS at item level) were entered into the model as target variable, while activity type, year level, time, a triple interaction between these variables, and vocabulary knowledge were included as factors. The results indicated that this time the triple interaction between activity type, class and time only approached statistical significance (F (4, 495)= 2.173, p= .071) (see Table 39). However, a closer examination of the Bonferroni pairwise contrasts suggested that the fourth graders that completed construction-focused activities obtained greater gains from the treatment, whereas fifth graders obtained similar gains regardless of the activity type (see Table 40).

Table 39.

WWFR: The influence of activity type on learners' outcomes with vocabulary knowledge as covariate

Source	F	df1	df2	Sig.
Corrected Model	24.136	8	145	.000
Activity type	.094	1	103	.760
Level	3.298	1	28	.080
Time	114.051	1	139	.000
Activity type * Level * Time	2.173	4	495	.071
Vocabulary knowledge	56.800	1	13	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Binary scores

Table 40.

WWFR: Time pairwise contrasts by year level and activity type when including vocabulary knowledge as covariate

Activity type	Year level	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
Meaning-focused	Year 4	Pretest - Posttest	-.064	.018	-3.489	40	.001	-.101	-.027
	Year 5	Pretest - Posttest	-.099	.023	-4.371	691	.000	-.143	-.054
Construction-focused	Year 4	Pretest - Posttest	-.108	.025	-4.376	181	.000	-.157	-.059
	Year 5	Pretest - Posttest	-.114	.023	-4.992	394	.000	-.158	-.069

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

4.4.2 Viewing distribution

Separate models were calculated to assess the influence of viewing distribution in year 4 (1-4 episodes a week), and in the groups that watched either two or four episodes a week (in fourth and fifth grade). These models were fitted separately since 1-fourth and 3-fourth did not have a counterpart in year 5. To start with, a compound symmetry structure GLMM (binary logistic regression) was performed with fourth graders by fitting a model with student identification as subjects, as well as time and word as repeated measures. The dichotomous values (FKS at item level) were entered into the model as target variable, while viewing distribution, time and their interaction were included as factors. As shown in Table 41, the results showed significant effects for time (F (1, 2078)= 111.066, p< .001) and the interaction between viewing distribution and time (F (3, 862)= 4.805, p= .003). The Bonferroni pairwise contrasts in Table 42 indicated that all the groups improved significantly over time, but also suggested that 4-fourth obtained greater gains from the treatment. Yet, learners' gains did not seem to increase with the number of episodes watched a week, since 3-fourth obtained the lowest gains (see Figure 27). Considering that these results may also be influenced by each group's proficiency level, a new model was fitted with vocabulary knowledge (EFL PVT) as covariate (see Table 43).

Table 41.

WWFR: The influence of viewing distribution in fourth graders

Source	F	df1	df2	Sig.
Corrected Model	18.650	7	188	.000
Viewing distribution	.785	3	65	.507
Time	114.066	1	2078	.000
Viewing distribution * Time	4.805	3	862	.003

Probability distribution: Binomial
Link function: Logit[a]
a. Target: Binary scores

Table 42.

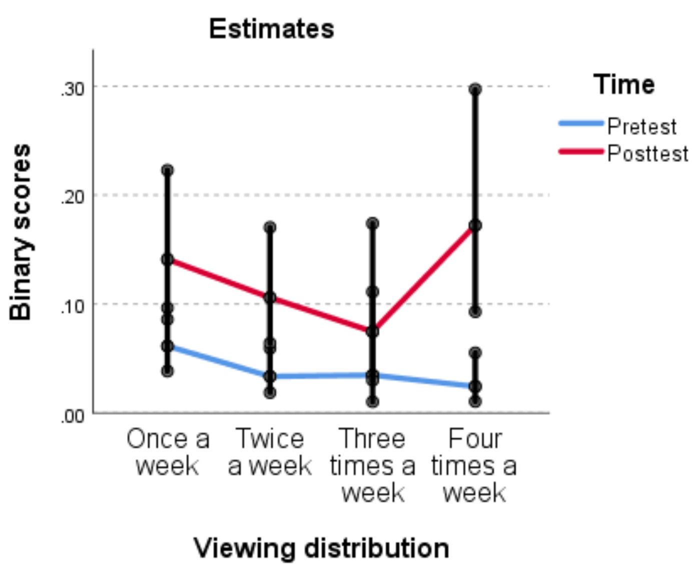
WWFR: Time pairwise contrasts by viewing distribution in fourth graders

Viewing distribution	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
Once a week	Pretest - Posttest	-.080	.024	-3.316	172	.001	-.127	-.032
Twice a week	Pretest - Posttest	-.072	.022	-3.334	662	.001	-.115	-.030
Three times a week	Pretest - Posttest	-.040	.013	-2.971	47	.005	-.067	-.013
Four times a week	Pretest - Posttest	-.148	.044	-3.349	91	.001	-.236	-.060

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 27.

WWFR: The trajectory of the fourth-grade viewing distribution groups over time



As shown in Table 43, the compound symmetry structure GLMM (binary logistic regression) that included vocabulary knowledge as covariate confirmed that, in fourth grade, watching four episodes a week was conducive to greater gains. In addition, the Bonferroni pairwise contrasts (see Table 44) suggested that watching less than four episodes a week did not affect learners' outcomes, implying the presence of a threshold (see Figure 28). Specifically, watching four episodes a week appeared to enhance learning.

Table 43.

WWFR: The influence of viewing distribution in fourth graders with vocabulary knowledge as covariate

Source	F	df1	df2	Sig.
Corrected Model	14.819	8	51	.000
Viewing distribution	.178	3	31	.910
Time	104.120	1	346	.000
Viewing distribution * Time	4.712	3	101	.004
Vocabulary knowledge	10.083	1	4	.030

Probability distribution: Binomial
Link function: Logit
a. Target: Binary scores

Table 44.

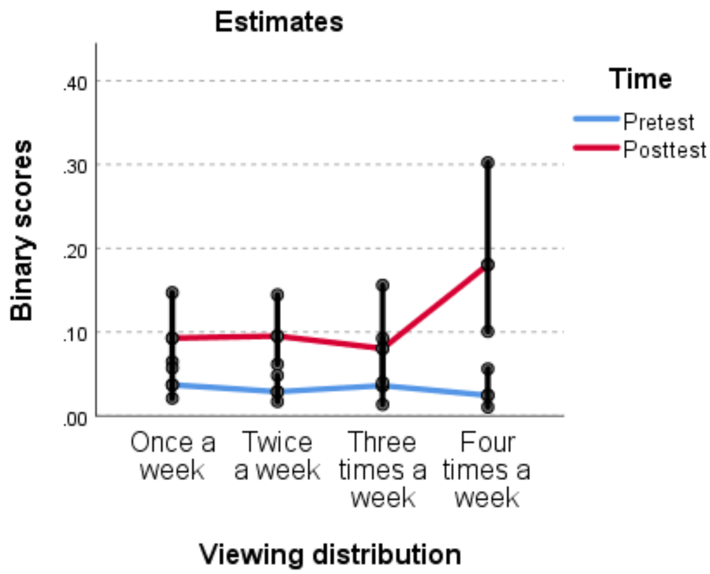
WWFR: Time pairwise contrasts per fourth-grade viewing distribution group with vocabulary knowledge as covariate

Viewing distribution	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Once a week	Pretest - Posttest	-.056	.018	-3.180	88	.002	-.091	-.021
Twice a week	Pretest - Posttest	-.066	.019	-3.548	520	.000	-.103	-.030
Three times a week	Pretest - Posttest	-.044	.012	-3.716	50	.001	-.068	-.020
Four times a week	Pretest - Posttest	-.156	.043	-3.636	45	.001	-.242	-.070

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 28.

WWFR: The trajectory of the fourth-grade viewing distribution groups over time with vocabulary knowledge as covariate



As mentioned earlier, a separate model was fitted to assess the influence of viewing distribution in fourth and fifth grade. Specifically, the groups that watched two and four episodes a week were included since only these time distributions were implemented in both year levels. To this aim, a compound symmetry structure GLMM (binary logistic regression) was performed with these groups by fitting a model with student identification as subjects, as well as time and word as repeated measures. The dichotomous values (FKS at item level) were included as target variable, while viewing distribution, year level, time and their interaction were entered as factors. The analyses revealed significant effects for level (F (1, 98)= 14.005, p< .001), time (F (1, 2743)= 111.706, p< .001) and the triple interaction between viewing distribution, year level and time (F (4, 424)= 4.623, p= .001) (see Table 45). As displayed in Table 46, watching four episodes a week seemed to lead to higher gains, however the difference between viewing distribution groups was more evident in year 4.

Table 45.

WWFR: The influence of viewing distribution in fourth and fifth graders (twice and four episodes a week)

Source	F	df1	df2	Sig.
Corrected Model	20.836	7	257	.000
Viewing distribution	.004	1	98	.952
Level	14.005	1	98	.000
Time	111.706	1	2743	.000
Viewing distribution * Level * Time	4.623	4	424	.001

Probability distribution: Binomial
Link function: Logit
a. Target: Binary scores

Table 46.

WWFR: Time pairwise contrasts by year level and viewing distribution groups

Viewing distribution	Year level	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Twice a week	Year 4	Pretest - Posttest	-.073	.022	-3.332	1327	.001	-.115	-.030
	Year 5	Pretest - Posttest	-.112	.028	-4.043	81	.000	-.167	-.057
Four times a week	Year 4	Pretest - Posttest	-.148	.044	-3.345	210	.001	-.235	-.061
	Year 5	Pretest - Posttest	-.127	.028	-4.607	1032	.000	-.181	-.073

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Then, to further explore the influence of viewing distribution in fourth and fifth graders, a new model was fitted by adding vocabulary knowledge (EFL PVT) as covariate. As shown in Table 47, the compound symmetry structure GLMM (binary logistic regression) yielded significant effects for the interaction between viewing distribution, year level and time (F (4, 244)= 2.457, p= .046), corroborating the results obtained in the previous analysis. The Bonferroni pairwise comparisons confirmed that fourth graders obtained greater gains when watching four episodes a week. Yet, in this analysis, 4-fourth's gains were magnified as a result of their great improvement despite their low vocabulary knowledge. With respect to fifth graders, they seemed to benefit from the treatment regardless of the number of episodes they had to watch a week (see Table 48).

Table 47.

WWFR: The influence of viewing distribution in fourth and fifth graders with vocabulary knowledge as covariate

Source	F	df1	df2	Sig.
Corrected Model	43.769	8	181	.000
Viewing distribution	.974	1	35	.331
Year level	.812	1	32	.374
Time	122.142	1	2727	.000
Viewing distribution * Level * Time	2.457	4	244	.046
Vocabulary knowledge	222.615	1	92	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Binary scores

Table 48.

WWFR: Time pairwise contrasts by year level and viewing distribution groups with vocabulary knowledge as covariate

Viewing distribution	Year level	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Twice a week	Year 4	Pretest - Posttest	-.076	.023	-3.346	330	.001	-.121	-.031
	Year 5	Pretest - Posttest	-.107	.022	-4.749	265	.000	-.151	-.062
Four times a week	Year 4	Pretest - Posttest	-.196	.052	-3.803	29	.001	-.301	-.091
	Year 5	Pretest - Posttest	-.105	.024	-4.403	530	.000	-.152	-.058

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

4.4.3 The influence of treatment-related factors: Summary of findings

This section assessed the influence of treatment-related factors (i.e. after-viewing activity type and viewing distribution) on vocabulary learning at the level of written-word form recall. The main findings are enlisted as follows:

- The analyses did not yield significant main effects for activity type nor viewing distribution.
- The implementation of construction-focused activities (intentional condition) led to higher gains specifically in the case of fourth graders.
- In fifth graders, both activity types seemed to result in similar gains (meaning-focused vs. construction-focused activities).

- Fourth graders appeared to be more sensitive to the effects of viewing distribution. More precisely, watching four episodes a week enhanced the recall of written-word forms.
- Fifth graders benefitted from the treatment regardless of the number of episodes they had to watch a week.

4.5 Written-word form recall: The influence of cognitive and language-related factors

This section focuses on the analyses that assessed the influence of cognitive and language-related factors on written-word form recall. In these analyses, only the experimental groups from school 1 were included (i.e. 2-fourth, 3-fourth, 4-fourth, 2-fifth and 4-fifth). It is also important to note that the presence of collinearity between variables was ruled out before running the tests to ensure that all the independent variables could be included in the analyses (Pallant, 2016).

4.5.1 Cognitive factors

In order to assess the influence of cognitive factors on written-word form recall, a Generalized Linear Mixed Model (binary logistic regression) with repeated measures (time) compound-symmetry structure was calculated. The model was built with learners' scores at pretest and posttest (FKS) by setting 36 (maximum score) as denominator. The fixed effects included in the analyses were class, time, PSTM, complex working memory, visual processing speed (high vs. low), and all possible two-way interactions. A backward (or step back) elimination procedure was used to determine the best fitted model. Thus, the non-significant interactions and factors were removed from the model one by one. In this case, the analyses indicated that neither working memory nor visual processing speed contributed to the learning process significantly (p >.05), therefore, only PSTM was kept in the best fitted model (see Table 49).

Table 49.

WWFR: Best fitted model obtained to assess the influence of cognitive factors on fourth and fifth graders' outcomes

Model Term	Coeff.	SE	t	Sig.	95% CI Lower	95% CI Upper	Exp (Coeff.)	95% CI for Exp(Coeff.) Lower	95% CI for Exp(Coeff.) Upper
Intercept	-3,105	,7692	-4,037	,000	-4,635	-1,575	,045	,010	,207
3-fourth	-1,214	,5588	-2,173	**,032**	-2,322	-,107	**,297**	,098	,899
4-fourth	-,148	,4490	-,331	**,742**	-1,048	,751	**,862**	,351	2,120
2-fourth	-,786	,3708	-2,120	**,035**	-1,518	-,054	**,456**	,219	,947
2-fifth	,209	,4015	,521	**,604**	-,594	1,013	**1,233**	,552	2,753
4-fifth	0^b	.	.	**.**	.	.	.	.	.
PSTM	,201	,0729	2,759	**,007**	,056	,346	**1,223**	1,058	1,414
$Time^+$	,465	,4522	1,029	**,305**	-,427	1,358	**1,593**	,652	3,888
[3-fourth]*[$Time^+$]	-,004	,3025	-,013	**,990**	-,601	,593	**,996**	,548	1,810
[4-fourth]*[$Time^+$]	-1,288	,3650	-3,529	**,001**	-2,009	-,567	**,276**	,134	,567
[2-fourth]*[$Time^+$]	-,273	,2931	-,930	**,354**	-,851	,306	**,761**	,427	1,358
[2-fifth]*[$Time^+$]	,108	,2360	,457	**,649**	-,360	,576	**1,114**	,698	1,778
[4-fifth]*[$Time^+$]	0^b	.	.	**.**	.	.	**.**	.	.
PSTM*[$Time^+$]	-,152	,0425	-3,569	**,000**	-,236	-,068	**,859**	,790	,935

Probability distribution: Binomial
Link function: Logit
a. Target: FKS score /36
b. This coefficient is set to zero because it is redundant.
+Category of reference: Posttest.

As presented in Table 50, the results revealed a statistically significant main effect for class (F (4, 104)= 3.253, p= .015), as well as significant interactions between class and time (F (4, 170)= 4966, p= .001), and PSTM and time (F (1, 151)= 12.741, p< .001). As regards PSTM, the exponential coefficient indicated that when PSTM scores increased by one, the odds of a correct response increased by 22%. In addition, learners' accuracy in written-word form recall increased by 2,74% at posttest per each additional point in PSTM[15] (forward digit span test) (see Table 49). In short, the contribution of PSTM to learners' improvement from pretest to posttest was found to be significant but relatively low.

[15] 1/(1,223-0,859) = 2,74%

Table 50.

WWFR: The influence of cognitive factors on FKS scores in fourth and fifth graders

Source	F	df1	df2	Sig.
Corrected Model	19,514	11	147	,000
Class	3,253	4	104	,015
PSTM	3,145	1	89	,080
Time	,199	1	170	,656
Class * Time	4,966	4	170	,001
PSTM * Time	12,741	1	151	,000

Probability distribution: Binomial
Link function: Logit
a. Target: FKS score/36

4.5.2 Language-related factors

In order to assess the influence of language-related factors on written-word form recall, a series of Generalized Linear Mixed Model (binary logistic regression) with repeated measures (time) compound-symmetry structure were calculated. The models were built with learners' scores at pretest and posttest (FKS) by setting 36 (maximum score) as denominator. The first model only included L2-related factors without interactions in order to calculate the approximate contribution of each factor to the learning of written-word forms. Initially, the predictor variables entered into the model were as follows: class, time, vocabulary knowledge, listening skills, English text segmentation, and ER efficacy. The step back elimination procedure indicated that ER efficacy did not contribute to the model significantly, therefore this factor was removed to obtain the best fitted model (see Table 51). The exponential coefficients indicated that vocabulary knowledge significantly predicted learning, increasing the odds of a correct response by 9,3% per each additional word known in the EFL picture vocabulary test. Likewise, when the Movers listening test score increased by one point, the possibility of writing words accurately increased by 11%. In addition, English text segmentation was found to be a significant albeit weak predictor of vocabulary learning at the level of written-word form recall. Specifically, the odds of a correct response in the dictation test increased by 1,6% per each additional word identified in the segmentation task. On the whole, listening skills and vocabulary knowledge accounted for greater variance in learners' performance as regards written-word form recall.

Table 51.

WWFR: Best fitted model obtained to assess the influence of L2-related factors on FKS scores (without interactions)

Model Term	Coeff	Std. Error	t	Sig.	95% CI Lower	95% CI Upper	Exp (Coeff)	95% CI for Exp(Coeff) Lower	95% CI for Exp(Coeff) Upper
Intercept	-5.023	.3599	-13.958	.000	-5.739	-4.307	.007	.003	.013
3-fourth	-.158	.3877	-.407	**.686**	-.938	.623	**.854**	.391	1.865
2-fourth	.702	.3652	1.922	**.067**	-.054	1.457	**2.017**	.948	4.294
2-fourth	-.212	.3485	-.609	**.544**	-.906	.482	**.809**	.404	1.619
2-fifth	-.231	.2116	-1.091	**.277**	-.649	.188	**.794**	.522	1.206
4-fifth	0^b	.	.	**.**	.	.	**.**	.	.
EFL PVT	.089	.0133	6.717	**.000**	.062	.116	**1.093**	1.064	1.123
English segmentation	.016	.0077	2.062	**.042**	.001	.031	**1.016**	1.001	1.032
Listening skills	.105	.0315	3.324	**.001**	.042	.168	**1.110**	1.043	1.182
Time+	-1.314	.1364	-9.634	**.000**	-1.590	-1.039	**.269**	.204	.354

Probability distribution: Binomial
Link function: Logit
a. Target: FKS /36
b. This coefficient is set to zero because it is redundant.
+ Category of reference: Posttest.

Then, a new model was built with class, time, vocabulary knowledge, segmentation in English, ER efficacy, listening skills, and all possible two-way interactions. By following a step back procedure, the analyses revealed that, again, ER efficacy did not contribute to the model significantly, therefore it was removed from the best fitted model (see Table 52).

Table 52.

WWFR: Best fitted model obtained to assess the influence of L2-related factors on FKS scores

Model Term	Coeff.	SE	t	Sig.	95% CI Lower	95% CI Upper	Exp (Coeff.)	95% CI for Exp(Coeff.) Lower	95% CI for Exp(Coeff.) Upper
Intercept	-5,360	,3429	-15,630	,000	-6,040	-4,679	,005	,002	,009
3-fourth	-,080	,3674	-,219	**,828**	-,823	,662	**,923**	,439	1,939
4-fourth	,933	,3678	2,537	**,020**	,165	1,701	**2,542**	1,180	5,480
2-fourth	-,092	,3735	-,247	**,806**	-,842	,657	**,912**	,431	1,930
2-fifth	-,157	,2088	-,751	**,454**	-,569	,256	**,855**	,566	1,292
4-fifth	0^b	.	.	**.**	.	.	**.**	.	.
Vocabulary knowledge	,094	,0127	7,391	**,000**	,068	,120	**1,099**	1,071	1,127
English segmentation	,021	,0078	2,710	**,008**	,006	,037	**1,021**	1,006	1,037
Listening skills	,102	,0318	3,198	**,002**	,038	,165	**1,107**	1,039	1,180

Time+	,122	,3040	,403	**,688**	-,478	,723	**1,130**	,620	2,061
[3-fourth]*[Time+]	-,344	,3204	-1,075	**,284**	-,978	,289	**,709**	,376	1,335
[4-fourth]*[Time+]	-1,658	,3802	-4,362	**,000**	-2,414	-,903	**,190**	,089	,405
[2-fourth]*[Time+]	-,645	,3795	-1,700	**,096**	-1,410	,119	**,525**	,244	1,126
[2-fifth]*[Time+]	-,155	,2618	-,591	**,557**	-,678	,369	**,857**	,507	1,446
[4-fifth]*[Time+]	0[b]	.	.	**.**	.	.	**.**	.	.
En. segmentation*[Time+]	-,026	,0059	-4,365	**,000**	-,037	-,014	**,975**	,964	,986

Probability distribution: Binomial
Link function: Logit
a. Target: FKS score/36
b. This coefficient is set to zero because it is redundant.
+ Category of reference: Posttest

As summarized in Table 53, the model showed a significant main effect of vocabulary knowledge (F (1, 40)= 54.629, p< .001), indicating that the odds of recalling a word form correctly increased by 9.9% per each correct answer at the Picture EFL Vocabulary test (see Table 52). The results also revealed a significant main effect for listening skills (F (1, 62)= 10.225, p= .002), showing that when students' score at the listening skills test increased by one, the odds of writing a target word correctly increased by 10.7%. Class also emerged as a statistically significant factor (F (1, 152)= 4.536, p= .035), but so did its interaction with time (F (4, 81)= 5.273, p= .001) and the interaction between English segmentation and time (F (1, 97)= 19.052, p< .001). As regards the class*time interaction, the Bonferroni pairwise comparisons indicated that the between-groups differences at each testing time did not reach statistical significance (p > .05) but, as observed in previous analyses, all the groups showed significant improvement over time (p < .011). Hence, this significant interaction implied that the groups differed as regards the extent to which they benefitted from the treatment. With respect to the significant interaction between English segmentation and time, the odds of writing words accurately increased by 21,73% at posttest per each additional word identified at the English text segmentation task[16] (see Table 51). Indeed, Figure 29 illustrates the stronger relationship between English text segmentation and written-word form recall at posttest.

[16] 1/(1.021-0,975) = 21,73%

Table 53.

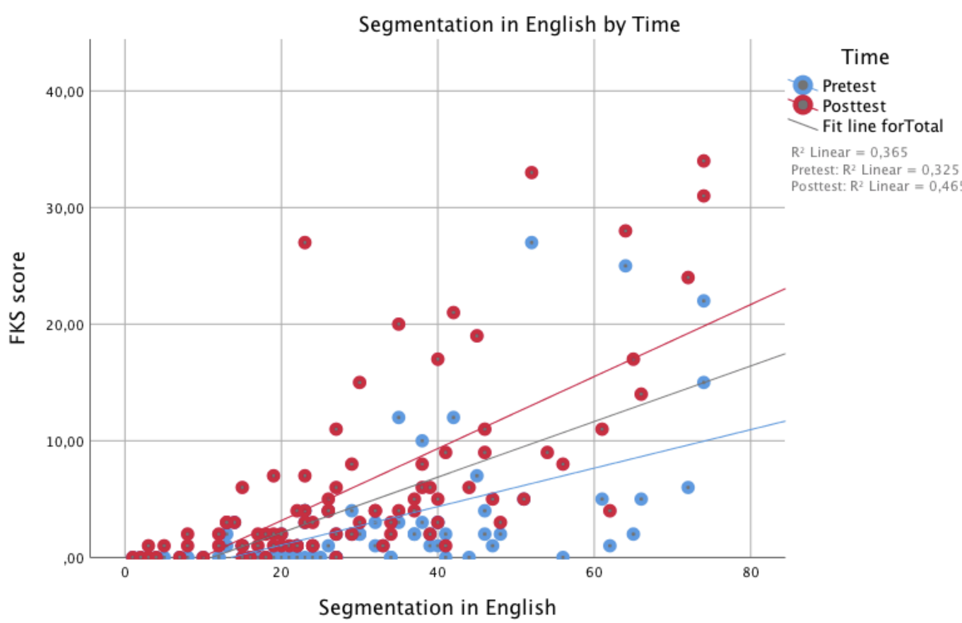
WWFR: The influence of L2-related factors on FKS scores

Source	F	df1	df2	Sig.
Corrected Model	35,475	13	68	,000
Class	,707	4	37	,593
Vocabulary knowledge	54,629	1	40	,000
Segmentation in English	1,249	1	86	,267
Listening skills	10,225	1	62	,002
Time	4,536	1	152	,035
Class * Time	5,273	4	81	,001
English segmentation * Time	19,052	1	97	,000

Probability distribution: Binomial
Link function: Logit
a. Target: FKS score /36

Figure 29.

WWFR: Interaction between English segmentation and time



Then, a new model was built to assess the influence of L1-related factors on vocabulary learning by entering the variables as follows: class, time, SR efficacy and L1 text segmentation. No interactions were included to calculate the approximate contribution of each factor to learners' outcomes. As displayed in Table 54, Spanish segmentation was shown to be a marginally significant albeit weak predictor of written-word form recall (*p*= .057), while SR efficacy was only found to be marginally significant (*p*=.073).

Table 54.

WWFR: The influence of L1-related factors on FKS scores (without interactions)

Model Term	Coef	Std. Error	t	Sig.	95% CI Lower	95% CI Upper	Exp (Coef)	95% CI for Exp(Coef) Lower	95% CI for Exp(Coef) Upper
Intercept	-2.918	.5895	-4.951	.000	-4.082	-1.755	.054	.017	.173
3-fourth	-.662	.5266	-1.258	**.210**	-1.702	.377	**.516**	.182	1.458
4-fourth	.548	.5313	1.032	**.305**	-.506	1.603	**1.730**	.603	4.967
2-fourth	-.383	.4588	-.835	**.405**	-1.289	.523	**.682**	.276	1.687
2-fifth	.173	.4305	.403	**.688**	-.680	1.027	**1.189**	.507	2.792
4-fifth	0^b	.	.	**.**	.	.	**.**	.	.
Time⁺	-1.013	.1340	-7.562	**.000**	-1.280	-.747	**.363**	.278	.474
Spanish segmentation	.022	.0118	1.913	**.057**	-.001	.046	**1.023**	.999	1.047
SR efficacy	.006	.0033	1.804	**.073**	-.001	.013	**1.006**	.999	1.013

Probability distribution: Binomial
Link function: Logit
a. Target: FKS /36
b. This coefficient is set to zero because it is redundant.
+ Category of reference: Posttest.

Next, a new model was fitted to study the effects of L1-related factors in comparison with the L2 variables that were found to be significant in the analyses above. The following independent variables were entered into the model: class, time, vocabulary knowledge, English segmentation, listening skills, Spanish segmentation, SR efficacy and all possible two-way interactions. By following a step back procedure to obtain the best fitted model (see Table 55), Spanish segmentation was removed as it was not shown to contribute to the model significantly. As summarized in Table 56, the GLMM revealed significant main effects for vocabulary knowledge (F (1, 31)= 50.935, p< .001), listening skills (F (1, 59)= 9.365, p= .003), time (F (1, 147)= 4.599, p= .034), and Spanish reading efficacy (F (1, 26)= 5.272, p= .030). However, the latter was found to interact with class significantly (F (4, 43)= 10.597, p< .001), confirming that 2-fourth and 3-fourth relied more on SR efficacy to benefit from the treatment (see Figure 30). Once again, the model revealed statistically significant interactions between class and time (F (4, 89)= 5.160, p= .001), as well as English segmentation and time (F (1, 113)= 18.602, p< .001), which were already reported in previous models.

Table 55.

WWFR: Best fitted model obtained to assess the influence of L1 and L2-related factors on FKS scores

					95% CI		Exp	95% CI for Exp(Coeff.)	
Model Term	Coeff.	SE	t	Sig.	Lower	Upper	(Coeff.)	Lower	Upper
Intercept	-4,926	,3631	-13,566	,000	-5,643	-4,208	,007	,004	,015
3-fourth	-3,104	,6524	-4,758	**,000**	-4,393	-1,815	**,045**	,012	,163
4-fourth	,899	,6029	1,491	**,142**	-,308	2,106	**2,457**	,735	8,216
2-fourth	-,295	,4631	-,638	**,525**	-1,211	,620	**,744**	,298	1,859
2-fifth	-1,007	,3566	-2,823	**,005**	-1,712	-,302	**,365**	,181	,739
4-fifth	0^b	.	.	**.**	.	.	**.**	.	.
Vocabulary knowledge	,089	,0124	7,137	**,000**	,063	,114	**1,093**	1,065	1,121
English segmentation	,023	,0081	2,904	**,005**	,007	,040	**1,024**	1,007	1,040
SR efficacy	-,003	,0023	-1,313	**,192**	-,008	,002	**,997**	,992	1,002
Listening skills	,098	,0322	3,060	**,003**	,034	,163	**1,103**	1,035	1,177
Time$^+$	,103	,2994	,344	**,732**	-,489	,695	**1,108**	,613	2,003
SR efficacy*[3-fourth]	,033	,0053	6,170	**,000**	,022	,044	**1,034**	1,023	1,044
SR efficacy*[4-fourth]	-,001	,0081	-,109	**,914**	-,018	,017	**,999**	,982	1,017
SR efficacy*[2-fourth]	,002	,0038	,504	**,616**	-,006	,009	**1,002**	,994	1,009
SR efficacy*[2-fifth]	,008	,0031	2,734	**,007**	,002	,014	**1,008**	1,002	1,015
SR efficacy*[4-fifth]	0^b	.	.	**.**	.	.	**.**	.	.
[3-fourth]*[Time$^+$]	-,409	,3130	-1,308	**,193**	-1,028	,210	**,664**	,358	1,233
[4-fourth]*[Time$^+$]	-1,647	,3784	-4,352	**,000**	-2,395	-,898	**,193**	,091	,407
[2-fourth]*[Time$^+$]	-,629	,3737	-1,684	**,099**	-1,381	,122	**,533**	,251	1,130
[2-fifth]*[Time$^+$]	-,124	,2504	-,497	**,621**	-,624	,376	**,883**	,536	1,456
[4-fifth]*[Time$^+$]	0^b	.	.	**.**	.	.	**.**	.	.
English segmentation*[Time$^+$]	-,025	,0059	-4,313	**,000**	-,037	-,014	**,975**	,964	,986

Probability distribution: Binomial
Link function: Logit
a. Target: FKS score/36
b. This coefficient is set to zero because it is redundant.
+ Category of reference: Posttest.

Table 56.

WWFR: the influence of L1 and L2-related factors on FKS scores

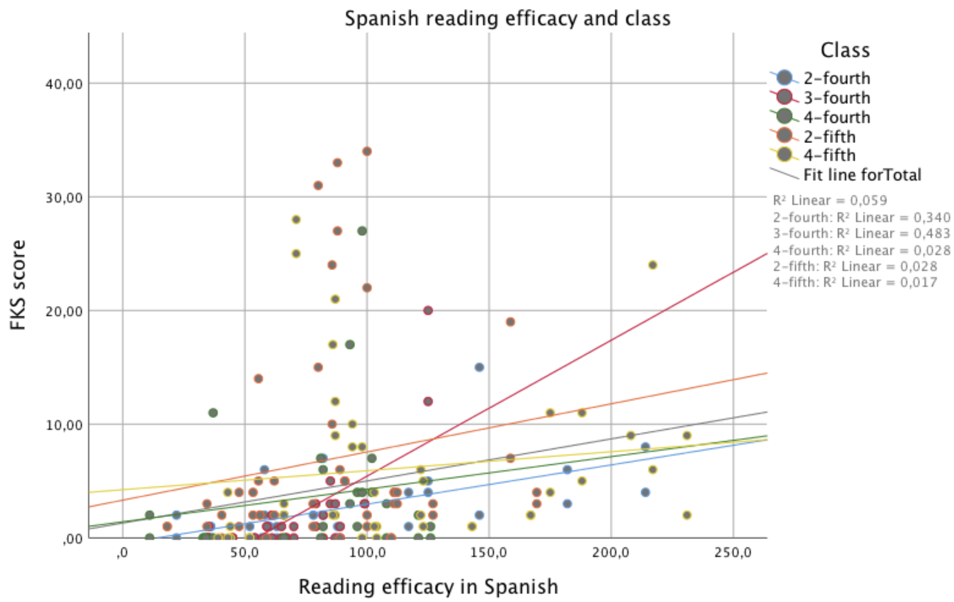
Source	F	df1	df2	Sig.
Corrected Model	38,286	18	99	,000
Class	7,707	4	142	,000
Vocabulary knowledge	50,935	1	31	,000
English segmentation	1,891	1	50	,175
SR efficacy	5,272	1	26	,030
Listening skills	9,365	1	59	,003
Time	4,599	1	147	,034
Class * SR efficacy	10,597	4	43	,000
Class * Time	5,160	4	89	,001
English segmentation * time	18,602	1	113	,000

Probability distribution: Binomial
Link function: Logit
a. Target: FKS score/36

In summary, L2 vocabulary knowledge and listening skills were found to be stronger predictors of vocabulary learning at the level of written-word form recall. As for SR efficacy, this factor seemed to be more relevant in the case of the fourth graders that had to watch two or three episodes a week. As suggested in Figure 30, L1 reading skills did not play a prominent role in 4-fourth and fifth graders' performance. Thus, the extent to which SR efficacy influenced the outcomes might have depended on learners' age and proficiency level, as well as the number of episodes they had to watch a week.

Figure 30.

WWFR: Interaction between Spanish reading efficacy and class



4.5.3 The influence of cognitive and language-related factors: Summary of findings

Overall, the results reported in this section indicate that:

- Among the cognitive factors, PSTM was the only variable that played a role in written-word form recall. Yet, its contribution did not seem to increase significantly from pretest to posttest.
- Within the group of language-related factors, the analyses indicated that vocabulary knowledge and listening skills were the strongest predictors of learners' progress over time. As for English text segmentation, its influence increased at posttest.
- The extent to which SR efficacy predicted learning seemed to change as a function of year level and viewing distribution. More precisely, the fourth graders that watched either two or three episodes a week appeared to rely significantly more on SR efficacy to benefit from the treatment.

4.6 Written-word form recall: The influence of treatment, cognitive and language-related factors

All the factors found to be significant in previous analyses were entered into a new model as independent variables to examine which ones remain as significant predictors. To this aim, we ran a series of compound-symmetry structure GLMMs (binary logistic regressions) with student identification as subjects, together with time and word as repeated measures. FKS score at item level (binomial distribution) was set as outcome variable, while the fixed factors selected for these analyses were as follows: viewing distribution, year level, activity type, time, vocabulary knowledge, listening skills, English segmentation, SR efficacy, PSTM and all possible two-way interactions between time and the rest of the factors. In these analyses, only the participants that watched two or four episodes a week were included since these were the only viewing time distributions that were implemented in both year levels.

The step back procedure indicated that SR efficacy and PSTM were no longer significant predictors of written-word form recall (see Table 57). Still, SR efficacy was unlikely to contribute to the model significantly if the group that relied on L1 reading skills the most had been excluded from these analyses (3-fourth). With respect to PSTM, previous analyses already anticipated that its contribution to the learning process was weaker in comparison with language-related factors. Therefore, its effects may have been overpowered by the rest of the factors.

Table 57.

WWFR: Best fitted model built to assess the influence of treatment, cognitive, and language-related factors

					95% CI		Exp	95% CI for Exp(Coeff.)	
Model Term	Coeff.	SE	t	Sig.	Lower	Upper	(Coeff.)	Lower	Upper
Intercept	-5,058	,3769	-13,420	,000	-5,803	-4,313	,006	,003	,013
Twice a week	-,540	,2291	-2,356	,027	-1,012	-,067	,583	,363	,935
Four times a week	0^b	.	.	.	.	.	.	.	.
Meaning-focused activity	-,464	,2447	-1,897	,067	-,964	,036	,629	,381	1,036
Construction-focused act.	0^b	.	.	.	.	.	.	.	.
Year 4	,562	,2790	2,014	,049	,004	1,120	1,754	1,004	3,066
Year 5	0^b	.	.	.	.	.	.	.	.
Vocabulary knowledge	,093	,0131	7,064	,000	,066	,119	1,097	1,068	1,127

Segmentation in English	,016	,0096	1,688	,100	-,003	,036	1,016	,997	1,036
Listening skills	,128	,0406	3,138	,004	,045	,210	1,136	1,046	1,234
Time+	-,287	,3164	-,907	,364	-,907	,333	,751	,404	1,396
[Meaning foc.]*[Time+]	,599	,2236	2,680	,007	,161	1,038	1,820	1,174	2,822
[Construction foc.]*[Time+]	0[b]	.	.	.	.	.	.	.	.
[Year 4]*[Time+]	-1,109	,2941	-3,769	,000	-1,685	-,532	,330	,185	,588
[Year 5]*[Time+]	0[b]	.	.	.	.	.	.	.	.
Eng. segmentation*[Time+]	-,025	,0058	-4,221	,000	-,036	-,013	,976	,965	,987

Probability distribution: Binomial
Link function: Logit
a. Target: Binary FKS score
b. This coefficient is set to zero because it is redundant.
+ Category of reference: Posttest.

As summarized in Table 58, the analyses revealed significant main effects for viewing distribution (F (1, 24)= 5.552, p= .027), vocabulary knowledge (F (1, 36)= 49.898, p< .001), listening skills (F (1, 32)= 9.849, p= .004) and time (F (1, 4813)= 5.327, p= .021). With respect to the significant interactions, the outcomes are similar to those reported in previous models. The results revealed significant interactions between activity type and time (F (1, 1615)= 7.181, p= .007), segmentation in English and time (F (1, 4813)= 17.814, p< .001), as well as year level and time (F (1, 1853)= 14.206, p<. 001). As for the latter interaction, the presence of language-related factors magnified the outcomes obtained by fourth graders, implying that year 4 showed great improvement from pretest to posttest despite their significantly lower proficiency level (see contrast estimates in Table 59). Similarly, the effects of viewing distribution were more clearly observed when the language-related factors were entered into the same model (see Table 60 and Figure 31). More precisely, the results suggest that shorter lags between episodes (4 episodes a week) may have moderated the effects of language-related factors to facilitate written-word form recall, especially in the case of 4-fourth. Taken together, these analyses seem to indicate that treatment and language-related factors were better predictors of written-word form recall in comparison with cognitive factors (see Table 61).

Table 58.

WWFR: The influence of treatment, cognitive and language-related factors on FKS scores

Source	F	df1	df2	Sig.
Corrected Model	42,293	10	121	,000
Viewing distribution	5,552	1	24	,027
Activity type	,548	1	25	,466
Level	,001	1	51	,975
Vocabulary picture	49,898	1	36	,000
Segmentation in English	,176	1	32	,678
Listening skills	9,849	1	32	,004
Time	5,327	1	4813	,021
Activity type * Time	7,181	1	1615	,007
Year level * Time	14,206	1	1853	,000
English segmentation * Time	17,814	1	4813	,000

Probability distribution: Binomial
Link function: Logit
a. Target: Item-level FKS score

Table 59.

WWFR: Bonferroni pairwise contrasts for the interaction between year level and time in a model that included L2-related factors.

		Pairwise Contrasts	Contrast Estimate	SE	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Time	Pretest	Fourth grade - Fifth grade	-,023	,012	-1,960	167	,052	-,046	,000
	Posttest	Fourth grade - Fifth grade	,070	,037	1,879	53	,066	-,005	,145
Year level	Year 4	Pretest - Posttest	-,151	,029	-5,211	67	,000	-,209	-,093
	Year 5	Pretest - Posttest	-,058	,012	-4,859	1568	,000	-,082	-,035

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 60.

WWFR: Bonferroni pairwise contrasts for viewing distribution groups in a model that included L2-related factors.

Viewing distribution Pairwise Contrasts	Contrast Estimate	SE	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Twice a week - Four times a week	-,040	,018	-2,243	21	,036	-,078	-,003

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 31.

WWFR: The effects of viewing distribution shown by a model that fitted language and treatment-related factors.

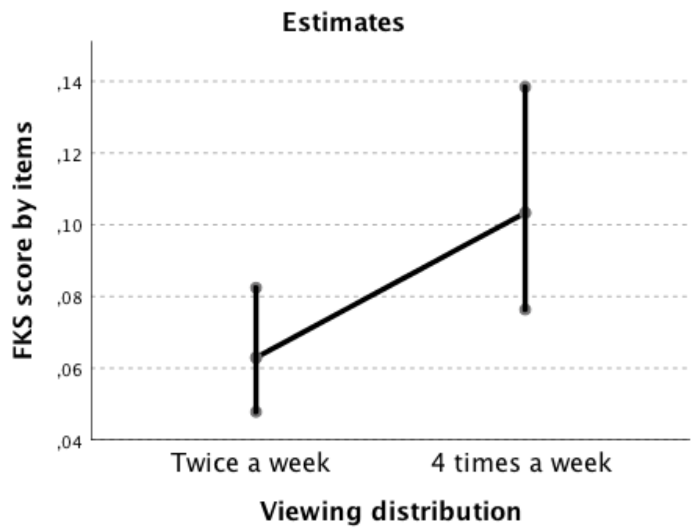


Table 61.

Summary: Predictors of written-word form recall

Factors	Outcomes
Viewing distribution	***Significant.*** All the viewing distribution groups improved from pretest to posttest significantly. However, shorter lags between episodes (i.e. watching 4 episodes a week) seemed to lead to greater gains, especially in year 4.
Year level	***Significant***. Both year levels improved from pretest to posttest significantly, and fifth graders obtained greater gains at posttest. However, the outcomes also indicated that year 4 (particularly 4-fourth) showed great improvement despite their significantly lower L2 proficiency level.
After-viewing activity type	***Significant.*** Both activity types were conducive to significant learning over time. Yet, the use of construction-focused activities seemed to foster greater gains, especially in fourth graders.
Time	***Significant.***
Vocabulary knowledge	***Significant.*** One of the strongest predictors of written-word form recall.
English segmentation	***Significant.*** It was a stronger predictor at posttest.
Spanish segmentation	***Non-significant.***
English reading efficacy	***Non-significant.***

Spanish reading efficacy	***Significant.*** The implementation of longer lags between episodes and a significantly lower L2-proficiency level resulted in 3-fourth and 2-fourth's greater reliance on SR efficacy to benefit from the treatment.
Listening skills	***Significant.*** One of the strongest predictors of written-word form recall.
Phonological short-term memory	***Statistically significant*** albeit weaker predictor of written-word form recall. Its influence was more prominent at posttest.
Working memory	***Non-significant.***
Visual processing speed	***Non-significant.***

4.7 Written-word form recall: The influence of context and word-related factors

When exploring the relative gains of each target word (see Table 62), it seems reasonable to assume that they differ as regards their learning burden. Thus, whether a word is easier or more difficult to learn may be associated to the influence of context and word-related factors (Barclay, 2021; Peters, 2020). The analyses in this section attempt to shed some light on this issue by examining the role played by word distribution (i.e. spacing effect), frequency of occurrence, regularity, length, and concreteness. It is important to acknowledge that one of the main disadvantages of using authentic materials is the inability to manipulate the target items. Thus, given that this study prioritized ecological validity over the control of context and word-related factors, the results reported in this section should be interpreted with due caution. It is important to note that the GLMMs (Binary logistic regressions) were conducted at item level since this is recommended for unbalanced samples (number of items per category).

Table 62.

WWFR: Relative gains per word

Word	Relative gains (%)	Regularity[a]	Length[b]	Concreteness[c]	Frequency[d]	Nº of episodes[e]
Mermaid	.9	1	2	1	2	2
Pleased	.9	1	2	1	2	2
Wide	3.4	1	1	1	2	2
Wobbly	4.2	1	1	1	2	2
Trolley	4.2	1	2	2	1	1
Stripy	4.3	1	1	2	1	1
Creaky	4.3	1	1	1	1	1
Web	4.4	2	1	1	2	1
Hairy	4.5	1	1	1	2	1
Cabbage	5.9	1	2	2	1	2
Slipper	5.9	2	2	2	1	1
Careful	6.0	1	2	1	2	2
Lead	6.1	1	1	1	1	1
Pillow	6.1	1	1	2	1	1
Useful	6.7	1	1	1	1	2
Fairy	6.8	1	1	1	2	2
Suitcase	8.0	1	2	2	1	1
Sausage	8.8	1	2	2	1	2
Costume	8.9	1	2	1	1	1
Fluffy	9.3	1	1	1	2	1
Puddle	9.3	1	1	2	2	2
Mud	9.5	1	1	2	1	2
Bandage	11.4	1	2	2	1	1
Wing	12.2	2	1	2	1	2
Clever	13.0	2	1	1	2	2
Track	13.0	2	1	1	2	2
Wand	13.0	1	1	2	1	1
Busy	13.7	1	1	1	2	1
Leaf	13.8	1	1	2	2	1
Sticky	17.3	2	1	1	1	2
Pea	20.4	1	1	2	2	2
Kitten	22.1	2	1	2	1	2
Shell	22.2	1	1	2	2	1
Handbag	27.0	2	2	2	2	1
Drop	31.5	2	1	1	2	1
Forest	43.1	2	1	1	1	2

a 1= Less consistent with L1 patterns, 2= More consistent with L1 patterns.
b 1= Shorter words (<= 6 letters), 2= Longer words (7+).
c 1= Less concrete words (<4.62), 2= More concrete words (>4.63).
d 1= Less frequent (3-5), 2= More frequent (6+).
e 1= Repetitions concentrated in one episode, 2= Repetitions distributed in multiple episodes.

4.7.1 The role of word distribution in the outcomes

As explained in the methodology section, half of the target words were encountered in a single video, while the other half were distributed in multiple episodes (2-3). This distinction was labelled as word distribution to compare massed and distributed encounters. Although these groups were comparable as regards frequency of occurrence, an optimal comparison between these two conditions would include the same set of target words in each category. Yet, the analyses on word distribution were considered to be relevant for two main reasons. Firstly, the words encountered in multiple episodes were explicitly tested by the construction-focused activities after their first encounter. Thus, the facilitating effect of the intentional after-viewing activity detected in previous analyses might be associated to potential higher levels of attention on subsequent encounters. Secondly, word frequency effects have been found to be higher in massed condition (Uchihara et al., 2019; Fievez et al., 2020), therefore, it would also be interesting to test this assumption.

Thus, a series of repeated-measures (word and time) compound-symmetry structure GLMMs (binary logistic regression) with student identification as subjects were run in order to explore the two assumptions mentioned above. To this aim, FKS binary score (at item level) was set as outcome variable, whereas time, activity type, word distribution (one vs. multiple episodes), frequency of occurrence (3-5 vs. 6+ repetitions), and some interactions of interest were entered into the model as independent variables: 1) Word distribution*time, 2) Word distribution*time*activity type, 3) Word distribution*time*frequency, 4)Activity type*time*frequency, 5)Activity type*frequency. The interactions that involved activity type and frequency (4 and 5) were included to test the assumption that test announcement (the completion of construction-focused activities in this case) increases learners' sensitivity to frequency effects (Uchihara et al., 2019). In addition, word concreteness and length were added as covariates. The non-significant interactions and factors were removed one by one from the analyses until obtaining the best fitted model. This was the case of activity type and the interactions where this factor was included. More precisely, this non-significant interaction indicated that testing the words that occurred in multiple episodes after their first encounter(s) (by means of the construction-focused activities) did not lead to higher gains. Moreover, the competition of construction-focused activities did not increase learners' sensitivity to frequency effects. Likewise, the non-significant interaction between word distribution and time suggests that none of the conditions (one or multiple episodes) was more conducive to learning.

As shown in Table 63, the analyses yielded a significant interaction between time, word distribution and frequency (F (4, 1864)= 18.085, p< .001). The Bonferroni pairwise contrasts (see Table 64) indicated that the beneficial effects of a higher frequency of encounters seems to be associated to the words that are massed in a single episode, while the words that are encountered less frequently might be better learnt in multiple episodes (see Figure 32).

Table 63.

The influence of word distribution on written-word form recall

Source	F	df1	df2	Sig.
Corrected Model	22.710	9	3496	.000
Time	57.910	1	1141	.000
Word distribution	2.261	1	7866	.133
Frequency	.463	1	8342	.496
Length	1.044	1	7087	.307
Concreteness	51.853	1	6229	.000
Time * Word distribution *Frequency	18.085	4	1864	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Binary scores

Table 64.

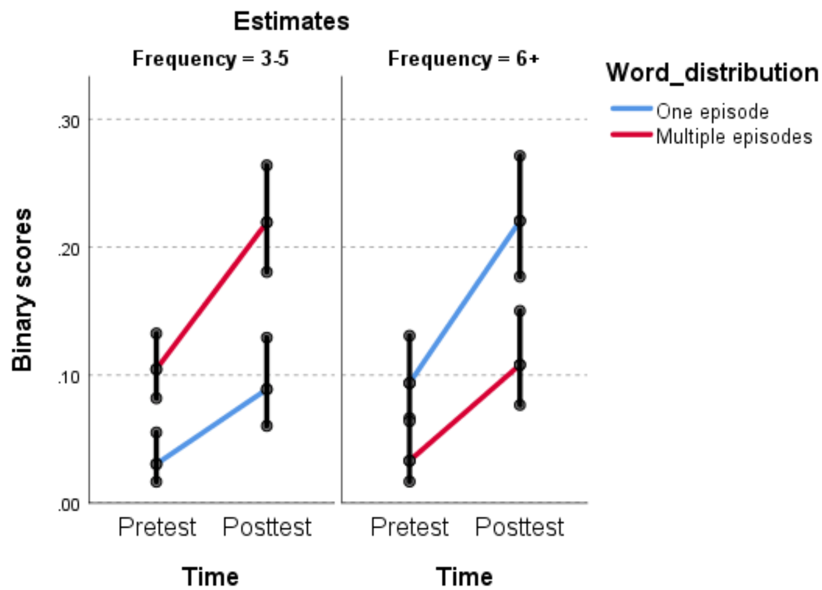
Bonferroni pairwise contrasts between word distribution categories by time and frequency

Time	Frequency	Word distribution Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Pretest	3-5	One vs. multiple episodes	-.074	.009	-7.865	6449	.000	-.093	-.056
	6+	One vs. multiple episodes	.061	.011	5.689	2560	.000	.040	.082
Posttest	3-5	One vs. multiple episodes	-.130	.014	-9.659	7081	.000	-.157	-.104
	6+	One vs. multiple episodes	.113	.016	6.845	3829	.000	.080	.145

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 32.

WWFR: Interaction between word distribution, time and frequency of occurrence.



4.7.2 The role of frequency and word-related factors in written-word form recall

A series of repeated-measures (word and time) compound-symmetry structure GLMMs (binary logistic regression) with student identification as subjects were run in order to explore the effects of frequency, regularity, concreteness and length. To facilitate the interpretation of the outcomes and improve model fit, word characteristics were transformed to categorical variables by using the visual binning tool in SPSS (equal percentiles). FKS binary score (at item level) was set as outcome variable, whereas time, frequency of occurrence (3-5 vs. 6+ repetitions), concreteness (low vs. high [4.63+]), word length (shorter [<=6 letters] vs. longer words [7 letters+]), and all possible two-way interactions were entered into the model as independent variables. The non-significant interactions were removed one by one from the analyses until obtaining the best fitted model. When the two-way interactions suggested the presence of a three-way interaction, the model was also fitted with three-way interactions to test this possibility.

As shown in Table 65, the results showed significant main effects for time (F (1, 955)= 41.459, p< .001), concreteness (F (1, 1141)= 80919, p< .001), length (F (1, 800)= 48.927, p< .001), frequency (F (1, 1219)= 44.306, p< .001), and regularity (F (1, 4459)=

24.297, $p < .001$). Regarding frequency, the results suggest that a higher number of encounters leads to higher gains (see Table 66). In addition, the results showed statistically significant interactions between time and length (F (1, 2204)= 5.729, p= .017), time, length and regularity (F (2, 2690)= 35.825, $p < .001$), and time, concreteness and regularity (F (3, 8339)= 4.504, p= .004).

Table 65.

The influence of context and word-related factors on written-word form recall

Source	F	df1	df2	Sig.
Corrected Model	15.919	12	5535	.000
Time	41.459	1	955	.000
Concreteness	80.919	1	1141	.000
Length	48.927	1	800	.000
Frequency	44.306	1	1219	.000
Regularity	24.297	1	4459	.000
Time * Length	5.729	1	2204	.017
Time * Length * Regularity	35.825	2	2690	.000
Time * Concreteness * Regularity	4.504	3	8339	.004

Probability distribution: Binomial
Link function: Logit a. Target: Binary scores

Table 66.

Bonferroni time pairwise comparisons between frequency groups

Frequency Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-5 - 6+	-.045	.007	-6.524	1133	.000	-.058	-.031

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

The statistically significant interaction between time and length indicated that shorter words were easier to learn, which is why they resulted in higher gains (see Table 67). The time*length*regularity interaction confirmed these results and added that the facilitating effect of word length is evidently enhanced when the orthographic patterns (sound-symbol correspondence) of the target words are more consistent with the regular patterns of L1 Spanish, which is a transparent language (see Table 68 and Figure 33). Likewise, the time* concreteness*regularity interaction suggests that the facilitating effect of concreteness is increased by word regularity (see Table 69).

Table 67.

Time pairwise contrasts per word length categories

Length (Binned)	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
<= 6,00	Pretest - Posttest	-.125	.015	-8.581	1081	.000	-.153	-.096
7,00+	Pretest - Posttest	-.079	.014	-5.560	565	.000	-.107	-.051

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 68.

Time pairwise contrasts by regularity and length

Length	Regularity	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
<= 6	Less consistent	Pretest - Posttest	-.075	.012	-6.083	526	.000	-.099	-.051
	More consistent	Pretest - Posttest	-.172	.020	-8.551	3103	.000	-.212	-.133
7+	Less consistent	Pretest - Posttest	-.056	.013	-4.459	756	.000	-.080	-.031
	More consistent	Pretest - Posttest	-.099	.019	-5.301	909	.000	-.135	-.062

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 69.

Time pairwise contrasts by regularity and concreteness

Concreteness	Regularity	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
<= 4,62	Less consistent	Pretest - Posttest	-.042	.010	-4.211	547	.000	-.062	-.022
	More consistent	Pretest - Posttest	-.109	.019	-5.820	1501	.000	-.146	-.073
4,63+	Less consistent	Pretest - Posttest	-.097	.014	-6.744	766	.000	-.125	-.069
	More consistent	Pretest - Posttest	-.219	.030	-7.353	1864	.000	-.277	-.160

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 33.

WWFR: Interaction between regularity, length and time

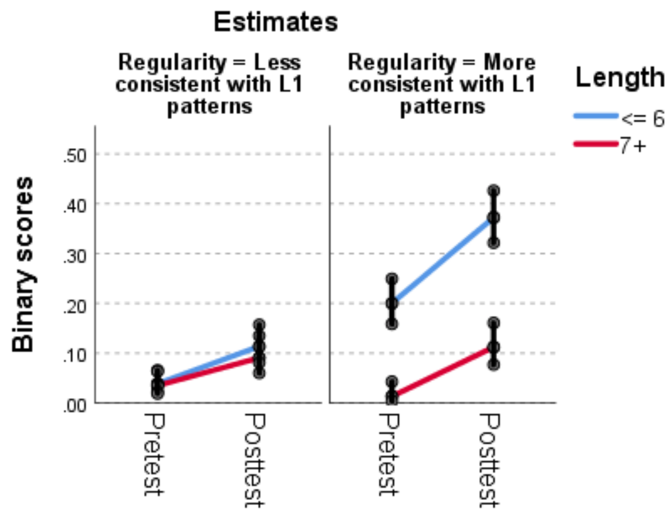
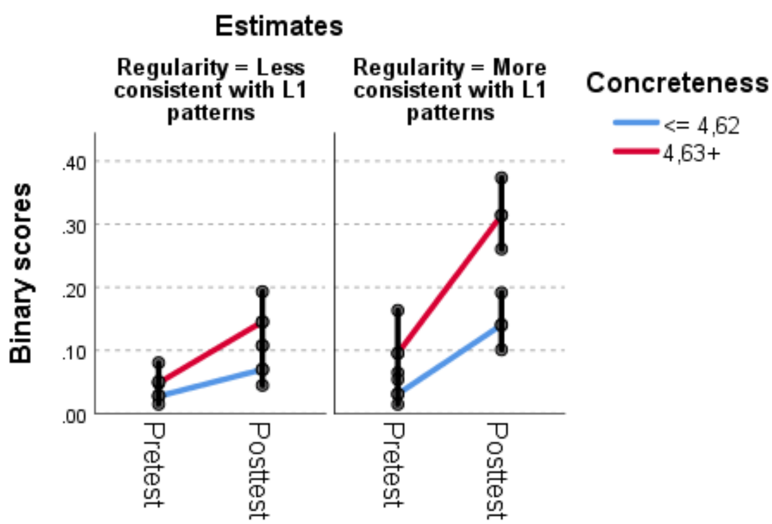


Figure 34.

WWFR: Interaction between regularity, time and concreteness



The results reported in this section explain why there was great variability in relative gains among the target words. To illustrate, *mermaid* (0.9%) and *pleased* (0.9%) obtained the lowest gains. These two words were long (7 letters), less concrete (4.5 and 2.37, respectively), and less consistent with the regular patterns of L1 Spanish, which are three of the factors that were found to increase the learning burden. Although *pleased* and *mermaid* were repeated six times, these repetitions occurred in multiple episodes. The results above indicated that frequency effects are more prominent when the repetitions are concentrated in a single episode, which is why these six repetitions did not seem to be effective. By contrast, the two words that obtained the highest relative gains were *drop* (31.5%) and *forest* (43.1%), which were shorter and more consistent with L1-Spanish patterns. In the case of *drop*, it was found to be less concrete (4.21) but its six encounters (higher frequency) were concentrated in a single episode. As for the word *forest*, its fewer repetitions (three) might have been compensated by its concreteness.

4.7.3 The influence of context and word-related factors: Summary of findings

The analyses reported in this section explain the great variability in relative gains among the 36 target words selected in this investigation. All in all, the analyses indicated that the following context and word-related factors were conducive to higher gains in written-word form recall.

- Regularity. The words that were more consistent with the regular patterns of L1 Spanish.
- Concreteness. Higher concreteness ratings.
- Word length. Shorter words, especially when they are more consistent with the regular patterns of L1-Spanish.
- Frequency. Higher number of encounters, especially when the repetitions were concentrated in a single episode.

Additionally, the analyses revealed that when the words were less frequent in the input, they were better learned in multiple episodes. As regards the potential role of construction-focused activities in the learning of the words whose occurrences were distributed in multiple episodes, the analyses indicated that being tested after the first encounter(s) did not enhance learning. By the same token, completing construction-focused activities did not increase learners' sensitivity to frequency effects.

4.8 Written-word form recall: Discussion

This section focused on vocabulary learning at the level of written-word form recall from captioned-video viewing. In addition, it assessed the influence of treatment, learner, context and word-related factors on the outcomes. To start with, the results revealed that, in comparison with the control group, all the experimental groups benefitted from the treatment significantly. Although the literature has consistently supported the beneficial effects of captioned videos on vocabulary learning (Montero Perez, 2022; Muñoz, 2022), this is an important finding since there is much less evidence collected from primary school learners (Montero Perez & Rodgers, 2019; Muñoz, 2022). On the whole, the gains were found to be relatively low, which may have been the result of the incidental learning conditions (Hulstijn, 2003, 2013; Webb, 2020), the higher demands of vocabulary recall (González-Fernández & Schmitt, 2020), the lack of vocabulary pre-teaching (Gesa, 2019; Pujadas & Muñoz, 2019), and learners' little knowledge of orthographic patterns in English (Birch & Fulop, 2021; Sun-Alperin & Wang, 2008).

The results also showed that fifth graders scored significantly higher than fourth graders at both testing times. Likewise, this group obtained greater gains from the treatment. Overall, these findings match those obtained in previous studies where late primary school learners outperformed the younger participants (e.g. Koolstra & Beentjes, 1999; Lekkai, 2014). This result may be attributed to learners' significantly higher proficiency level and higher vocabulary knowledge, corroborating the presence of a Matthew effect as in previous studies on vocabulary learning from audiovisual input, where the rich get richer (Montero Perez, 2022; Montero Perez et al., 2013; Pujadas & Muñoz, 2019; Stanovich, 1986). Yet, when taking a closer look at the gains obtained in each class, there seemed to be great variability among participants, which may be associated to the influence of multiple factors on the outcomes (Montero Perez, 2022; Muñoz, 2022).

4.8.1 Treatment-related factors

The treatment-related factors studied for the purpose of this study were the use of after-viewing activities (i.e. meaning-focused vs. construction-focused) and viewing distribution (i.e. lag effects operationalized as the number of episodes watched a week). As regards after-viewing activity type, the results indicated that the use of both, meaning-focused and construction-focused activities fostered vocabulary learning; however, the implementation of construction-focused activities (intentional condition) was shown to be conducive to higher gains over time. This falls in line with the results obtained by Montero Perez et al. (2015), where test announcement increased learners' attention to unknown words and resulted in better outcomes. By the same token, this finding is congruent with the results of the studies that have implemented pre-teaching activities to enhance learning (e.g. Pujadas, 2019; Teng, 2022). Yet, the further analyses that included vocabulary knowledge (EFL PVT) as covariate also indicated that the benefits of construction-focused activities were more prominent in fourth graders. This may imply that in the case of the younger participants, written-word form recall may be more difficult to achieve in incidental learning conditions (Kim & Webb, 2022a; Webb, 2020) due to their lower L2 proficiency level and still developing cognitive and L1 literacy skills (Holmes & Myles, 2019). As the literature suggests, a lower vocabulary coverage results in a more effortful processing of input, leaving little cognitive resources available to notice and pick up unknown vocabulary items (Kormos, 2017; Lin & Siyannova-Chanturia, 2015). Likewise, the benefits of intentional activities may be more evident when the target language items convey greater levels of difficulty (Montero Perez et al., 2015), which may explain why the difference between activity groups was more evident in year 4.

All in all, the results confirm that the use of construction focused-activities may be used as an enhancement technique (Montero Perez et al., 2015) to promote vocabulary learning, especially in younger or less proficient students. Still, it is important to acknowledge that the use of captioned videos may be complemented by more effective intentional vocabulary activities, such as flashcards (Barclay, 2021; Webb et al., 2020). Indeed, based on Nation and Webb's (2011) technique feature analysis, the multiple-choice format used in the construction-focused activities may be considered as relatively effective, and probably, one of the main disadvantages is that the participants were not explicitly anticipated on the target items that were going to be assessed. Nonetheless, as explained in the methodology section, the format allowed the administration of both types of activities in the same classroom without alerting the participants about their different objectives.

As concerns viewing distribution, the comparisons between fourth graders indicated that 4-fourth obtained greater gains from the treatment. Yet, the increase in vocabulary gains was not found to be linear given that 3-fourth obtained the lowest gains. The further analyses that included vocabulary knowledge as a covariate attempted to disentangle the role of viewing distribution by controlling for L2 knowledge. The results suggested that the performance of the participants that watched between one and three episodes a week did not differ as a function of viewing distribution. However, a shorter distance between episodes (i.e. watching four episodes a week) clearly boosted learners' outcomes at immediate posttest as in previous studies on vocabulary learning with primary and secondary school participants (e.g. Küpper-Tetzel et al., 2014; Serrano & Huang, 2018, 2021). Hence, the findings appeared to indicate the presence of a potential threshold in fourth graders (four episodes a week, ISI-1.75). By contrast, the advantage of 4-fifth over 2-fifth was found to be limited, suggesting that fourth graders were more sensitive to the effects of viewing distribution. Hence, these results relate to Suzuki et al.'s (2019) assertion that shorter lags between episodes may be suitable to learn more difficult language aspects, which may have been the case of fourth graders and written-word form recall. In like manner, the advantage of shorter lags between episodes corroborates the results obtained by Serfaty and Serrano's (2022a) in that a small amount of spacing was found to be more appropriate for slower and lower proficiency students.

Considering that half of the words were encountered in a single episode, which means that the distance between encounters was the same for all the viewing distribution groups, the results may also be interpreted in light of Greving and Richter's (2021) findings and the concept of narrow viewing (Rodgers & Webb, 2011). Specifically, when texts (or episodes as in the current investigation) are separated by short spacing, they are perceived as more connected and easier to understand. Thus, a shorter distance between episodes might have facilitated the processing of input, leaving more cognitive resources available to promote the learning of unknown words (Kormos, 2017; Lin & Siyannova-Chanturia, 2015).

All in all, the advantage of 4-fourth over the rest of the participants in the same year level has important implications. The implementation of shorter lags between episodes seemed to moderate the influence of language-related factors. This may explain why 4-fourth's performance was not strongly associated to SR efficacy, and their progress was magnified when fitting a model with language-related factors. Therefore, it may be assumed that when learners are less proficient, concentrating the episodes in a shorter period of time may facilitate the viewing process and, to a certain extent, compensate for their knowledge gaps. These results partially concur with those of Collins and White (2012), who found that the concentration of L2 instruction moderated the influence of individual differences in young learners. Nonetheless, it is worth mentioning that the participants of the present investigation were not tested at delayed posttest, thus it was not possible to examine whether the short-lag advantage was kept over time.

4.8.2 Cognitive and language-related factors

Among the cognitive factors assessed for the purpose of this study, the analyses yielded a significant interaction between PSTM and time, indicating that this factor had a slightly higher influence at posttest. This is congruent with the literature, which suggests that, overall, PSTM plays a more significant role at early L2 learning stages (Wen & Jackson, 2022; Wright, 2015) in the areas of vocabulary and lexically driven grammar learning (Wright, 2015). In addition, they seem to fall in line with the fact that PSTM plays a more important role in incidental learning (Kormos & Sáfár, 2008), while complex working memory may be more relevant under explicit L2 learning conditions (Kormos & Sáfár, 2008; Suárez et al., 2021; Wen & Jackson, 2022) and heavier cognitive demands (Li et al., 2019). Hence, learners' accuracy in the dictation test may not have been predicted by the complex cognitive processes entailed in the integration of verbal and non-verbal information, as in other word knowledge dimensions such as form-meaning mapping (Suárez & Gesa, 2019). It might also be the case that the use of construction-focused activities did not lead to fully intentional learning conditions. Even when the participants tried to commit some words to memory, they might have primarily focused on comprehension since the activity also included comprehension questions and the learners were not explicitly anticipated on the target words. As observed in Montero Perez et al.'s (2018) study, learners may prioritize meaning over word learning despite the announcement of an upcoming vocabulary test. In fact, previous research suggests that the simultaneous attention to comprehension and intentional vocabulary learning increases the cognitive load, therefore lower-proficiency learners may prioritize one of these aspects when processing the input (Pujadas & Muñoz, 2020). Thus, their gains might have been the result of semi-incidental learning conditions, which may explain why only PSTM and not complex WM played a significant role in written-word form recall.

Additionally, it may be hypothesized that the simultaneous processing of captions and audio facilitated input processing, reducing the cognitive demands, which is why learners' outcomes were not significantly predicted by complex WM (Gregersen & MacIntyre, 2014; Mayer et al., 2020; Pattemore & Muñoz, 2020). This corroborates the results obtained by Pattemore and Muñoz's (2020), which indicated that complex WM was only associated to the non-captions condition. Alternatively, the lack of significance of complex working memory might also be related to the fact that working memory capacity does not reach adult-like levels before the age of 14 (Gathercole et al., 2004; Wright, 2015), which may explain why under that age the outcomes may be inconsistent (Wright, 2015). As for visual processing speed, learners' performance at the dictation task was neither found to be influenced by the speed at which learners processed non-verbal visual input, nor their visual scanning ability, visual discrimination, multitasking or their capacity to direct sustained attention to task (Flanagan & Alfonso, 2017; Weiss et al., 2019). The fact that visual processing did not emerge as a significant predictor of written-word form recall may be expected since learners' outcomes might have relied more on their ability to process bimodal verbal input rather than imagery.

It is important to note that the further analyses indicated that PSTM was not a strong predictor of written-word form recall when L2-related factors were fitted in the same model. First of all, based on Porter's (2017) findings, the weaker effects of PSTM may be associated to the enhancement of vocabulary learning through the simultaneous presentation of aural and written representations. In addition, these results may be explained in light of the Simple view of reading model (Gough & Tunmer, 1986; Hoover & Gough, 1990; Tunmer & Chapman, 2012) given that L2 reading comprehension is mainly predicted by L2-related factors (e.g. vocabulary knowledge and listening skills) rather than L1 literacy and cognitive skills (Alderson et al., 2016; Pattemore & Serra, 2021; Sparks, 2021). This is somehow connected with vocabulary learning from captioned-video viewing since reading is a key component of the viewing process. Then, it might have been learners' L2-related factors the ones that determined the cognitive effort involved in the comprehension process and the availability of attentional resources to notice and pick up words from the input (Kim & Webb, 2022a; Kormos, 2017; Lin & Siyanova-Chanturia, 2015; Montero Perez, 2020). In like manner, when learners are more proficient in the target language, the L2 knowledge stored in their long-term memory supports the learning of unknown items (Kim & Webb, 2022a; Montero Perez, 2020). To give an example, more proficient L2 learners may have greater knowledge of phoneme-grapheme correspondences in a language (Birch & Fulop, 2021), which may enhance vocabulary learning at the level of written-word form recall.

Specifically, the analyses on language-related factors revealed that vocabulary knowledge, listening skills, English text segmentation and SR efficacy significantly influenced learners' gains over time. Given that L2 comprehension in different modalities has been found to have a strong relationship with learners' vocabulary knowledge (Montero Perez, 2020; Miralpeix & Muñoz, 2018; Stæhr, 2008), it is unsurprising that this factor emerged as one of the strongest predictors of written-word form recall. In previous studies on audiovisual input, vocabulary knowledge has consistently been identified as a strong predictor of L2 learning (e.g. Alexiou, 2015; Montero Perez et al. 2013, 2018; Peters et al., 2016; Peters & Webb, 2018; Pujadas & Muñoz, 2019), confirming the rich-get-richer principle. The significant effects of L2 listening skills may also be associated to their key role in learners' comprehension of the L2 (Gough & Tunmer, 1986; Hoover & Gough, 1990; Sparks, 2021; Tunmer & Chapman, 2012), and the importance of aural-word form representations in the dictation test. Although listening skills have not been widely studied as a factor in L2 learning from audiovisual input, the few investigations that have examined its effects have provided evidence of its significant contribution to the learning process (Pattemore & Muñoz, 2020; Pujadas & Muñoz, 2019; Suárez & Gesa, 2019).

The fact that ER efficacy was not found to be a statistically significant predictor of written-word form recall may be associated to the specific contribution of lower-level reading skills, which were measured by means of the instrument on English text segmentation. Thus, learners' capacity to efficiently decode the text (with aural support) aided the learning of written word forms. As for SR efficacy, the extent to which learners relied on this factor depended on age and viewing distribution. More precisely, the groups of fourth graders that watched fewer episodes a week relied more on SR reading efficacy to learn from the treatment. Thus, along with their lower L2 proficiency level and still developing L1 literacy skills, the greater distance between the episodes may have made the viewing process more effortful. On the whole, fourth graders might have relied on SR efficacy to compensate, to a certain extent, for their L2 knowledge gaps (Yamashita, 2002). Equally important, the results seem to demonstrate that in early stages, learners progressively assimilate and accommodate their linguistic infrastructure to the characteristics of the L2 (Birch & Fulop, 2021; Jiang et al., 2019; Perfetti et al., 2007), which is a process that evolves according to learners' L2 proficiency level and familiarity with the characteristics of target language (Jiang et al., 2019). The fact SR efficacy and not Spanish text segmentation was shown to be a significant predictor of written-word form recall may be associated to the fact that the SR efficacy test integrated the assessment of lower-level and higher-level reading skills. Therefore, learners' reliance on SR efficacy might not only have supported input decoding but also the application of comprehension strategies to cope with the task demands.

4.8.3 Context and word-related factors

The context and word-related variables examined in this investigation were frequency of occurrence, regularity, concreteness and length (number of letters). Overall, the results indicated that all these factors significantly affected word learnability. With respect to frequency of occurrence, the results revealed that a higher number of encounters increased the odds of recalling written-word forms. Nevertheless, as suggested in the literature, frequency is one of the many factors that affect language learning (Ellis & Wulff, 2015) and may be moderated by other variables (Uchihara et al., 2019). Although the analyses were unable to demonstrate a relationship between activity type and frequency of occurrence, the results confirmed that frequency effects are more evident when the repetitions are concentrated in a single episode (Fievez et al., 2020; Uchihara et al., 2019). Yet, as in the case of the word *forest* exemplified in the analyses section, three repetitions distributed in two episodes were enough to result in the highest relative gains. Therefore, it may be assumed that the rest of the factors (word length, concreteness and regularity) compensated for the lack of repetitions. What this means is that, in the case of the items that are shorter, more concrete, and more consistent with L1 patterns, fewer encounters are sufficient to foster learning. Conversely, a higher number of repetitions may be required to learn the words that have a heavier learning burden (Barclay & Pellicer-Sánchez, 2022).

In regard to regularity, the analyses clearly indicated that the words whose representations were more consistent with the regular orthographic patterns of L1-Spanish were easier to learn. This result is not surprising since the literature has consistently demonstrated that regularity is an important source of difficulty for EFL learners (Krepel et al., 2020, 2021; Muñoz, 2017b; Sun-Alperin & Wang, 2008). The participants' little knowledge of phoneme-grapheme correspondences in English may have affected their capacity to recall written-word form representations. In fact, a general revision of their errors suggested that learners' attempts mainly relied on one-to-one correspondences; thus, our findings seem to be consistent with those of Sun-Alperin and Wang (2008), who found that the errors of L1 Spanish learners of English were influenced by the patterns of the L1. Therefore, the findings of the present study appear to support the view that, at early stages, learners' linguistic infrastructure progressively assimilates and accommodates to the patterns of the L2 (Birch & Fulop, 2021; Perfetti et al., 2007).

It is also important to mention that learners' accuracy to write the words that were less consistent with L1 patterns did not seem to be aided by word length. Although the literature posits that at early stages learners may rely in word memorization (Birch, 2015), the limited difference between longer and shorter words suggested that this does not necessarily apply to the learning of more complex items by primary school learners. Yet, none of the activities involved the recall of written word forms, therefore, the implementation of more explicit activities or instruction in this regard might eventually increase learners' outcomes (Marian et al., 2021; Pérez Cañado, 2006; Porter, 2020; Pujadas & Muñoz, 2019). Nonetheless, a different picture was observed between word length and regularity in the words that were more consistent with the patterns of the L1. Specifically, the rate of accuracy was significantly higher in shorter words. This result partially concurs with the outcomes obtained by Krepel et al. (2020), where the regular words resulted in higher gains in aural form recall, which was attributed to the use of bimodal verbal input and the creation of stronger knowledge representations. It can thus be suggested that the use of captioned videos with young learners may primarily foster written-word form recall in more regular words, particularly when they are shorter. On the whole, the results concerning word length point to the advantage of shorter words, confirming that longer words are more difficult to learn (e.g. Ellis & Beaton, 1993a; Barclay & Pellicer-Sánchez, 2022). As the literature suggests, this difficulty may be accounted by the additional time needed to recognize and process longer words (Grabe, 2009), the likelihood of containing more complex patterns (Ellis & Beaton, 1993a), and the greater effort required to store them in PSTM (Birch, 2015).

Concerning word concreteness, the analyses indicated that higher concreteness ratings increased the odds of recalling written-word forms. The advantage of concreteness emerged in both, the words that were less and more consistent with the patterns of the L1. However, the influence of this factor was more prominent in the case of the more regular words. The results from this study confirm previous findings regarding the beneficial effects of word concreteness in intentional and incidental conditions (e.g. De Groot & Keijzer, 2000; Ellis & Beaton, 1993b; Puimège & Peters, 2019b), which are associated to their greater saliency (Crossley et al., 2016) and imageability (Peters, 2020). This is particularly relevant in the case of vocabulary learning from captioned-video viewing since more concrete words tend to be graphically represented onscreen (Peters, 2020). Previous research has already demonstrated that concreteness also affects the learning of written-word forms (e.g. Puimège & Peters, 2019b; Van Zeeland & Schmitt, 2013). This may be interpreted in light of the Dual Coding Theory where the strength of the relationship between verbal and non-verbal representations relies on word concreteness (Clark & Paivio, 1991); therefore, this factor accounts for learners' capacity to evoke these representations (Clark & Paivio, 1991). Indeed, previous studies have corroborated that the use of verbal input and supporting imagery leads to higher levels of accuracy in word spelling (e.g. Sadoski et al., 2004). Thus, our findings further support the idea that the use of multimodal input strengthens the encoding of information and their further retrieval (Mitchell & Rule, 2022, p.40).

V. Written-word form and meaning recognition

This section focuses on vocabulary learning from captioned-video viewing at the level of written-word form and meaning recognition (i.e. receptive form-meaning mapping). Specifically, it attempts to answer the following research questions:

1) To what extent does viewing distribution (i.e. shorter vs. longer lags) influence young L2 learners' gains from captioned video viewing?

2) In comparison with meaning-focused activities, what are the effects of construction-focused after-viewing activities on L2 learning through captioned-video viewing?

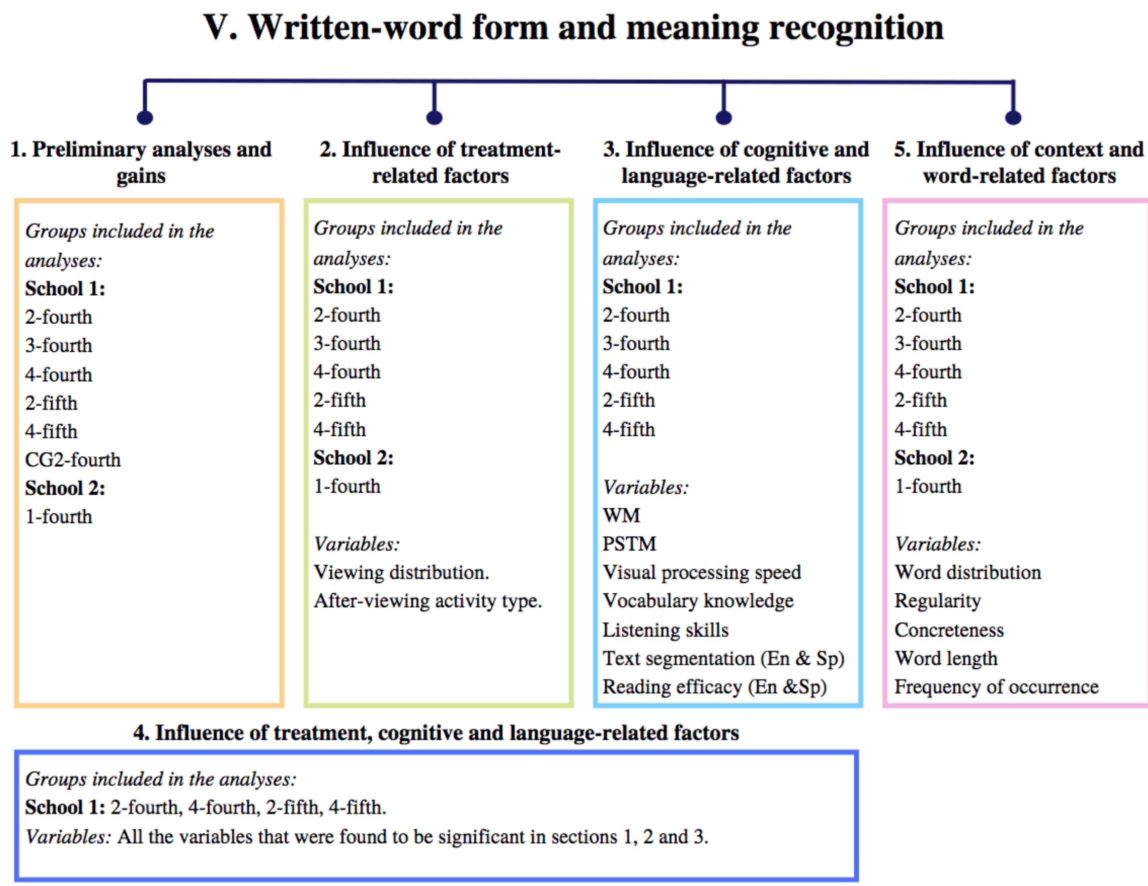
3) To what extent do learner characteristics influence young L2 learners' gains from captioned-video viewing? (i.e. age, vocabulary knowledge, cognitive abilities [phonological short-term memory, complex working memory, and visual processing speed], L1 and L2 reading skills [reading efficacy and text segmentation] and L2 listening skills).

4) To what extent do context and word-related factors (frequency of occurrence, regularity, word-length, and concreteness) influence vocabulary learning?

The overview of this section is displayed in Figure 35. As explained earlier, written-word form and meaning recognition (WWFMR) measured by means of a multiple-choice test was tested at pretest, posttest and delayed posttest. The procedures followed for the analyses are similar to those used in written-word form recall. Considering that the normality tests indicated that students' scores were not normally distributed, the dependent variable (written-word form and meaning recognition) was square root transformed (SQRT) to be used in the analyses where normal distribution is required (e.g. ANOVAs).

Figure 35.

Section 5 overview



5.1 Written-word form and meaning recognition: Preliminary analyses

To start with, a set of between-groups comparisons were run with the aim of establishing whether the groups were comparable in terms of written word-form and meaning recognition at pretest (See Figure 36). An independent-samples T-test showed that the difference between fourth and fifth graders was only marginally significant (t (135)= 1.847, p=.067, r=.1), suggesting a moderate advantage for fifth graders. In addition, a One-Way ANOVA revealed that the differences between classes as regards pretest results did not reach significance (F (6)= 1.052, p=.395, η^2 = .046), suggesting that the groups from each year level were comparable as concerns previous knowledge (see descriptive statistics in Table 70). The comparisons are summarized in Table 71.

Figure 36.

WWFMR: Learners' scores over time

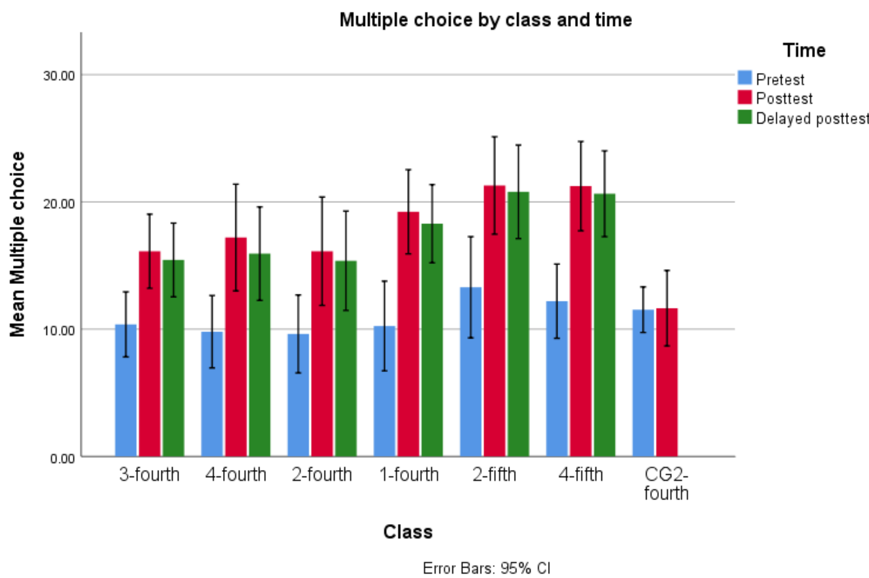


Table 70.

WWFMR: Descriptive statistics

		Pretest		Posttest		Delayed posttest		Posttest gains		Delayed posttest gains	
		M	(SD)	M	(SD)	M	(SD)	M	(SD)	M	(SD)
Class	1-fourth	10.3	(8.3)	19.2	(7.5)	18.3	(6.7)	8.5	(4.5)	8.9	(4.4)
	2-fourth	9.6	(5.7)	16.1	(8.0)	15.4	(7.3)	6.5	(4.9)	5.8	(3.7)
	3-fourth	10.4	(4.8)	16.1	(5.5)	15.4	(5.4)	5.8	(2.4)	5.1	(2.3)
	4-fourth	9.8	(5.1)	17.2	(7.6)	15.9	(6.6)	7.4	(4.6)	6.1	(3.2)
	2-fifth	13.3	(9.4)	21.3	(9.1)	20.8	(8.7)	8.0	(4.6)	7.5	(3.8)
	4-fifth	12.2	(7.1)	21.2	(8.5)	20.6	(8.2)	9.0	(5.0)	8.4	(4.1)
	CG2-fourth	11.52	(3.4)	11.65	(5.77)	-	-	0.12	(3.85)	-	-
Year level	Year 4*	10.0	(6.3)	17.3	(7.2)	16.4	(6.6)	7.2	(4.3)	6.6	(3.8)
	Year 5	12.7	(8.2)	21.3	(8.7)	20.7	(8.3)	8.5	(4.8)	8.0	(3.9)

*Without CG2-fourth.

Table 71.

Summary: Between-groups comparisons in WWFMR at pretest

Factor	Statistically sig. differences between year levels	Statistically sig. differences between classes
Written-word form and meaning recognition at pretest.	Year 5 > Year 4 (marginally significant)	No significant differences between groups

In addition, Pearson correlations were performed in order to explore the relationships between learners' outcomes over time and the continuous variables assessed for the purpose of this study (cognitive and language-related factors) (see Table 72). The results revealed stronger relationships between learners' accuracy in written-word form and meaning recognition and the language-related factors, especially in the case of L2 vocabulary knowledge, where the shared variance ranged between 62% and 66% (large effect size). In comparison with the correlations performed in written-word form recall, the relationship with vocabulary knowledge was found to be stronger at the level of receptive form-meaning mapping (R^2 = .37 - .40 vs. R^2 = .62 - .66, respectively).

Table 72.

Correlations between WWFMR with learner-related factors

		Pretest	Posttest	Delayed posttest
Pretest	Pearson Correlation	1	**,804**(R^2=.64)**	**,815**(R^2=.66)**
	Sig. (2-tailed)		,000	,000
	N	120	118	117
Posttest	Pearson Correlation	**,804**(R^2=.64)**	1	**,974** (R^2=.94)**
	Sig. (2-tailed)	,000		,000
	N	118	118	115
PSTM	Pearson Correlation	,270**(R^2=.07)	,382**(R^2=.14)	,379**(R^2=.14)
	Sig. (2-tailed)	,009	,000	,000
	N	93	93	93
WM	Pearson Correlation	,250* (R^2=.05)	,326** (R^2=.10)	,316**(R^2=.09)
	Sig. (2-tailed)	,016	,001	,002
	N	93	93	93
Visual processing speed	Pearson Correlation	,234*(R^2=.05)	,164 (R^2=.02)	,239*(R^2=.05)
	Sig. (2-tailed)	,024	,116	,021
	N	93	93	93
Vocabulary knowledge	Pearson Correlation	**,790** (R^2=.62)**	**,789**(R^2=.62)**	**,814**(R^2=.66)**
	Sig. (2-tailed)	,000	,000	,000
	N	112	110	109
Listening skills	Pearson Correlation	**,676**(R^2=.45)**	**,640**(R^2=.40)**	**,675**(R^2=.45)**
	Sig. (2-tailed)	,000	,000	,000

	N	91	91	91
Spanish segmentation	Pearson Correlation	,350**(R^2=.12)	,435**(R^2=.18)	,449**(R^2=.20)
	Sig. (2-tailed)	,001	,000	,000
	N	114	112	111
English segmentation	Pearson Correlation	**,512** (R²=.26)**	**,622**(R²=.38)**	**,629**(R²=.39)**
	Sig. (2-tailed)	,000	,000	,000
	N	113	111	110
Spanish reading efficacy	Pearson Correlation	,495**(R^2=.24)	**,502**(R²=.25)**	**,519**(R²=.26)**
	Sig. (2-tailed)	,000	,000	,001
	N	92	92	92
English reading efficacy	Pearson Correlation	**,512**(R²=.26)**	**,530**(R²=.28)**	**,527**(R²=.27)**
	Sig. (2-tailed)	,000	,000	,000
	N	92	92	92

**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

5.2 Written-word form and meaning recognition: Progress over time

To compare the trajectories of both year levels over time, we ran a compound symmetry structure GLMM (binary logistic regression) with student identification as subjects, and word and time as repeated measures. The model was built with learners' dichotomous scores (by items) as dependent variable, and the following fixed effects: time, year level and their interaction. In this model, the scores obtained by the control group were not included. The results yielded significant effects for year level (F (1,97)= 6.611, p= .012) and time (F (2,9650)= 194.067, p< .001). The interaction between year level and time did not reach significance (F (2,9650)= .885, p= .413) but was kept in the model for further exploration (see Table 73). The Bonferroni pairwise contrasts displayed in Table 74 indicate that both year levels showed significant improvement from pretest to posttest, and from pretest to delayed posttest. In addition, the statistically significant difference between fifth and fourth graders was kept over time and did not experiment great variability (see Figure 37). On the whole, the treatment appeared to be equally beneficial for both year levels.

Table 73.

WWFMR: The influence of time and year level on the results

Source	F	df1	df2	Sig.
Corrected Model	84.583	5	512	.000
Level	6.611	1	97	.012
Time	194.067	2	9650	.000
Level * Time	.885	2	9650	.413

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

Table 74.

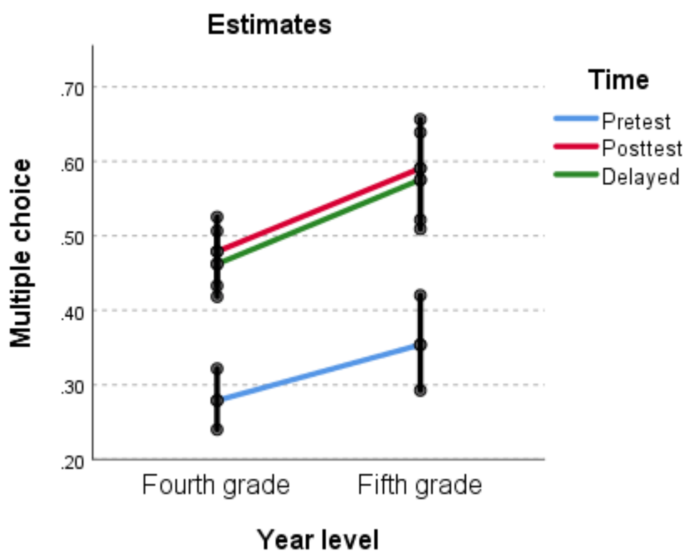
WWFMR: Year level and time comparisons

Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
Year 4 – year 5	-.103	.040	-2.571	96	.012	-.183	-.024
Pretest - Posttest	-.220	.012	-18.468	4117	.000	-.249	-.192
Pretest - Delayed	-.204	.010	-20.342	4578	.000	-.226	-.181
Posttest - Delayed	.016	.004	3.754	12774	.000	.008	.025

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 37.

WWFMR: The trajectory of each year level over time



5.2.1 Comparisons between control and experimental groups' performance

In order to assess the performance of the control and the experimental groups from pretest to posttest, we ran a compound symmetry structure GLMM (binary logistic regression) with student identification as subjects, and time as repeated measures. The analysis was calculated with learners' dichotomous scores in written-word form and meaning recognition (i.e. at item level); while class, time, and their interaction were entered into the model as fixed factors. As shown in Table 75, the results revealed significant effects for time (F (1,4308)= 359.948, p< .001), and the interaction between class and time (F (6,4762)= 14.168, p= .012). The Bonferroni pairwise contrasts in Table 76 indicated that all the groups showed significant improvement from pretest to posttest, including the control group. However, the gains shown by the latter were significantly lower.

Table 75.

WWFMR: Control and experimental groups' progress over time

Source	F	df1	df2	Sig.
Corrected Model	34.634	13	443	.000
Class	1.887	6	142	.087
Time	359.948	1	4308	.000
Class * Time	14.168	6	4762	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

Table 76.

WWFMR: Time pairwise contrasts per class

Class	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-fourth	Pretest - Posttest	-.160	.016	-9.752	9706	.000	-.192	-.128
4-fourth	Pretest - Posttest	-.206	.032	-6.500	2007	.000	-.268	-.144
2-fourth	Pretest - Posttest	-.181	.033	-5.512	2107	.000	-.245	-.116
2-fifth	Pretest - Posttest	-.222	.025	-8.764	5089	.000	-.272	-.173
4-fifth	Pretest - Posttest	-.251	.027	-9.143	2879	.000	-.305	-.197
1-fourth	Pretest - Posttest	-.237	.026	-9.096	3800	.000	-.288	-.186
CG2-fourth	Pretest - Posttest	-.036	.015	-2.442	9706	.015	-.066	-.007

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

5.2.2 Comparisons between experimental groups

In order to compare the performance of each class over time, a Generalized Linear Mixed Model (GLMM; Binary Logistic regression) with repeated measures (time) compound-symmetry structure was run. The model was calculated with written-word form and meaning recognition scores as target variable, which included the total scores obtained by each participant at each testing time. The maximum score per test (36) was set as denominator. The fixed effects included in the analysis were class, time, and their interaction. As summarized in Table 77, the results showed a non-significant main effect for class (F (5,129)= 1.743, p= .129), whereas time (F (2,141)= 221,357, p< .001), and the interaction between class and time (F (10,157)= 2.042, p= .032) reached statistical significance. Specifically, the Bonferroni adjusted results revealed that the differences between groups were not significant over time (from pretest to delayed posttest) (p > .05). In addition, all the groups were shown to improve significantly from pretest to posttest, as well as from pretest to delayed posttest (p< .05). However, the mean delayed-posttest score was significantly lower than the posttest score in 3-fourth (p= .030) and 4-fourth (p= .031) (see Table 78 and Figure 38), whereas in 1-fourth, 2-fourth and 2-fifth, the analyses did not yield significant word-knowledge decay. In the case of 4-fifth, the difference between posttest and delayed posttest was marginally significant (p= .054).

Table 77.

WWFMR: Groups' progress over time

Source	F	df1	df2	Sig.
Corrected Model	33,422	17	200	,000
Class	1,743	5	129	,129
Time	221,357	2	141	,000
Class * Time	2,042	10	157	,032

Probability distribution: Binomial
Link function: Logit
a. Target: Written-word form and meaning recognition/36

Table 78.

WWFMR: Interaction between class and time

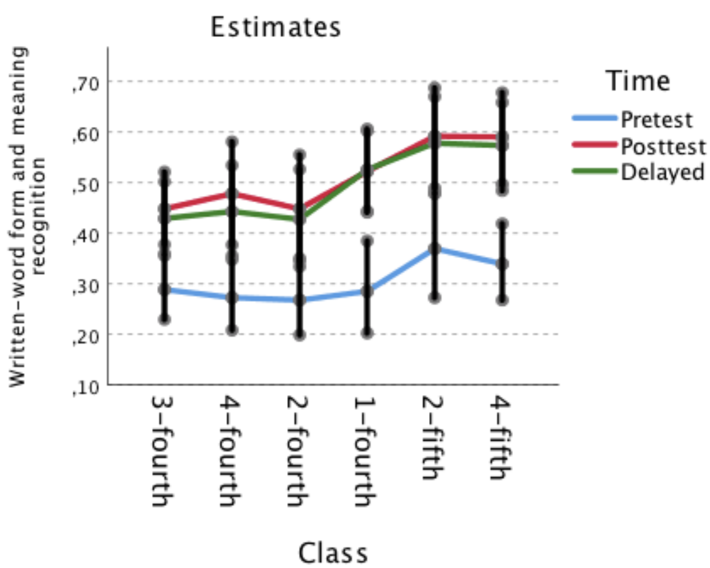
Class	Time Pairwise Contrasts	Contrast Estimate	SE	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-fourth	Pretest - Posttest	-,160	,016	-9,752	337	,000	-,199	-,120
	Pretest - Delayed	-,141	,016	-9,003	337	,000	-,176	-,105
	Posttest - Delayed	,019	,009	2,181	337	,030	,002	,036
4-fourth	Pretest - Posttest	-,206	,032	-6,500	107	,000	-,277	-,134
	Pretest - Delayed	-,170	,022	-7,674	150	,000	-,224	-,117
	Posttest - Delayed	,035	,016	2,166	313	,031	,003	,067
2-fourth	Pretest - Posttest	-,181	,033	-5,512	72	,000	-,256	-,106
	Pretest - Delayed	-,160	,025	-6,489	84	,000	-,220	-,100
	Posttest - Delayed	,021	,013	1,618	239	,107	-,005	,046

1-fourth	Pretest - Posttest	-,237	,025	-9,405	52	,000	-,295	-,179
	Pretest - Delayed	-,240	,025	-9,643	53	,000	-,302	-,179
	Posttest - Delayed	-,003	,011	-,305	337	,761	-,025	,018
2-fifth	Pretest - Posttest	-,222	,025	-8,764	69	,000	-,280	-,164
	Pretest - Delayed	-,208	,021	-9,969	73	,000	-,260	-,157
	Posttest - Delayed	,014	,008	1,769	337	,078	-,002	,029
4-fifth	Pretest - Posttest	-,251	,027	-9,143	50	,000	-,315	-,188
	Pretest - Delayed	-,234	,022	-10,628	53	,000	-,289	-,180
	Posttest - Delayed	,017	,009	1,936	210	,054	,000	,034

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 38.

WWFMR: Groups' progress over time



After that, in order to confirm the outcomes above, we ran a compound symmetry structure GLMM (binary logistic regression) at item level with student identification as subjects, as well as time and word as repeated measures. The dichotomous values were entered into the model as target variable, while time, class and their interaction were included as factors. As shown in Table 79, the analyses yielded a non-significant main effect for class (F (5,131)= 1.743, p= .129), but statistically significant effects for time (F (2,7715)= 221.165, p< .001) and the interaction between class and time (F (10, 7902)= 2.035, p= .026), confirming the results above. Taken together, the results indicate that the treatment had a significant effect on all the experimental groups, obtaining mean relative gains (Horst et al., 1998) of 34.9% at pretest and 32.8% at delayed posttest (see Table 80).

The non-significant difference between posttest and delayed posttest in the groups that watched either one or two episodes a week suggests that longer lags between sessions prevented word-knowledge decay, therefore, these outcomes are further explored in the sections below. By the same token, the great variability in relative gains displayed in Table 80 implies that different factors influenced learners' performance. Thus, the roles of a series of treatment- and learner-related variables are also examined in the following sections. The results are summarized in Table 81.

Table 79.

WWFMR: GLMM by items to study learners' progress over time

Source	F	df1	df2	Sig.
Corrected Model	33,416	17	495	,000
Time	221,165	2	7715	,000
Class	1,743	5	131	,129
Time * Class	2,035	10	7902	,026

Probability distribution: Binomial
Link function: Logit
a. Target: Written-word form and meaning recognition.

Table 80.

WWFMR: Absolute and relative gains

	Minimum	Maximum	Mean (SD)
Posttest gains (N) (Mean nº of items)	-2,00	22,00	7,72 (4,53)
Delayed posttest gains (Mean nº of items)	-4,00	21,00	7,19 (3,89)
Absolute posttest gains (%)	-5,56	61,11	21,46 (12,58)
Absolute delayed posttest gains (%)	-11,11	58,33	19,99 (10,81)
Relative posttest gains (%)	-9,52	100,00	34,90 (23,53)
Relative delayed posttest gains (%)	-19,05	100,00	32,87 (21,33)

Table 81.

Summary: Written-word form and meaning recognition over time

Analysis	Outcome
Significant improvement from pretest to posttest (p<.05).	Fourth and fifth grade. In all experimental groups: 1-fourth, 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth.
Significant improvement from pretest to delayed posttest (p<.05).	Fourth and fifth grade. In all experimental groups: 1-fourth, 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth.
Significant knowledge decay from posttest to delayed posttest (p<.05).	Only 3-fourth, 4-fourth and 4-fifth*.

*Marginally significant.

5.3 Written-word form and meaning recognition: The influence of treatment-related factors

5.3.1 After-viewing activity type

In order to explore the role of after-viewing activity type (see descriptive statistics in Table 82), a compound symmetry structure GLMM (binary logistic regression) was performed with student identification as subjects, as well as time and word as repeated measures. The dichotomous values (at item level) were entered into the model as target variable, while activity type, year level, time and all possible two-way and three-way interactions between these variables were included as factors. By following a step back procedure, the non-significant interactions were removed one by one until obtaining the best fitted model. As shown in Table 83, the results showed non-significant effects for activity type (F (1, 116)= 1.077, p= .302), but statistically significant effects for time (F (2, 9250)= 206.117, p< .001), year level (F (1, 86)= 7.603, p= .007), as well as a significant interaction between activity type and time (F (2, 8954)= 5.932, p= .003). The Bonferroni pairwise comparisons (see Table 84) showed that both activity types led to significant gains from pretest to posttest, and from pretest to delayed posttest. However, the gains were higher in the case of the construction-focused activities (intentional condition) (see Figure 39). Although at pretest the difference between activity type groups approached significance (p=.052), the distance was significantly reduced at posttest (p=.599) and delayed-posttest (p=.678) (see Figure 39).

Table 82.

WWFMR: Descriptive statistics per activity type

		Pretest		Posttest		Delayed posttest	
		Mean	(SD)	Mean	(SD)	Mean	(SD)
Activity type	Meaning focused	12.25	(8.09)	19.10	(8.12)	18.29	(7.80)
	Construction- focused	10.03	(6.21)	18.84	(8.03)	18.14	(7.52)

Table 83.

WWFMR: The influence of activity type on learners' outcomes

Source	F	df1	df2	Sig.
Corrected Model	74.614	6	313	.000
Level	7.603	1	86	.007
Time	206.117	2	9250	.000
Activity type	1.077	1	116	.302
Activity type * Time	5.932	2	8954	.003

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

Table 84.

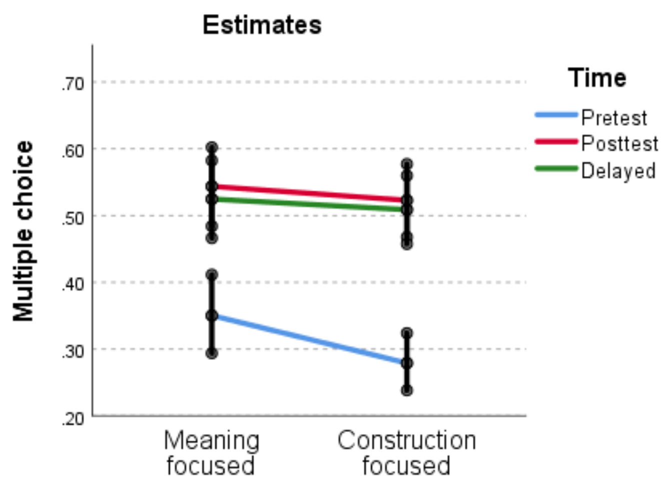
WWFMR: Activity type and time pairwise contrasts

Category	Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Meaning focused	Pretest - Posttest	-.194	.013	-14.761	12773	.000	-.225	-.162
	Pretest - Delayed	-.174	.011	-16.028	12773	.000	-.199	-.150
	Posttest - Delayed	.019	.006	3.161	12773	.002	.007	.031
Construction focused	Pretest - Posttest	-.244	.019	-12.993	1937	.000	-.289	-.199
	Pretest - Delayed	-.230	.016	-14.380	2066	.000	-.266	-.194
	Posttest - Delayed	.014	.006	2.207	12773	.027	.002	.026
Pretest	Meaning- Construction	.071	.036	1.957	136	.052	-.001	.143
Posttest	Meaning- Construction	.021	.040	.527	135	.599	-.058	.099
Delayed	Meaning- Construction	.016	.038	.416	126	.678	-.060	.091

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 39.

WWFMR: Interaction between activity type and time



Then, a new model was built to further explore the influence of activity type by adding vocabulary knowledge (EFL PVT) as a covariate. Again, a compound symmetry structure GLMM (binary logistic regression) was performed with student identification as subjects, as well as time and word as repeated measures. The dichotomous values were entered into the model as target variable, while activity type, year level, time, all possible two-way and three-way interactions between these variables, and vocabulary knowledge were included as factors. The best fitted model was obtained by a backward elimination procedure. As shown in Table 85, the results confirmed the findings of the previous model and yielded a more evident advantage for construction-focused activities (see Table 86 and Figure 40). In addition, the results suggested that the use of construction-focused activities led to higher retention (see Table 86).

Table 85.

WWFMR: The influence of activity type on learners' outcomes with vocabulary knowledge as covariate

Source	F	df1	df2	Sig.
Corrected Model	87.519	6	340	.000
Time	171.361	2	6056	.000
Activity type	.020	1	107	.888
Vocabulary Knowledge	192.163	1	94	.000
Activity type * Time	3.807	2	6672	.022

Probability distribution: Binomial
Link function: Logit a. Target: Multiple choice

Table 86.

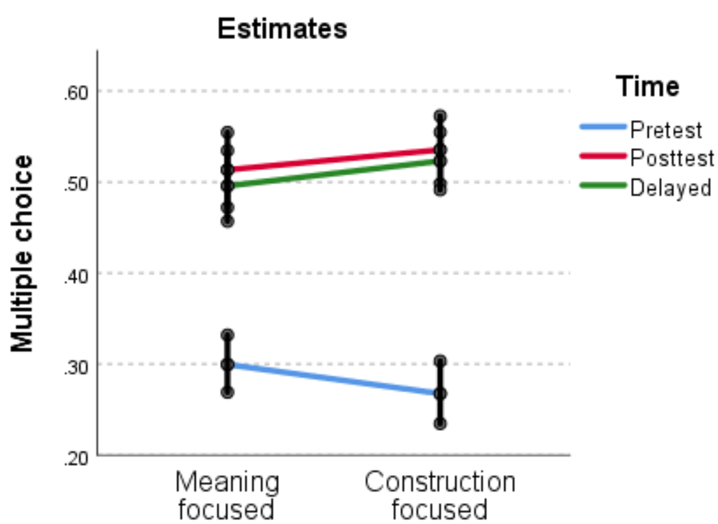
WWFMR: Time pairwise contrasts by activity type when including vocabulary knowledge as covariate

Activity type	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Meaning focused	Pretest - Posttest	-.214	.015	-14.227	1949	.000	-.250	-.178
	Pretest - Delayed	-.196	.013	-15.015	2104	.000	-.226	-.167
	Posttest - Delayed	.018	.006	2.979	11909	.003	.006	.029
Construction focused	Pretest - Posttest	-.268	.022	-12.452	2467	.000	-.319	-.216
	Pretest - Delayed	-.256	.019	-13.614	2683	.000	-.298	-.214
	Posttest - Delayed	.012	.007	1.664	10618	.096	-.002	.027

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 40.

WWFMR: Interaction between activity type and time when adding vocabulary knowledge as covariate



5.3.2 Viewing distribution

As in written-word form recall, separate models were calculated to assess the influence of viewing distribution in year 4 (1-4 episodes a week), and in the groups that watched either two or four episodes a week (in fourth and fifth grade). These models were fitted separately since 1-fourth and 3-fourth did not have a counterpart in year 5. To start with, a compound symmetry structure GLMM (binary logistic regression) was performed with fourth graders by fitting a model with student identification as subjects, as well as time and word as repeated measures. The dichotomous values were entered into the model as target variable, while viewing distribution, time and their interaction were included as factors. As shown in Table 87, the results showed significant effects for time (F (2, 4017)= 125.541, p< .001), while the interaction between viewing distribution and time was only found to be marginally significant (F (6, 5795)= 1.942, p= .070). The Bonferroni pairwise contrasts in Table 88 indicated that all the groups improved significantly over time, but also suggested that 1-fourth obtained greater gains from pretest to delayed posttest. Likewise, the results confirmed that watching one or two episodes a week led to slightly higher retention from posttest to delayed posttest. Considering that these results may also be influenced by each group's proficiency level, a new model was fitted with vocabulary knowledge (EFL PVT) as covariate (see Table 89).

Table 87.

WWFMR: The influence of viewing distribution in fourth graders

Source	F	df1	df2	Sig.
Corrected Model	29.091	11	369	.000
Viewing distribution	.419	3	78	.740
Time	125.541	2	4017	.000
Viewing distribution * Time	1.942	6	5795	.070

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

Table 88.

WWFMR: Time pairwise contrasts by viewing distribution in fourth graders

Viewing distribution	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
Once a week	Pretest - Posttest	-.238	.025	-9.421	2326	.000	-.298	-.177
	Pretest - Delayed	-.240	.025	-9.612	2380	.000	-.296	-.184
	Posttest - Delayed	-.002	.011	-.232	7476	.817	-.023	.018
Twice a week	Pretest - Posttest	-.181	.033	-5.512	794	.000	-.254	-.107
	Pretest - Delayed	-.160	.025	-6.489	654	.000	-.219	-.101
	Posttest - Delayed	.021	.013	1.618	6846	.106	-.004	.046
Three times a week	Pretest - Posttest	-.160	.016	-9.752	7476	.000	-.199	-.121
	Pretest - Delayed	-.141	.016	-9.003	7476	.000	-.176	-.106
	Posttest - Delayed	.019	.009	2.181	7476	.029	.002	.036
Four times a week	Pretest - Posttest	-.206	.032	-6.500	817	.000	-.277	-.135
	Pretest - Delayed	-.170	.022	-7.674	882	.000	-.224	-.117
	Posttest - Delayed	.035	.016	2.166	5193	.030	.003	.067

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 89.

WWFMR: The influence of viewing distribution in fourth graders with vocabulary knowledge as covariate

Source	F	df1	df2	Sig.
Corrected Model	34.256	12	360	.000
Vocabulary knowledge	91.074	1	71	.000
Viewing distribution	.825	3	76	.484
Time	111.302	2	3716	.000
Viewing distribution * Time	1.733	6	4585	.109

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

As shown in Table 89, the compound symmetry structure GLMM (binary logistic regression) that included vocabulary knowledge as covariate indicated that neither viewing distribution (F (3, 76)= .825, p= .484), nor the interaction between viewing distribution and time (F (6, 4585)= 1.733, p= .109) reached statistical significance. Likewise, when fitting a model without the interaction between viewing distribution and time, the results confirmed that viewing distribution was no longer significant. When exploring the Bonferroni pairwise contrasts (see Appendix 37) the results suggested a similar tendency, that is watching one episode a week led to slightly higher gains from pretest to delayed posttest, and resulted in slightly higher retention. Still, this outcome was exclusively detected in 1-fourth since there did not seem to be a significant difference between the performance of 2-fourth and 4-fourth. All in all, the significant interaction between viewing distribution and time that was obtained in the first model might be either attributed to the slightly higher vocabulary knowledge of 1-fourth or the fact that viewing distribution effects were not robust, so they were overridden by the effects of vocabulary knowledge.

As mentioned above, a separate model was fitted to assess the influence of viewing distribution in fourth and fifth grade. Specifically, the groups that watched two and four episodes a week were included since only these time distributions were implemented in both year levels. A compound symmetry structure GLMM (binary logistic regression) was performed with these groups by fitting a model with student identification as subjects, as well as time and word as repeated measures. The dichotomous values were included as target variable, while viewing distribution, year level, time and all possible interactions were entered as factors. After following a backward elimination procedure, the results confirmed that viewing distribution did not affect the results (see Table 90). Then, the same outcomes were observed when fitting a new model with vocabulary knowledge as covariate (see Appendix 38).

Table 90.

WWFMR: The influence of viewing distribution in fourth and fifth graders (two and four episodes a week)

Source	F	df1	df2	Sig.
Corrected Model	60.513	6	243	.000
Viewing distribution	.014	1	70	.908
Level	6.642	1	92	.012
Time	157.611	2	6481	.000
Level * Time	3.539	2	5945	.029

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

5.3.3 The influence of treatment-related factors: Summary of findings

This section assessed the influence of treatment-related factors (i.e. after-viewing activity type and viewing distribution) on vocabulary learning at the level of written-word form and meaning recognition. The key findings are as follows:

- The analyses did not yield significant main effects for viewing distribution.
- Although both types of activities resulted in significant improvement over time, the use of construction-focused activities (intentional condition) led to higher gains.
- Learners' benefitted from the treatment regardless of the number of episodes they had to watch a week. However, in year four, it seemed that watching one episode a week led to slightly higher gains from pretest to delayed posttest and prevented, to a certain extent, word-knowledge decay. However, the advantage of 1-fourth was not robust and was only detected when vocabulary knowledge was not entered into the same model.

5.4 Written-word form and meaning recognition: The influence of cognitive and language-related factors

This section focuses on the analyses that explored the influence of cognitive and language-related factors on written-word form and meaning recognition. These analyses only included the experimental groups from school 1, that is 2-fourth, 3-fourth, 4-fourth, 2-fifth and 4-fifth. It is also important to mention that collinearity tests were performed before running the tests to ensure that all the independent variables could be included in the analyses (Pallant, 2016).

5.4.1 Cognitive factors

In order to assess the influence of cognitive factors on written-word form and meaning recognition, a series of Generalized Linear Mixed Models (binary logistic regression) with repeated measures (time) compound-symmetry structure were calculated. The model was built with learners' dichotomous scores (at item level). The fixed effects included in the analyses were class, time, PSTM, complex working memory, visual processing speed (high vs. low), and all possible two-way interactions. A backward (or step back) elimination procedure was used to determine the best fitted model (see Table 91). Thus, the non-significant interactions and factors were removed from the model one by one. As summarized in Table 92, the results revealed significant main effects for PSTM (F (1, 128)= 8.085, p= .005), complex WM (F (1,112)= 6.152, p= .015), and time (F (2, 2028)= 141.420, p< .001), as well as significant interaction between visual processing speed and time (F (2,3913)=

4.678, p= .009). As for PSTM, the exponential coefficient indicated that when PSTM scores increased by one, the odds of a correct response increased by 14%. Similarly, learners' accuracy in written-word form and meaning recognition increased by 11% per each additional point in complex WM (backward digit span test). With respect to the significant interaction between time and visual processing speed, the results suggest that higher visual processing speed fostered greater retention from posttest to delayed posttest (see Table 93 and Figure 41).

Table 91.

WWFMR: Best fitted model obtained to assess the influence of cognitive factors on fourth and fifth graders' outcomes

Model Term	Coef	Std. Error	t	Sig.	95% CI Lower	95% CI Upper	Exp (Coef)	95% CI for Exp(Coef) Lower	95% CI for Exp(Coef) Upper
Intercept	-2.006	.5229	-3.837	.000	-3.040	-.972	.134	.048	.378
PSTM	.132	.0464	2.843	.005	.040	.224	1.141	1.041	1.251
Complex WM	.107	.0432	2.480	.015	.022	.193	1.113	1.022	1.212
Pretest	-.813	.0775	-10.488	.000	-.965	-.661	.444	.381	.516
Posttest	.014	.0299	.480	.631	-.044	.073	1.014	.957	1.076
Delayed	0[b]	.	.	.	.	.	.	.	.
3-fourth	-.339	.2178	-1.555	.122	-.769	.091	.713	.464	1.096
4-fourth	-.234	.2525	-.928	.355	-.734	.265	.791	.480	1.304
2-fourth	-.385	.2393	-1.610	.110	-.859	.088	.680	.424	1.092
2-fifth	.239	.2381	1.004	.319	-.236	.714	1.270	.790	2.042
4-fifth	0[b]	.	.	.	.	.	.	.	.
VPS[i]	-.053	.1615	-.328	.743	-.374	.268	.948	.688	1.307
[Pretest]*[VPS][i]	-.041	.1013	-.408	.683	-.240	.157	.959	.787	1.170
[Posttest]* [VPS][i]	.104	.0365	2.850	.004	.032	.176	1.110	1.033	1.192
[Delayed]* [VPS][i]	0[b]	.	.	.	.	.	.	.	.

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice
b. This coefficient is set to zero because it is redundant.
[i] VPS=visual processing speed; Reference category= High VPS

Table 92.

WWFMR: The influence of cognitive factors on fourth and fifth graders' scores

Source	F	df1	df2	Sig.
Corrected Model	37.534	11	207	.000
PSTM	8.085	1	128	.005
Complex WM	6.152	1	112	.015
Time	141.420	2	2028	.000
Class	1.971	4	104	.104
Visual processing speed	.036	1	87	.851
Time * Visual processing speed	4.678	2	3913	.009

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

Table 93.

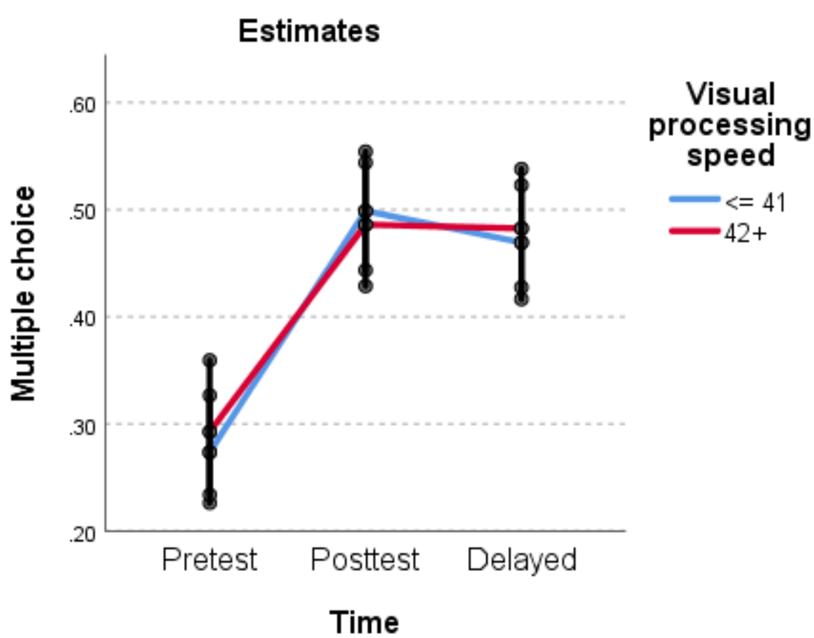
WWFMR: Time pairwise contrasts by visual processing speed groups

Visual processing speed	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
<= 41	Pretest - Posttest	-.225	.016	-13.732	2797	.000	-.265	-.186
	Pretest - Delayed	-.196	.014	-13.712	2984	.000	-.228	-.164
	Posttest - Delayed	.030	.005	5.637	10032	.000	.019	.040
42+	Pretest - Posttest	-.194	.021	-9.135	2972	.000	-.244	-.143
	Pretest - Delayed	-.190	.015	-12.387	2667	.000	-.224	-.156
	Posttest - Delayed	.004	.007	.480	7148	.631	-.011	.018

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 41.

WWFMR: Interaction between visual processing speed and time



5.4.2 Language-related factors

In order to assess the influence of language-related factors on written-word form and meaning recognition, a series of Generalized Linear Mixed Model (binary logistic regression) with repeated measures (time) compound-symmetry structure were calculated. The models were built with learners' scores at pretest and posttest by setting 36 (maximum score) as denominator. To start with, class, time, and only L2-related factors were entered into the model as independent variables: vocabulary knowledge, English segmentation, ER efficacy and listening skills. No interactions were included in order to compare the contribution of each language-related factor to the learning of form-meaning links. The non-significant main effects were removed from the model one by one until the best fitted model was obtained (see Table 94). The results yielded significant main effects for vocabulary knowledge (F (1, 91)= 108.143, p< .001) and listening skills (F (1, 164)= 10.107, p= .002). The exponential coefficients in Table 94 indicate that the odds of a correct response increased by 8.9% per each additional word known in the EFL picture vocabulary test, and by 4.6% per each additional point at the listening test. Thus, among the L2-related factors, vocabulary knowledge appeared to be the strongest predictor of written-word form and meaning recognition.

Table 94.

WWFMR: Best fitted model obtained to assess the influence of L2-related factors on learners' scores (without interactions).

Model Term	Coef	Std. Error	t	Sig.	95% CI Lower	95% CI Upper	Exp (Coef)	95% CI for Exp(Coef) Lower	95% CI for Exp(Coef) Upper
Intercept	-1.916	.1398	-13.712	.000	-2.193	-1.640	.147	.112	.194
Pretest	-.914	.0548	-16.684	.000	-1.023	-.804	.401	.359	.448
Posttest	.077	.0210	3.647	.000	.035	.118	1.080	1.036	1.125
Delayed	0^b	.	.	.	.	.	.	.	.
Vocabulary knowledge	.085	.0082	10.399	.000	.069	.102	1.089	1.071	1.107
Listening skills	.045	.0140	3.179	.002	.017	.072	1.046	1.017	1.075

Probability distribution: Binomial
Link function: Logit
a. Target: Written-word form and meaning recognition/36
b. This coefficient is set to zero because it is redundant.

Then, a new model was built with L2-related factors and interactions: class, time, vocabulary knowledge, English segmentation, ER efficacy, listening skills, and all possible two-way interactions. By following a step back procedure, the analyses revealed that neither ER efficacy nor class contributed to the model significantly, therefore they were removed from the best fitted model (see Table 95). As summarized in Table 96, the model showed a significant main effect for vocabulary knowledge (F (1, 73)= 77.046, p< .001), indicating that the odds of recognizing a written-word form and its meaning correctly increased by 8.7% per each correct answer at the EFL Picture Vocabulary test. The results also revealed a marginally significant main effect for listening skills (F (1, 130)= 3.617, p= .059), showing that when students' score at the listening skills test increased by one, the odds of recognizing a target word and its meaning correctly increased by 3%. Time was also found to be significant (F (2, 86)= 22.137, p< .001), as well as the interaction between segmentation in English and time (F (2,41)= 6.241, p= .004). As shown in Figure 42, the strength of the relationship between segmentation in English and time increased over time.

Table 95.

WWFMR: Best fitted model obtained to assess the influence of L2-related factors.

					95% CI		Exp	95% CI for Exp(Coeff)	
Model Term	Coeff	SE	t	Sig.	Lower	Upper	(Coeff)	Lower	Upper
Intercept	-1,916	,1443	-13,273	,000	-2,202	-1,630	,147	,111	,196
Vocabulary knowledge	,083	,0095	8,778	,000	,064	,102	1,087	1,066	1,107
English segmentation	,007	,0044	1,564	,122	-,002	,016	1,007	,998	1,016
Listening skills	,030	,0157	1,902	,059	-,001	,061	1,030	,999	1,063
Pretest	-,584	,0953	-6,128	,000	-,775	-,393	,558	,461	,675
Posttest	,012	,0477	,260	,796	-,082	,107	1,012	,921	1,113
Delayed posttest	0[b]	.	.	.	.	.	.	.	.
English seg. *[Pretest]	-,011	,0033	-3,334	,003	-,018	-,004	,989	,982	,996
English seg. *[Posttest]	,002	,0018	1,114	,271	-,002	,006	1,002	,998	1,006
English seg. *[Delayed]	0[b]	.	.	.	.	.	.	.	.

Probability distribution: Binomial
Link function: Logit
a. Target: Written-word form and meaning recognition/36
b. This coefficient is set to zero because it is redundant.

Table 96.

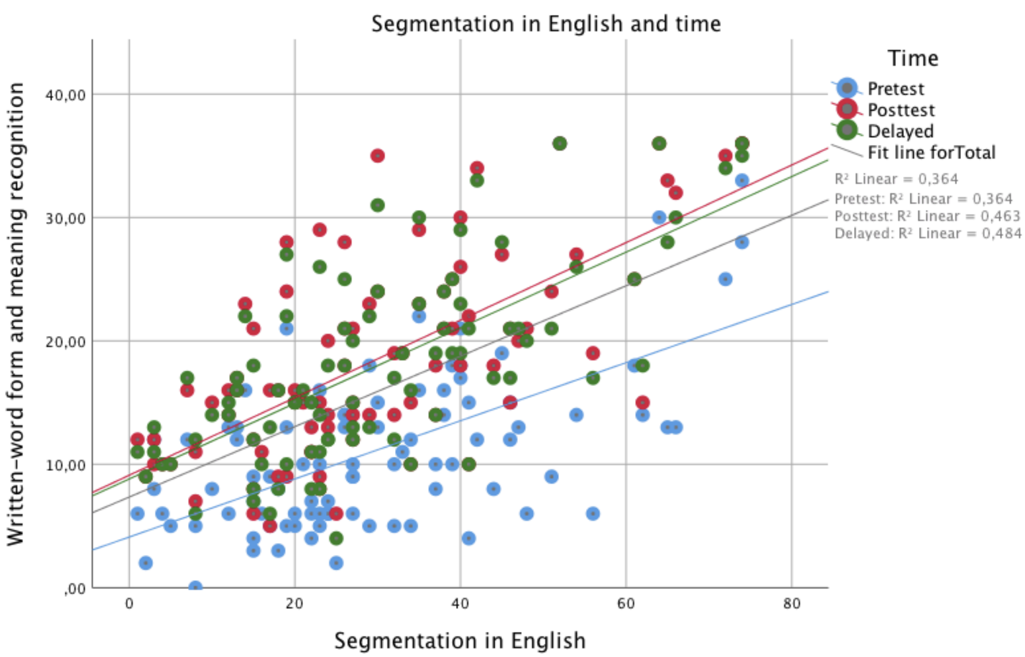
WWFMR: The effects of L2-related factors

Source	F	df1	df2	Sig.
Corrected Model	69,351	7	106	,000
Vocabulary	77,046	1	73	,000
English segmentation	,819	1	86	,368
Listening skills	3,617	1	130	,059
Time	22,137	2	86	,000
English segmentation * Time	6,241	2	41	,004

Probability distribution: Binomial
Link function: Logit
a. Target: Written-word form and meaning recognition/36

Figure 42.

WWFMR: Relationship between English segmentation and time



As for the L1-related factors, we first built a model without interactions to calculate the contribution of each factor to learners' performance over time. Therefore, the following fixed factors were entered into the model: class, time, Spanish text segmentation and SR efficacy. The results revealed significant main effects for both L1-related factors: Spanish text segmentation (F (1, 140)= 6.735, p= .010) and SR efficacy (F (1, 193)= 14.369, p< .001).

The exponential coefficients indicated that learners' scores increased by 1.6% per each additional word recognized at the Spanish text segmentation test, and by 0.7% per each additional point at the SR efficacy test (see Table 97).

Table 97.

WWFMR: The influence of L1-related factors on learners' scores (without interactions)

Model Term	Coef	Std. Error	t	**Sig.**	95% CI Lower	95% CI Upper	**Exp (Coef)**	95% CI for Exp(Coef) Lower	95% CI for Exp(Coef) Upper
Intercept	-1.096	.1893	-5.786	**.000**	-1.469	-.722	**.334**	.230	.486
Pretest	-.833	.0510	-16.345	**.000**	-.936	-.731	**.435**	.392	.481
Posttest	.065	.0186	3.509	**.001**	.029	.102	**1.068**	1.029	1.107
Delayed	0^b	.	.	**.**	.	.	**.**	.	.
Spanish text segmentation	.016	.0061	2.595	**.010**	.004	.028	**1.016**	1.004	1.028
SR efficacy	.007	.0017	3.791	**.000**	.003	.010	**1.007**	1.003	1.010

Probability distribution: Binomial
Link function: Logit
a. Target: Written-word form and meaning recognition/36
b. This coefficient is set to zero because it is redundant.

Then, a new model was built to study the effects of L1-related factors in comparison with the L2 variables that were found to be significant in the analyses above. The independent variables entered into the model were as follows: class, time, vocabulary knowledge, English segmentation, listening skills, Spanish segmentation, SR efficacy and all possible two-way interactions. By following a step back procedure to obtain the best fitted model (see Table 98), Spanish segmentation had to be removed since it was not shown to contribute to the model significantly (p>.05). As summarized in Table 99, the GLMM revealed significant main effects for vocabulary knowledge (F (1, 56)= 73.330, p< .001), listening skills (F (1, 133)= 3.982, p= .048) and time (F (2, 83)= 140.160, p< .001). In the presence of SR efficacy, English segmentation was only found to be marginally significant (F (1, 108)= 3.591, p= .061). As for SR efficacy, this factor was found to interact with class significantly (F (4, 68)= 2,943, p= .026), confirming that 2-fourth and 3-fourth relied more on Spanish reading efficacy to benefit from the treatment (see Figure 43).

Table 98.

WWFMR: Best fitted model obtained to assess the influence of language related factors.

Model Term	Coeff.	SE	t	**Sig.**	95% CI Lower	95% CI Upper	**Exp (Coeff.)**	95% CI for Exp(Coeff.) Lower	95% CI for Exp(Coeff.) Upper
Intercept	-2,131	,3006	-7,090	**,000**	-2,734	-1,528	**,119**	,065	,217
3-fourth	-,045	,4831	-,094	**,925**	-1,009	,918	**,956**	,364	2,505
4-fourth	,584	,3508	1,665	**,099**	-,112	1,281	**1,794**	,894	3,599
2-fourth	-,036	,3590	-,099	**,922**	-,752	,681	**,965**	,471	1,977
2-fifth	-,326	,2992	-1,090	**,279**	-,922	,270	**,722**	,398	1,310
4-fifth	0[b]	.	.	**.**	.	.	**.**	.	.
Vocabulary knowledge	,077	,0090	8,563	**,000**	,059	,095	**1,080**	1,061	1,099
SR efficacy	,001	,0018	,535	**,594**	-,003	,005	**1,001**	,997	1,005
Listening skills	,028	,0141	1,996	**,048**	,000	,056	**1,029**	1,000	1,058
English segmentation	,008	,0045	1,895	**,061**	,000	,017	**1,009**	1,000	1,017
Pretest	-,925	,0570	-16,210	**,000**	-1,039	-,810	**,397**	,354	,445
Posttest	,067	,0209	3,223	**,001**	,026	,109	**1,070**	1,027	1,115
Delayed posttest	0[b]	.	.	**.**	.	.	**.**	.	.
SR efficacy*[3-fourth]	,006	,0054	1,113	**,269**	-,005	,017	**1,006**	,995	1,017
SR efficacy*[4-fourth]	-,004	,0038	-,950	**,346**	-,011	,004	**,996**	,989	1,004
SR efficacy*[2-fourth]	,002	,0032	,568	**,572**	-,005	,008	**1,002**	,995	1,008
SR efficacy*[2-fifth]	,006	,0022	2,514	**,013**	,001	,010	**1,006**	1,001	1,010
SR efficacy*[4-fifth]	0[b]	.	.	.	.	.	.	.	.

Probability distribution: Binomial
Link function: Logit
a. Target: Written-word form and meaning recognition/36
b. This coefficient is set to zero because it is redundant.

Table 99.

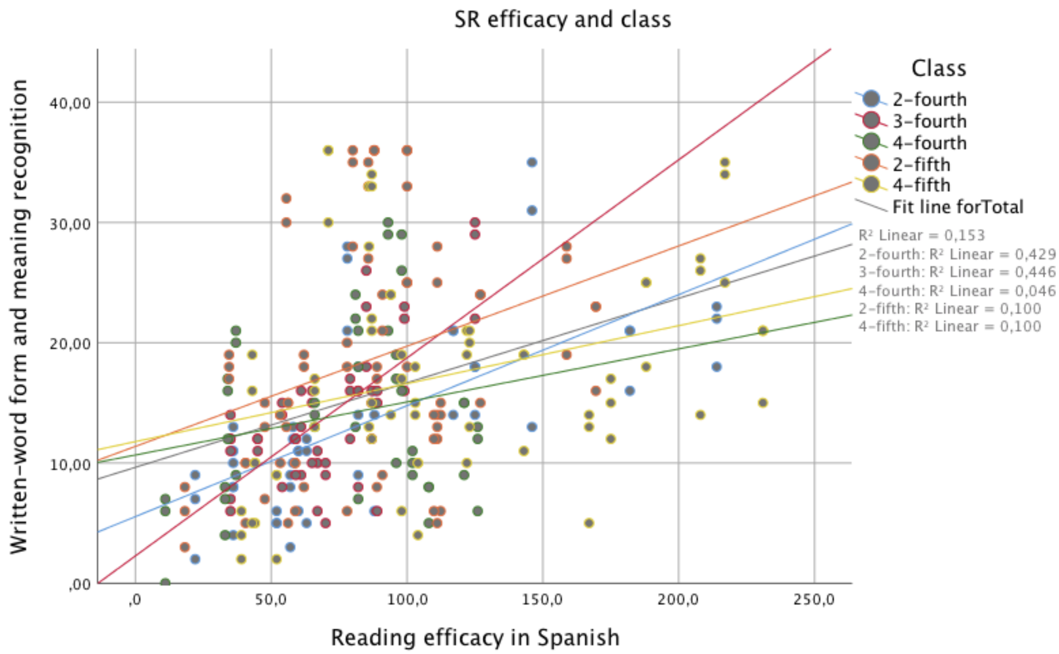
WWFMR: The effects of language-related factors

Source	F	df1	df2	Sig.
Corrected Model	38,782	14	101	,000
Class	2,549	4	91	,044
Vocabulary knowledge	73,330	1	56	,000
Spanish reading efficacy	3,625	1	72	,061
Listening skills	3,982	1	133	,048
English segmentation	3,591	1	108	,061
Time	140,160	2	83	,000
Class * SR efficacy	2,943	4	68	,026

Probability distribution: Binomial
Link function: Logit
a. Target: Written-word form and meaning recognition/36

Figure 43.

WWFMR: Relationship between SR efficacy and class



All in all, among the language-related factors, L2 vocabulary knowledge and listening skills were found to be stronger predictors of vocabulary learning at the level of written-word form and meaning recognition. Concerning SR efficacy, this factor appeared to play a more significant role in the case of the fourth graders that had to watch two or three episodes a week. As suggested in Figure 43, the groups did not rely on SR to the same extent, which is a finding that might be associated to learners' age and proficiency level, as well as the number of episodes they had to watch a week.

5.4.3 The influence of cognitive and language-related factors: Summary of findings

Taken together, the results reported in this section indicate that:

- All the cognitive factors played a significant role in vocabulary learning at the level of written-word form and meaning recognition: PSTM, complex WM and visual processing speed.

- Within the group of L2-related factors, the analyses indicated that vocabulary knowledge and listening skills were the strongest predictors of learners' progress over time. As for English text segmentation, this factor was only found to be marginally significant when fitting a model with SR efficacy.
- The extent to which SR efficacy predicted learning depended on year level and viewing distribution. Specifically, the fourth graders that watched either two or three episodes a week appeared to rely significantly more on SR efficacy to benefit from the treatment.

5.5 Written-word form and meaning recognition: The influence of treatment, cognitive and language-related factors

All the factors found to be significant in previous analyses were entered into a new model as independent variables to examine which ones remain as significant predictors. To this aim, we ran a series of compound-symmetry structure GLMMs (binary logistic regressions) with student identification as subjects, together with time and word as repeated measures. Learners' scores at item level (binomial distribution) were set as outcome variable, while the fixed factors selected for these analyses were as follows: viewing distribution, activity type, year level, time, vocabulary knowledge, listening skills, SR efficacy, English text segmentation, PSTM, complex working memory, visual processing speed and all possible interactions. In these analyses, only the participants that watched two or four episodes a week were included since these were the only viewing time distributions that were implemented in both year levels.

The step back procedure indicated that activity type and PSTM were no longer significant predictors of written-word form and meaning recognition. After eliminating the non-significant interactions and main effects, the best fitted model displayed in Table 100 was obtained. The results revealed significant main effects for year level (F (1, 89)= 4.020, p= .048), complex WM (F (1, 45)= 5.159, p= .028), vocabulary knowledge (F (1, 46)= 79.095, p< .001), listening skills (F (1, 94)= 7.663, p= .007), and time (F (2,1375)= 35.511, p< .001) (see Table 101). The exponential coefficients in Table 100 indicated that among the continuous variables that were found to have a significant main effect, vocabulary knowledge (9,9%) was the strongest predictor, followed by complex WM (7,9%) and listening skills

(4,5%). In addition, the analyses yielded significant interactions that confirmed the outcomes obtained in previous analyses: English segmentation and time (F (2, 414)= 7.354, p= .001) suggesting that English segmentation played a greater role at posttest and delayed posttest; Time and visual processing speed (F (2, 1222)= 4.830, p= .008) implying that higher visual processing speed leads to greater retention; and a triple interaction between viewing distribution, level and SR efficacy (F (3, 48)= 5.088, p= .004), corroborating that 2-fourth relied more on SR efficacy to benefit from the treatment. In addition, the significant interaction between SR efficacy and time indicated that the relationship between SR efficacy and written-word form and meaning recognition was stronger at posttest and delayed posttest (see Figure 44). Taken together, the analyses suggest that cognitive and language-related factors were stronger predictors of receptive form-meaning mapping than the treatment-related factors.

The key results obtained in this section are summarized in Table 102.

Table 100.

WWFMR: Best fitted model built to assess the influence of treatment, cognitive, and language-related factors.

Model Term	Coef	Std. Error	t	Sig.	95% CI Lower	95% CI Upper	Exp (Coef)	95% CI for Exp(Coef) Lower	95% CI for Exp(Coef) Upper
Intercept	-2.937	.4187	-7.015	.000	-3.784	-2.091	.053	.023	.124
Twice a week	-.322	.2417	-1.334	.187	-.805	.161	.724	.447	1.174
Four times a week	0[b]	.	.	.	.	.	.	.	.
Year 4	.459	.2289	2.005	.048	.004	.914	1.582	1.004	2.493
Year 5	0[b]	.	.	.	.	.	.	.	.
Complex WM	.076	.0333	2.271	.028	.009	.143	1.079	1.009	1.153
Vocabulary knowledge	.094	.0106	8.894	.000	.073	.115	1.099	1.075	1.122
English segmentation	.003	.0049	.705	.483	-.006	.013	1.003	.994	1.013
SR efficacy	.000	.0018	.054	.957	-.003	.004	1.000	.997	1.004
Listening skills	.044	.0158	2.768	.007	.012	.075	1.045	1.012	1.078
Pretest	-.706	.1061	-6.648	.000	-.914	-.497	.494	.401	.608
Posttest	-.096	.0586	-1.631	.103	-.210	.019	.909	.810	1.020
Delayed	0[b]	.	.	.	.	.	.	.	.
VPS[i]	-.018	.1065	-.173	.863	-.232	.195	.982	.793	1.215
English seg.*[Pretest]	-.015	.0041	-3.708	.000	-.023	-.007	.985	.977	.993
English seg.*[Posttest]	.004	.0025	1.700	.090	-.001	.009	1.004	.999	1.009
English seg.*[Delayed]	0[b]	.	.	.	.	.	.	.	.
SR efficacy*[Pretest]	.003	.0012	2.607	.010	.001	.006	1.003	1.001	1.006
SR efficacy*[Posttest]	.000	.0006	-.581	.561	-.001	.001	1.000	.999	1.001

SR efficacy*[Delayed]	0[b]	.	.	.	.	.	.	.	.
[Pretest]*[VPS][i]	-.251	.1148	-2.190	.029	-.477	-.026	.778	.621	.974
[Posttest]* [VPS][i]	.134	.0451	2.965	.003	.045	.222	1.143	1.046	1.249
[Delayed]* [VPS][i]	0[b]	.	.	.	.	.	.	.	.
SR efficacy*[2 a week]*[Year 4]	-.001	.0031	-.266	.792	-.007	.005	.999	.993	1.005
SR efficacy*[2 a week]*[Year 5]	.005	.0021	2.547	.013	.001	.009	1.005	1.001	1.009
SR efficacy*[4 a week]*[Year 4]	-.002	.0028	-.729	.470	-.008	.004	.998	.992	1.004
SR efficacy*[4 a week]*[Year 5]	0[b]	.	.	.	.	.	.	.	.

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice
b. This coefficient is set to zero because it is redundant.
[i] VPS=Visual processing speed; Reference category=High VPS

Table 101.

WWFMR: The influence of treatment, cognitive and language-related factors

Source	F	df1	df2	Sig.
Corrected Model	53.271	19	166	.000
Viewing distribution	1.779	1	62	.187
Level	4.020	1	89	.048
Complex WM	5.159	1	45	.028
Vocabulary knowledge	79.095	1	46	.000
English segmentation	.001	1	97	.973
SR efficacy	1.683	1	66	.199
Listening skills	7.663	1	94	.007
Time	35.511	2	1375	.000
Visual processing speed	.316	1	48	.577
English segmentation * Time	7.354	2	414	.001
SR efficacy * Time	4.000	2	551	.019
Time * Visual processing speed	4.830	2	1222	.008
Viewing distribution * Level * SR efficacy	5.088	3	48	.004

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

Figure 44.

Interaction between SR efficacy and time

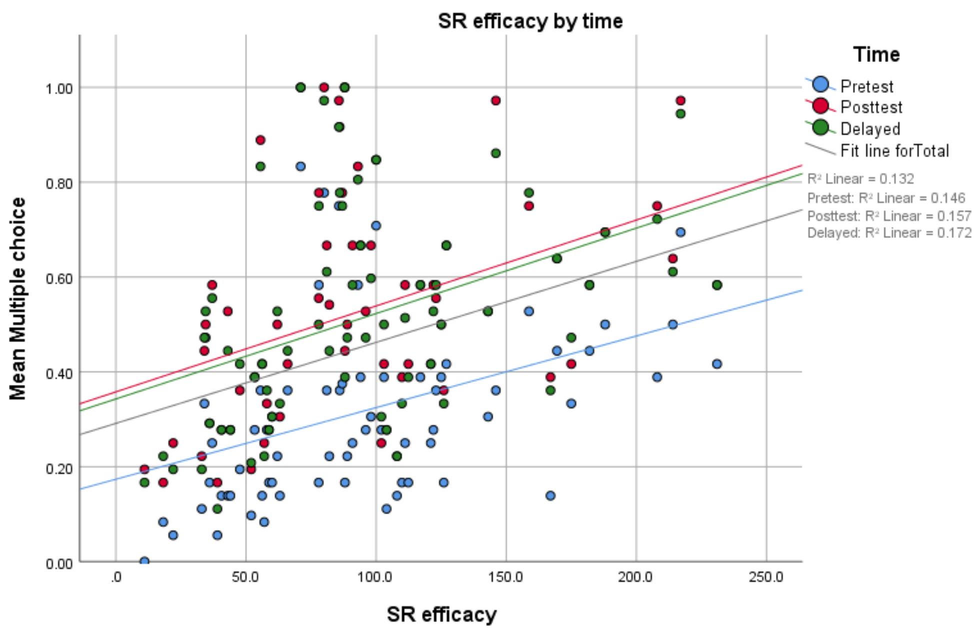


Table 102.

Summary: Predictors of written-word form and meaning recognition

Factors	Outcomes
Viewing distribution	***Significant*** in fourth graders and in interaction with SR efficacy. The results suggest that watching one episode a week led to slightly higher gains from pretest to delayed posttest, and fostered greater retention from posttest to delayed posttest. Yet, the results were not robust. As for the interaction with SR efficacy, 2-fourth and 3-fourth seemed to have relied more on SR efficacy. Therefore, shorter lags between episodes (four times a week) might have moderated the effects of SR efficacy and facilitated input processing in fourth graders.
Year level	***Significant***. Both year levels followed similar trajectories (significant improvement from pretest to posttest, and from pretest to delayed posttest, as well as significant decrease from posttest to delayed posttest); however, year-5 participants scored significantly higher at the three testing times.
Activity type	***Significant*** in interaction with time. Both activity types were conducive to significant gains over time. However, the use of construction-focused activities seemed to be more beneficial as regards written-word form and meaning recognition. Yet, the significant effects of activity type

	disappeared when fitting a model with cognitive and language-related factors.
Time	***Significant.***
Vocabulary knowledge	***Significant.*** It seemed to be the most important predictor of written-word and meaning recognition.
English segmentation	***Significant.*** Its relationship with written-word form and meaning recognition was shown to increase over time.
Spanish segmentation	***Significant.*** However, it had a weak influence on the outcomes and its effects were overridden by the L2-related factors.
English reading efficacy	***Non-significant.***
Spanish reading efficacy	***Significant*** in interaction with class, viewing distribution and year level. It seemed to play a more important role in year 4, especially when the distance between episodes was longer.
Listening skills	***Significant.***
PSTM	***Significant*** but its effect disappeared when compared with language-related factors.
Complex working memory	***Significant.*** It seemed to be a stronger predictor of written-word form and meaning recognition.
Visual processing speed	***Significant*** in interaction with time. Higher visual processing speed led to greater retention from posttest to delayed posttest.

5.6 Written-word and meaning recognition: The influence of context and word-related factors

As in written-word form recall, the analyses in this section attempt to shed some light on the role of context and language-related factors in the recognition of form-meaning mapping. Specifically, it will examine the influence of word distribution (i.e. spacing effect), frequency of occurrence, regularity, length, and concreteness. Given that the target items could not be manipulated, it is important to consider that some of the categories were unbalanced as regards their number of items. Therefore, to control for this factor, the series of GLMMs (Binary logistic regressions) were conducted at item level. The information displayed in Table 103 indicates that there was great variability in relative gains among the target items and suggests that there was not a clear pattern as regards context and word characteristics. Therefore, the analyses below may throw light on how these variables interacted and affected word learnability.

Table 103.

WWFMR: Relative gains per word

Word	Relative gains %	Regularity[a]	Length[b]	Concreteness[c]	Frequency[d]	Nº of episodes[e]
Track	7.69	2	1	1	2	2
Lead	8.65	1	1	1	1	1
Wide	11.21	1	1	1	2	2
Web	12.37	2	1	1	2	1
Mud	15.56	1	1	2	1	2
Hairy	17.24	1	1	1	2	1
Wand	17.65	1	1	2	1	1
Costume	19.54	1	2	1	1	1
Trolley	20.00	1	2	2	1	1
Pleased	20.39	1	2	1	2	2
Mermaid	22.95	1	2	1	2	2
Stripy	23.23	1	1	2	1	1
Shell	25.97	1	1	2	2	1
Suitcase	26.21	1	2	2	1	1
Wobbly	27.36	1	1	1	2	2
Fairy	28.42	1	1	1	2	2
Useful	30.49	1	1	1	1	2
Careful	31.88	1	2	1	2	2
Pillow	32.65	1	1	2	1	1
Clever	33.33	2	1	1	2	2
Slipper	35.11	2	2	2	1	1
Puddle	37.37	1	1	2	2	2
Kitten	37.50	2	1	2	1	2
Bandage	37.70	1	2	2	1	1
Drop	37.80	2	1	1	2	1
Leaf	37.93	1	1	2	2	1
Wing	40.54	2	1	2	1	2
Forest	41.38	2	1	1	1	2
Busy	42.86	1	1	1	2	1
Cabbage	44.19	1	2	2	1	2
Handbag	45.68	2	2	2	2	1
Sausage	46.51	1	2	2	1	2
Sticky	50.82	2	1	1	1	2
Pea	51.22	1	1	2	2	2

Creaky	67.31	1	1	1	1	1
Fluffy	75.76	1	1	1	2	1

a 1= Less consistent with L1 patterns, 2= More consistent with L1 patterns.
b 1= Shorter words (<= 6 letters), 2= Longer words (7+).
c 1= Less concrete words (<4.62), 2= More concrete words (>4.63).
d 1= Less frequent (3-5), 2= More frequent (6+).
e 1= Repetitions concentrated in one episode, 2= Repetitions distributed in multiple episodes.

5.6.1 The role of word distribution in the outcomes

As explained earlier, half of the target words were encountered in a single video, while the other half were distributed in multiple episodes (2-3). This distinction was labelled as word distribution to compare massed and distributed encounters. Although these groups were comparable with respect to frequency of occurrence, an optimal comparison between these two conditions would include the same set of target words in each category. Still, as in written-word form recall, the analyses on word distribution were considered to be relevant for two main reasons. Firstly, the words encountered in multiple episodes were tested by the construction-focused activities after their first encounter. Thus, there might be an interaction between activity type and the words that were encountered in multiple episodes. Secondly, the evidence suggests that word frequency effects are higher in massed condition (Uchihara et al., 2019; Fievez et al., 2020), therefore, it would also be interesting to test this assumption. Thus, a series of repeated-measures (word and time) compound-symmetry structure GLMMs (binary logistic regression) with student identification as subjects were calculated in order to explore the two assumptions mentioned above. To this aim, FKS binary score (at item level) was set as target variable, while time, activity type, word distribution (one vs. multiple episodes), frequency of occurrence (3-5 vs. 6+ repetitions), and some interactions of interest were entered into the model as independent variables: 1) Word distribution*time, 2) Word distribution*time*activity type, 3) Word distribution*time*frequency, 4) Activity type*frequency*time, and 5) Activity type*frequency. The interactions that involved activity type and frequency (4 and 5) were included to test the assumption that test announcement (the completion of construction-focused activities in this case) increases learners' sensitivity to frequency effects (Uchihara et al., 2019). In addition, we added word concreteness and length as covariates. The non-significant interactions and factors were removed one by one from the analyses until obtaining the best fitted model. This was the case of the interactions between activity type and frequency, whose non-significant effects indicated that the use of construction-focused activities did not enhance the role of word repetitions in vocabulary learning.

As shown in Table 104, the results revealed significant main effects for time (F (2, 6867)= 199.790, p< .001), length (F (1, 2444)= 100.206, p< .001), concreteness (F (1, 2970)= 5.473, p= .019), and word distribution (F (1, 2552)= 52.399, p< .001). The interaction between time and word distribution was found to be marginally significant (F (2, 12760)= 2.777, p= .062), indicating that although both distributions resulted in significant gains over time, the concentration of encounters in a single episode led to slightly higher gains at posttest and delayed posttest. Yet, retention was slightly higher when the repetitions were distributed in multiple episodes (see Table 105). As concerns the significant interaction between activity type, time and word distribution (F (5, 7406)= 2.997, p= .010), the Bonferroni pairwise contrasts suggested that in the construction-focused activities, the gains in the words that were encountered in multiple episodes were lower than the case of the items that were repeated in a single episode (see Table 106). This means that testing the spaced words through construction-focused activities after the first episode where they were encountered was not conducive to higher gains. Finally, the significant interaction between time, word distribution and frequency indicated that higher repetitions resulted in higher gains when they were concentrated in a single episode; this is why, in the case of the words that were repeated a higher number of times, the distance between word-distribution categories increased at posttest and delayed posttest (see Table 107 and Figure 45).

Table 104.

The influence of word distribution on written-word form and meaning recognition

Source	F	df1	df2	Sig.
Corrected Model	33.821	19	2329	.000
Activity type	.492	1	92	.485
Time	199.790	2	6867	.000
Length	100.206	1	2444	.000
Concreteness	5.473	1	2970	.019
Frequency	2.764	1	3820	.096
Word distribution	52.399	1	2552	.000
Time * Word distribution	2.777	2	12760	.062
Activity type * Time * Word distribution	2.997	5	7406	.010
Time * Word distribution * Frequency	43.357	5	8230	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

Table 105.

Bonferroni time pairwise contrasts between word distribution categories

Distribution	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
One episode	Pretest - Posttest	-.235	.014	-16.272	2768	.000	-.270	-.201
	Pretest - Delayed	-.214	.013	-16.121	2524	.000	-.244	-.184
	Posttest - Delayed	.021	.006	3.384	12760	.001	.009	.033
Multiple episodes	Pretest - Posttest	-.216	.013	-16.419	3835	.000	-.247	-.184
	Pretest - Delayed	-.201	.012	-16.534	6948	.000	-.228	-.173
	Posttest - Delayed	.015	.008	1.971	12760	.049	.000	.031

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 106.

Bonferroni time pairwise contrasts by activity type and word distribution

Activity type	Distribution	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Meaning focused	One episode	Pretest - Posttest	-.205	.019	-11.067	6598	.000	-.249	-.161
		Pretest - Delayed	-.177	.017	-10.286	6203	.000	-.216	-.138
		Posttest - Delayed	.028	.009	3.007	12760	.003	.010	.046
	Multiple episodes	Pretest - Posttest	-.193	.015	-13.098	12760	.000	-.228	-.157
		Pretest - Delayed	-.179	.016	-11.266	12533	.000	-.215	-.143
		Posttest - Delayed	.013	.010	1.370	12760	.171	-.006	.033
Construction focused	One episode	Pretest - Posttest	-.263	.022	-11.820	1387	.000	-.316	-.209
		Pretest - Delayed	-.248	.020	-12.258	1353	.000	-.294	-.203

	Posttest - Delayed	.014	.009	1.693	12760	.090	-.002	.031
Multiple episodes	Pretest - Posttest	-.238	.022	-10.876	2562	.000	-.291	-.186
	Pretest - Delayed	-.221	.018	-12.224	4948	.000	-.262	-.181
	Posttest - Delayed	.017	.012	1.416	8060	.157	-.007	.041

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 107.

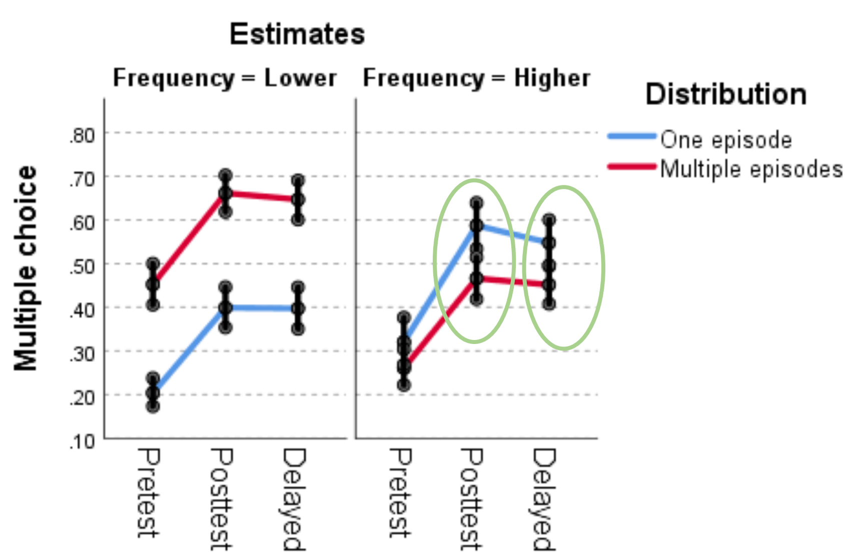
Bonferroni distribution pairwise contrasts by time and frequency

Time	Frequency	Distribution Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Pretest	Lower	One episode - Multiple episodes	-.249	.019	-12.981	2374	.000	-.286	-.211
	Higher	One episode - Multiple episodes	.059	.020	3.020	2521	.003	.021	.098
Posttest	Lower	One episode - Multiple episodes	-.263	.019	-13.873	3429	.000	-.300	-.225
	Higher	One episode - Multiple episodes	.121	.019	6.488	5238	.000	.085	.158
Delayed	Lower	One episode - Multiple episodes	-.250	.022	-11.508	3101	.000	-.292	-.207
	Higher	One episode - Multiple episodes	.096	.021	4.525	3751	.000	.054	.138

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 45.

Interaction between word distribution, time and frequency of occurrence



5.6.2 The role of frequency and word-related factors in written-word form and meaning recognition

These analyses only included learners' scores at pretest and posttest to obtain a clearer picture of the influence of frequency, regularity, concreteness and length at immediate posttest, and facilitate the comparison of the results obtained at the level of written-word form recall and written-word form and meaning recognition. To this aim, a series of repeated-measures (word and time) compound-symmetry structure GLMMs (binary logistic regression) with student identification as subjects were performed in order to study the effects of frequency, regularity, concreteness and length. To facilitate the interpretation of the outcomes and improve model fit, word characteristics were transformed to categorical variables by using the visual binning tool in SPSS (equal percentiles). FKS binary score (at item level) was set as outcome variable, while time (pretest and posttest), frequency of occurrence (3-5 vs. 6+ repetitions), concreteness (low vs. high [4.63+]), word length (shorter [<=6 letters] vs. longer words [7 letters+]), and all possible two-way interactions were entered into the model as factors. The non-significant interactions were removed one by one from the analyses until obtaining the best fitted model. When the two-way interactions suggested the presence of a three-way interaction, the model was also fitted with three-way interactions to test this assumption.

As displayed in Table 108, the results revealed significant main effects for regularity (F (1, 6470)= 110.409, p< .001), frequency (F (1, 3761)= 21161, p< .001), length (F (1, 4725)= 4.353, p= .037), and time (F (1, 3783)= 258.288, p< .001). Concerning regularity, the Bonferroni pairwise contrasts showed that the words whose orthographic patterns were more consistent with the regular patterns of L1-Spanish were easier to learn (see Table 109). The significant interaction between time and concreteness (F (1, 8557)= 12.888, p< .001) indicated that the words that were more concrete resulted in higher gains (see Table 110). The interaction between time, concreteness and length (F (3, 8428)= 6.267, p< .001) confirmed this outcome, however it added that the facilitating effect of concreteness was more prominent in the case of the words that were longer (see Table 111 and Figure 46). Finally, the significant interaction between time and frequency indicated that learners' knowledge of the less frequent words was initially higher; however, the difference between the more and less frequent words was reduced at posttest, suggesting that a higher number of repetitions triggered the learning of form-meaning links (see Table 112).

Table 108.

The influence of context and word-related factors on WWFMR

Source	F	df1	df2	Sig.
Corrected Model	45.119	10	7664	.000
Regularity	110.409	1	6470	.000
Frequency	21.161	1	3761	.000
Length	4.353	1	4725	.037
Concreteness	.057	1	7238	.810
Time	258.288	1	3783	.000
Time * Concreteness	12.888	1	8557	.000
Time * Concreteness * Length	6.267	3	8428	.000
Time * Frequency	10.231	1	8557	.001

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

Table 109.

Bonferroni pairwise contrasts between regularity groups

Regularity Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Less consistent vs. more consistent	-.124	.012	-10.470	6457	.000	-.148	-.101

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 110.

Time pairwise contrasts per concreteness categories

Concreteness	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
<= 4,62	Pretest - Posttest	-.182	.017	-10.581	5664	.000	-.216	-.148
4,63+	Pretest - Posttest	-.250	.015	-16.513	5418	.000	-.279	-.220

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 111.

Time pairwise contrasts by concreteness and length categories

Concreteness	Length	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
<= 4,62	<= 6	Pretest - Posttest	-.210	.016	-12.934	6315	.000	-.242	-.178
	7+	Pretest - Posttest	-.151	.027	-5.637	8380	.000	-.204	-.099
4,63+	<= 6	Pretest - Posttest	-.232	.016	-14.151	8557	.000	-.264	-.200
	7+	Pretest - Posttest	-.267	.023	-11.464	5763	.000	-.313	-.221

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 112.

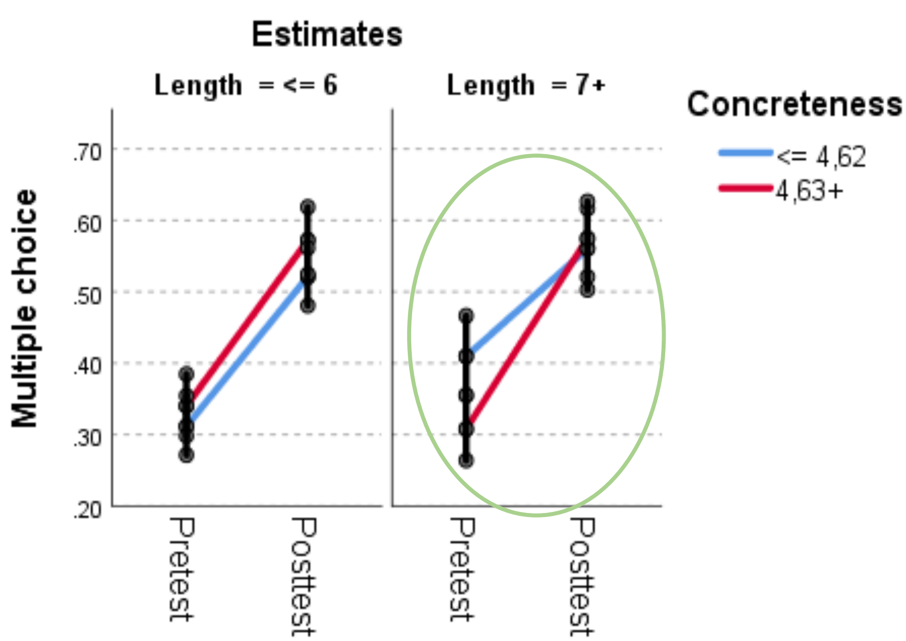
Frequency pairwise contrasts per testing time

Time	Frequency Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
Pretest	Lower - Higher	.074	.014	5.341	8557	.000	.047	.101
Posttest	Lower - Higher	.023	.013	1.773	8484	.076	-.002	.049

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 46.

Interaction between time, length and concreteness



It is important to point out that these results will be interpreted with caution since the results of the relative gains per word (Table 103) suggest that the outcomes of these analyses only reflect tendencies in the data. For instance, the word *fluffy*, which obtained the greatest relative gains (75.76%), was shorter, less concrete, less consistent with L1 patterns, and highly frequent. With this in mind, it may be assumed that word length and frequency of encounters enhanced its learnability. However, there are some factors that are not considered in the analyses. For instance, in the episode 'I will be especially, very careful', the word fluffy was used to describe a white fluffy coat, which was the main focus of the story. What this means is that its high frequency, shorter length and key role in the episode may have compensated for other factors that might have increased their learning burden, such as word regularity. Likewise, although this word was labelled as less concrete, the fluffy coat was graphically represented on screen, therefore, some factors that are beyond the scope of this investigation might well have influenced the outcomes.

5.6.3 The influence of context and word-related factors: Summary of findings

Taken together, the analyses reported in this section indicated that the following context and word-related factors led to higher gains in written-word form and meaning recognition:

- Regularity. The words that were more consistent with the regular patterns of L1 Spanish.
- Concreteness. Higher concreteness ratings. This factor was particularly relevant in the case of longer words.
- Word length. Shorter words. However, it is interesting to note that the learning of longer words was clearly boosted by higher levels of concreteness.
- Frequency. Higher number of encounters, especially when the repetitions were concentrated in a single episode.

Additionally, the analyses revealed that the concentration of encounters in a single episode led to slightly higher gains at pretest and posttest; while the distribution of encounters in multiple episodes resulted in marginally higher levels of retention from posttest to delayed posttest. As regards the potential role of construction-focused activities in the learning of the words whose occurrences were distributed in multiple episodes, the analyses indicated that being tested after the first encounter(s) did not boost learning. Likewise, the use of construction-focused activities did not enhance the role of frequency effects in the outcomes.

The results obtained in written-word form recall and written-word form and meaning recognition are summarized in Table 113.

Table 113.

Summary of the results in written-word form recall and written-word form and meaning recognition

Factors	Written-word form recall	Written-word form and meaning recognition
Viewing distribution	√√	√
Year level	√√	√√
Activity type	√√	√
Time	√√	√√
Vocabulary knowledge	√√	√√
English segmentation	√√	√√
Spanish segmentation	*	√
English reading efficacy	*	*
Spanish reading efficacy	√√	√√
Listening skills	√√	√√
PSTM	√	√
Complex working memory	*	√√
Visual processing speed	*	√√
Word distribution	*	√√
Frequency	√√	√√
Regularity	√√	√√
Length	√√	√√
Concreteness	√√	√√

√√ = significant
√ = significant but weak or unclear effects
* = non-significant

5.7 Written-word form and meaning recognition: Discussion

The main aim of this section was to determine the extent to which learners benefitted from captioned-video viewing as regards written-word form and meaning recognition. In addition, it attempted to explore the influence of treatment, learner, context and word-related factors on the outcomes. On the whole, the results indicated that both, fourth and fifth graders obtained significant gains from pretest to posttest, and from pretest to delayed posttest. Indeed, the significant effects observed for year level only reflected fifth graders' consistently higher scores over time, given that the treatment appeared to be similarly beneficial for both year levels. The results also showed that learners' gains were higher than in written-word form recall, which is expected considering the different demands at the level of recognition and recall (González-Fernández & Schmitt, 2020, Teng, 2019a). However, as it will be explained in subsequent sections, this does not mean that form-meaning mapping entails low cognitive demands (Montero Perez, 2022; Suárez & Gesa, 2019; Teng, 2019a).

As was expounded in the literature review, the simultaneous processing of audio and captions facilitate text decoding and allows viewers to devote greater attention to imagery (Pellicer-Sánchez, 2022; Tragant & Pellicer-Sánchez, 2019), which may help learners compensate for knowledge gaps (Durbahn et al., 2020, 2022; Peters & Muñoz, 2020). Thus, the positive outcomes of this study concerning written-word form and meaning recognition lend support to the beneficial effects of multimodality to enhance learners' capacity to figure out the meaning of unknown words. Likewise, one of the valuable findings of this study concerns primary school learners' capacity to make significant progress despite the absence of L1 translations (e.g. d'Ydewalle & Van de Poel, 1999; Gesa, 2019; Pujadas & Muñoz, 2019), glossaries (e.g. Fievez et al., 2021; Teng, 2022), and feedback. Yet, the great variability in relative gains among participants and words may be attributed to the influence of treatment, learner, and input-related factors (Montero Perez, 2022; Muñoz, 2022).

5.7.1 Treatment-related factors

Regarding after-viewing activity type, the analyses indicated that the use of construction-focused activities led to higher gains in receptive form-meaning mapping. However, when a series of learner-related factors were entered into the same model, the effects of activity type were no longer significant. Thus, the results imply that after-viewing activity type was a weak predictor of vocabulary learning, this is why its effects were overridden by learner-related factors. In addition, the items that occurred in multiple episodes and were tested through construction-focused activities after the first encounter(s) did not seem to receive greater attention in subsequent episodes. It may thus be hypothesized that the construction-focused activities designed for the purpose of this study were not sufficiently effective to boost learners' outcomes since the participants that completed comprehension-focused activities might have also devoted their attention to the meaning of unknown words to enhance comprehension and answer the questions. Therefore, the use of more effective intentional activities and strategies may be required to observe a clearer difference between meaning-focused and construction-focused activities (Laufer & Hulstijn, 2001; Nakata & Webb, 2016; Nation & Webb, 2011; Webb & Nation, 2017; Webb et al., 2020). Taken together, the construction-focused activities used in this study appeared to be more effective for recalling written-word forms than learning form-meaning links, which is partly consistent with Pujadas & Muñoz's (2019) findings concerning the effects of vocabulary pre-teaching.

As for the influence of viewing distribution, the results only reached significance when comparing the groups of fourth graders, suggesting that 1-fourth (ISI-7) led to slightly higher gains from pretest to delayed-posttest, and fostered greater retention from posttest to delayed posttest. This finding partially falls in line with those of Serfaty and Serrano (2022b), who found that the differential effects of ISI-7 were more evident in the case of vocabulary learning due to the lower complexity of the task in comparison with grammar learning, corroborating that longer lags may lead to better outcomes in simpler target language aspects or skills (Suzuki et al., 2019). However, it is important to acknowledge that the effects of viewing distribution in the learning of form-meaning links were not clear-cut since this effect disappeared when fitting a model with vocabulary knowledge, suggesting that 1-fourth's advantage (in gains and retention) was associated to their slightly higher vocabulary knowledge, or the weak effects of viewing distribution. In fact, learners' gains in 4-fourth were found to reach similar levels as those of 1-fourth, while 3-fourth did not perform better than 2-fourth; therefore, learners' scores did not increase nor decrease with viewing distribution in a linear order. Yet, in the first model, the results suggested that learners' retention had a negative relationship with viewing distribution. On the whole, it may be safer to conclude that viewing distribution had a weak influence on the extent to which learners benefitted from the treatment.

With respect to the significant interaction between viewing distribution, year level, and SR efficacy, the results confirmed the findings obtained in written-word form recall. Shorter lags between sessions appeared to moderate (to a certain extent) the influence of individual differences (Collins & White, 2012), namely SR efficacy; this is why students' scores in 2-fourth and 3-fourth depended more on their L1-reading skills. Hence, it might also be assumed that the advantage of 1-fourth in gains and retention (mentioned above) was counteracted by the facilitating effects of shorter lags regarding input processing (Greving & Richter, 2021). Yet, further studies should test these assumptions. In general, the results appear to support the assertion that in comparison to adults, longer lags may be less advantageous for young school learners, seeing that a greater effort to encode and retrieve knowledge might eventually result in greater retention but not necessarily in significantly greater gains (Greving & Richter, 2021; Kim & Webb, 2022b; Küpper-Tetzel et al., 2014; Serrano & Huang, 2018).

5.7.2 Cognitive and language-related factors

In relation to the cognitive factors, the analyses indicated that the three variables explored in this study significantly influenced learners' performance: complex working memory, PSTM and visual processing speed. These results corroborate the higher cognitive demands involved in form-meaning mapping since this word dimension draws on learners' capacity to integrate the meaning cues provided by each modality while viewing (Gesa, 2019; Mayer, 2014, 2022; Montero Perez et al., 2014; Peters et al., 2016; Pujadas & Muñoz, 2019; Suárez & Gesa, 2019). These findings may be explained in the framework of the Dual Coding theory (Paivio, 1986), and the Cognitive theory of Multimedia Learning (Mayer, 2014, 2022). As mentioned in the literature review, there are three cognitive processes that are essential for learning: selecting relevant verbal and pictorial information from the input, organizing the information in working memory to create coherent mental representations, and integrating these representations with previous knowledge (long-term memory) (Mayer, 2014, 2022). Therefore, learning is optimized when the learning experience considers the functioning of the human mind and is designed to reduce the cognitive load (Mayer, 2022). The studies on audiovisual input where complex WM has not emerged as a significant predictor of learners' L2 gains have attributed this outcome to learners' familiarity with viewing (Suárez et al., 2021), and the facilitating effects of onscreen text (Pattemore & Muñoz, 2020), which may prevent learners' cognitive overload (Kalyuga & Sweller, 2014). By contrast, in the studies where the learning burden has been increased through the use of non-words, the absence of captions (Montero Perez, 2020) or the implementation of intentional learning conditions (Teng & Zhang, 2021), complex working memory has been found to play a significant role in the outcomes. Therefore, despite the use of captions in the present investigation, the cognitive load might have been increased as a result of primary school learners' under-developed cognitive and literacy skills (Holmes & Myles, 2019), low proficiency level, and little familiarity with captioned videos in the L2 (Kalyuga & Sweller, 2014). Hence, consistent with the literature (Li et al., 2019), the significant effects of complex WM detected in this investigation may be associated to the heavier cognitive demands involved in the learning of form-meaning mapping from captioned-video viewing in this age group.

The significant effects of PSTM corroborates the importance of this factor in vocabulary learning (Wright, 2015) at lower proficiency levels (Montero Perez, 2020). However, as in written-word form recall, its effects were overridden when fitting a model with language-related factors. As for visual processing speed, the analyses yielded a significant interaction between visual processing speed and time, suggesting that higher processing speed fostered greater retention from posttest to delayed posttest. Thus, in light of the Dual Coding theory (Paivio, 1986), and the Cognitive theory of Multimedia Learning (Mayer, 2014, 2022), a more efficient processing of imagery may have strengthened the referential connections between verbal and non-verbal information, facilitating their further recall (Clark & Paivio, 1991).

In regard to the L2-related factors, the results showed significant effects for vocabulary knowledge, listening skills, and English text segmentation. This may be attributed to the fact that reading comprehension, a key component of the viewing process, is mainly explained by L2-related factors, namely decoding and linguistic comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990; Sparks, 2021; Tunmer & Chapman, 2012). As mentioned in the literature review, when the input matches learners' L2 proficiency, the processing of input is less effortful and learners may devote greater attention to unknown words to foster learning (Kormos, 2017; Lin & Siyanova-Chanturia, 2015). Based on the Dual-theoretical model of reading (Sadoski & Paivio, 2013), learners' decoding skills may be particularly relevant in written-word form and meaning recognition since their greater effort and attention to lower-level linguistic processes may hinder the associational and referential processing between the verbal and non-verbal codes (Sadoski et al., 2004; Sadoski & Paivio, 2013). Thus, readers' poor decoding skills may affect their capacity to link verbal and non-verbal input to fill knowledge gaps and enhance comprehension. This may explain why English text segmentation and not ER efficacy emerged as a significant predictor of vocabulary learning.

As for the contribution of vocabulary knowledge, it is important to note that the results confirmed the 'the rich get richer principle', which indicates that the greater vocabulary knowledge, the greater vocabulary gains (Montero Perez, 2022; Montero Perez et al., 2013; Stanovich, 1986). In most studies on audiovisual input, vocabulary knowledge has emerged as one of the strongest predictors of L2 learning (e.g. Alexiou, 2015; Montero Perez et al. 2013, 2018; Peters et al., 2016; Peters & Webb, 2018; Pujadas & Muñoz, 2019). Concerning listening skills, it is worth mentioning that while this factor was shown to predict learners' performance in receptive form-meaning mapping, the exponential coefficient suggested that its influence was higher in written word form recall. This finding is not surprising considering that in the dictation task, the aural word-form representations were used as prompts. Thus, learners' capacity to decode and comprehend the stream of speech (with the support of captions) appeared to be even more relevant at the level of written-word form recall.

In regard to the L1-related factors, Spanish text segmentation was not shown to be a strong predictor of written-word form and meaning recognition. Its weak contribution may be associated to the fact that this instrument is considered to be a measure of lower-level reading skills (Torres-Díaz et al., 2020), and the role of vocabulary and grammatical knowledge in reading (Alderson et al., 2015). As the literature suggests, L2-related factors may play a more significant role in L2 reading comprehension than L1-related factors (Alderson et al., 2016; Sparks, 2021). In addition, previous research has suggested that the simultaneous processing of L2 bimodal verbal input facilitates text decoding; therefore, learners' performance in reading-while-listening does not seem to be explained by L1-lower level reading skills, as in the case of the reading-only condition (Kormos et al., 2019). Thus, the stronger relationship between SR efficacy and learners' scores in written-word form and meaning recognition in 2-fourth and 3-fourth may be accounted by the wider scope of the SR efficacy test, which integrates lower-level and higher-level reading skills. Hence, the results suggest that the fourth graders, who were less proficient and were still developing their L1 reading skills, showed greater reliance on L1 reading skills to compensate, to a certain extent, for their knowledge gaps (Yamashita, 2002). In addition, this finding seems to further support the idea that at early L2 learning stages, learners may assimilate and accommodate their linguistic infrastructure to the characteristics of L2 (Birch & Fulop, 2021; Jiang et al., 2019; Perfetti et al., 2007), which is a process that evolves as a function of L2 proficiency and familiarity with the characteristics of the target language (Jiang et al., 2019).

5.7.3 Context and word-related factors

On the whole, the results revealed that all the context and word-related factors assessed in this study affected word learnability at the level of receptive form-meaning mapping. However, considering the characteristics of the words that obtained the highest and the lowest relative gains, these results should be interpreted with caution since they only report tendencies in the data. Concerning frequency effects, the results indicated that higher repetitions enhanced vocabulary learning. However, this factor seemed to be conducive to higher gains when the repetitions were concentrated in a single episode. This is congruent with the results obtained in written-word form recall and in earlier findings where frequency of occurrence has been shown to be moderated by input spacing (Uchihara et al., 2019). Nonetheless, this was not the case of activity type, given that in contrast to earlier findings, no evidence was detected on the potential relationship between the use of construction-focused activities (which worked as a sort of test announcement) and learners' higher sensitivity to frequency effects (Uchihara et al., 2019).

Although word distribution was only examined to assess its interaction with activity type and word frequency, it is worth noting that the results also indicated that the concentration of encounters in a single episode led to slightly higher gains, whereas the distribution of encounters in multiple episodes triggered marginally greater retention. This is partially congruent with the literature on distributed practice effects, which suggests that input spacing leads to the creation of stronger memory traces that prevent quick knowledge decay (Rogers, 2021). Yet, this finding may be somewhat limited by the intrinsic difficulty of the items in each group (massed and spaced).

Concerning word regularity, the results corroborated that this factor did not only aid the recall of written word forms but also the learning of form-meaning links. First of all, this finding might be accounted by the fact that the words that have more transparent orthographic patterns are easier to decode (Hamada & Koda, 2008); therefore, the participants might have had more attentional resources available to extract their meaning from the verbal and non-verbal input. In addition, the advantage of word regularity might be explained by the association between bimodal verbal input and the creation of stronger memory representations in the words that have more consistent orthographic patterns (Krepel et al., 2020).

As for word concreteness, the results reflect those of previous studies which showed that higher concreteness ratings facilitate vocabulary learning (De Groot & Keijzer, 2000; Ellis & Beaton, 1993b; Puimège & Peters, 2019b). The beneficial effects of concreteness may be attributed to word saliency (Crossley et al., 2016), imageability, and the strengthening of the referential connections between verbal and non-verbal representations (Clark & Paivio, 1991), which enhance learning (Mayer, 2022). In the case of audiovisual input, higher concreteness may also be translated to the presence of graphic representations on screen (Peters, 2020), increasing the odds of word learning (Rodgers, 2020). Thus, considering that lower proficiency learners may rely on imagery to achieve comprehension (Muñoz, 2022), it should not be surprising that concreteness emerged as a strong predictor of written-word form and meaning recognition. Yet, the actual effects of imagery are beyond the scope of this investigation.

As regards word length, an interesting picture emerged. To start with, the results are in accord with previous studies indicating that shorter words are easier to learn (Ellis & Beaton, 1993a; Barclay & Pellicer-Sánchez, 2022). As the literature suggests, longer words need additional time to be processed and recognized (Grabe, 2009), and are harder to store in the PSTM (Birch, 2015). Nonetheless, the significant interaction between word length, concreteness and time indicated that the words that were longer and more concrete obtained the highest gains, which was not the case of the longer words that had lower concreteness ratings. This tendency may be partially congruent with that of Puimège & Peters (2019b) who found that the longer words forms encountered in a non-captioned video were easier to recall. In their study, the word length advantage was associated to saliency in the aural input (Puimège & Peters, 2019b). While in our investigation the significant relationship between word length and concreteness may also be accounted by their greater saliency, the results also suggest that the intrinsic difficulty of longer words (Barclay, 2021) may have been counteracted by the effects of word concreteness, and the simultaneous encoding of information through the verbal and non-verbal channels (Clark & Paivio, 1991). In light of

Paivio's Dual Coding Theory (1986), their higher imageability (or even graphical representations in scene) may have strengthened the associational and referential connections between verbal and non-verbal representations, enhancing learning and recall (Clark & Paivio, 1991).

VI. L2 listening skills

This section focuses on the development of listening skills from captioned-video viewing in five groups of primary school learners. It is important to note that it does not intend to explore learners' comprehension of each video but rather the development of listening skills (i.e. generalization of learning) as a result of learners' extensive exposure to captioned videos. Specifically, the analyses reported in this section respond to the following research questions:

1) To what extent does viewing distribution (i.e. shorter vs. longer lags) influence young L2 learners' gains from captioned video viewing?

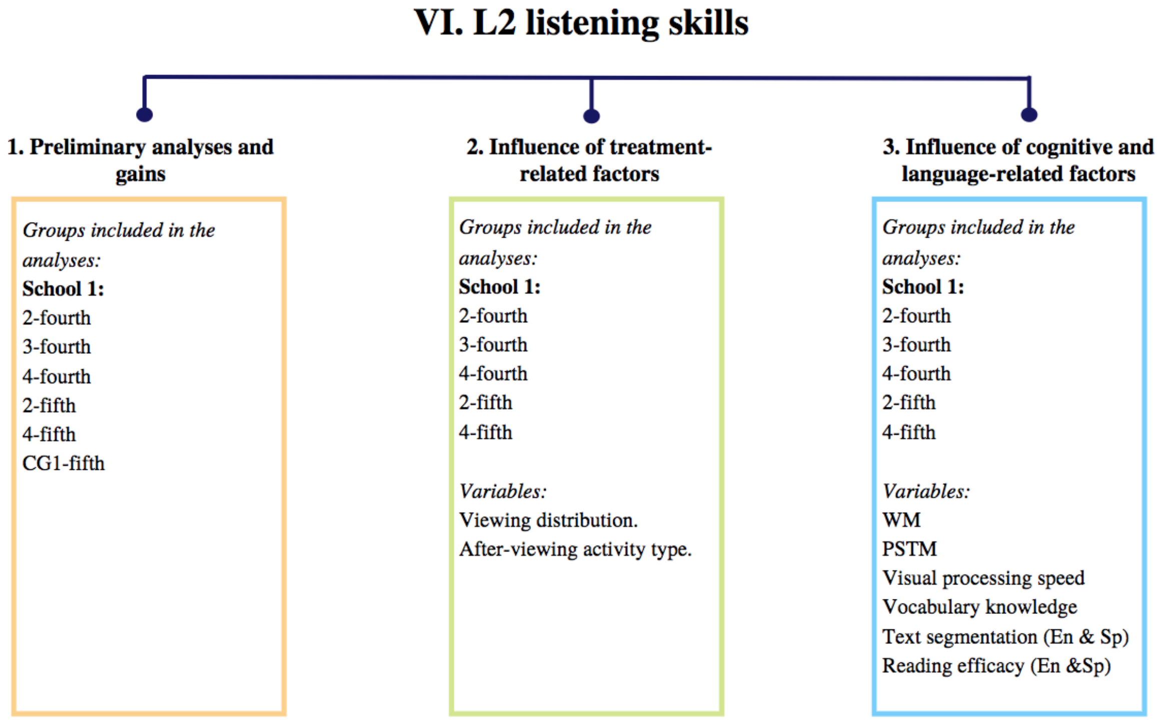
2) In comparison with meaning-focused activities, what are the effects of construction-focused after-viewing activities on L2 learning through captioned-video viewing?

3) To what extent do learner characteristics influence young L2 learners' gains from captioned-video viewing? (i.e. age, vocabulary knowledge, cognitive abilities [phonological short-term memory, complex working memory, and visual processing speed], and L1 and L2 reading skills [reading efficacy and text segmentation]).

The overview of this section is displayed in Figure 47. As explained in the methodology section, listening skills, measured by means of two Movers sample tests (Cambridge Assessment, 2018), was tested at pretest, posttest and delayed posttest (fourteen days after the administration of the posttest.

Figure 47.

Section 6 overview



6.1 Listening skills: Preliminary analyses

A set of between-groups comparisons were run with the aim of establishing whether the groups were comparable as regards L2 listening skills at pretest. To start with, the Independent-samples T-test performed to compare the two year levels indicated that fifth graders outperformed fourth graders at pretest (t (94)= 17.921, p<.001, r=.87) (see descriptive statistics in Table 114). Then, a One-Way ANOVA was run in order to determine whether the classes from each year level were comparable at pretest. The results revealed that the overall difference between groups (fourth and fifth graders) was statistically significant (F (5)= 3.361, p=.008, η^2 = .143). Specifically, the Tukey pairwise contrasts indicated that, although fifth graders were found to score higher, the only comparison that reached significance was between 2-fifth and 3-fourth (p=.017); whereas the difference between CG1-fifth and 3-fourth was only shown to be marginally significant (p=.061). In summary, these results indicate that the groups in each year level were comparable at pretest (see Table 115 and Figure 48).

Table 114.

Listening skills: Descriptive statistics

		Listening pretest		Listening posttest		Listening delayed	
		Mean	(SD)	Mean	(SD)	Mean	(SD)
Class	3-fourth	7.69	(3.63)	10.06	(3.55)	11.37	(4.08)
	4-fourth	8.64	(3.05)	11.33	(3.37)	10.33	(4.15)
	2-fourth	9.06	(3.38)	12.19	(3.31)	11.69	(3.18)
	2-fifth	11.83	(4.96)	13.75	(4.45)	14.54	(4.09)
	4-fifth	10.77	(4.03)	13.52	(3.16)	13.80	(3.66)
	CG1-fifth	11.56	(3.12)	10.06	(3.09)	.	.
Year level	Year 4	8.46	(3.35)	11.19	(3.46)	11.15	(3.78)
	Year 5[a]	11.31	(4.51)	13.63	(3.81)	14.16	(3.85)

a Without the control group

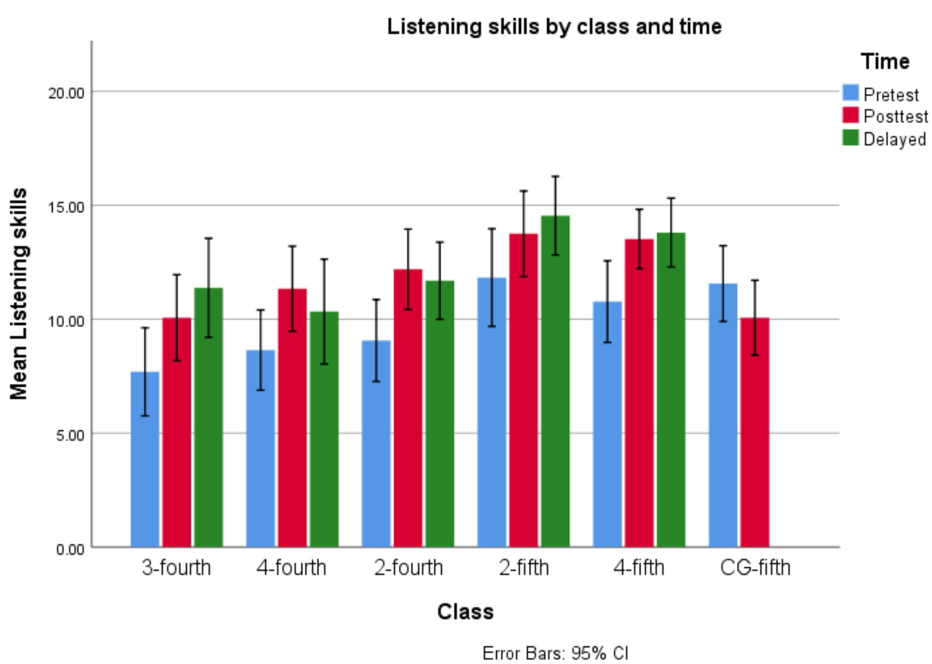
Table 115.

Summary: Between-groups comparisons in terms of listening pretest scores

Factor	Statistically sig. differences between year levels	Statistically sig. differences between classes
Listening pretest scores	Year 5 > Year 4	3-fourth < 2-fifth 3-fourth < CG1-fifth (marginally significant)

Figure 48.

Listening skills: Groups' performance over time



Additionally, Pearson correlations were run in order to explore the relationships between learners' outcomes at the three testing times and the continuous variables assessed in this study (cognitive and language-related factors) (see Table 116). The results revealed stronger relationships between the listening scores and L2-related factors, namely vocabulary knowledge and English text segmentation with a large effect size (R^2 > .25) (Larson-Hall, 2010). Although their relationship with L1-related factors reached statistical significance, their shared variance accounted for a medium (or medium-large) effect size (Larson-Hall, 2010). By the same token, the significant correlations between the listening scores and the cognitive factors indicated that the strength of their relationship was either weak or moderate. Among the three cognitive factors, the strongest correlations were found between PSTM and learners' listening scores with a medium effect size (R^2 >= .09) at pretest and posttest (Larson-Hall, 2010). Thus, the results obtained from the correlations suggest that learners' listening scores over time were mainly explained by vocabulary knowledge and English text segmentation.

Table 116.

Correlations between listening scores and learner-related factors

		Listening pretest	Listening posttest	Listening delayed
Listening pretest	Pearson Correlation	1	**.780** (R^2=60)**	**.774**(R^2=59)**
	Sig. (2-tailed)		.000	.000
	N	91	91	91
Listening posttest	Pearson Correlation	**.780** (R^2=60)**	1	**.723** (R^2=52)**
	Sig. (2-tailed)	.000		.000
	N	91	96	96
Listening delayed	Pearson Correlation	**.774** (R^2=59)**	**.723** (R^2=52)**	1
	Sig. (2-tailed)	.000	.000	
	N	91	96	96
PSTM	Pearson Correlation	.308** (R^2=09)	.383**(R^2=14)	.299** (R^2=08)
	Sig. (2-tailed)	.003	.000	.004
	N	90	93	93
Complex WM	Pearson Correlation	.248* (R^2=06)	.294**(R^2=08)	.196 (R^2=03)
	Sig. (2-tailed)	.018	.004	.059
	N	90	93	93
	Pearson Correlation	.290** (R^2=08)	.207* (R^2=04)	.208* (R^2=04)
	Sig. (2-tailed)	.006	.046	.046

Visual Processing speed	N	90	93	93
EFL PVT	Pearson Correlation	**.673**** **(R^2=45)**	**.679**** **(R^2=46)**	**.683******(R^2=46)**
	Sig. (2-tailed)	.000	.000	.000
	N	87	88	88
English segmentation	Pearson Correlation	**.683**** **(R^2=46)**	**.632******(R^2=39)**	**.641******(R^2=41)**
	Sig. (2-tailed)	.000	.000	.000
	N	86	89	89
Spanish segmentation	Pearson Correlation	.451** (R^2=20)	.435** (R^2=18)	.405** (R^2=16)
	Sig. (2-tailed)	.000	.000	.000
	N	87	90	90
SR efficacy	Pearson Correlation	**.504**** **(R^2=25)**	.470** (R^2=22)	.432** (R^2=18)
	Sig. (2-tailed)	.000	.000	.000
	N	90	92	92
ER efficacy	Pearson Correlation	.481** (R^2=23)	**.502**** **(R^2=25)**	.474** (R^2=22)
	Sig. (2-tailed)	.000	.000	.000
	N	90	92	92

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

6.2 Listening skills: Progress over time

In order to compare the trajectories of both year levels over time, we ran a compound symmetry structure GLMM (binary logistic regression) with student identification as subjects, and time as repeated measures. The model was built with learners' scores at the three testing times by setting 20 (maximum score) as denominator. The fixed effects included in the analyses were time, year level and their interaction. In this model, the scores obtained by the control group were not included. The results yielded significant main effects for time (F (2,190)= 63.966, p< .001), and year level (F (1,89)= 15.844, p< .001), while the interaction between these two factors did not reach significance (see Table 117). However, this interaction was kept in the model to further explore the trajectory of each year level (see Figure 49). As shown in Table 118, the experimental groups showed significant improvement from pretest to posttest, and from pretest to delayed posttest, regardless of their year level. In addition, the results indicated that learners' scores did not decrease significantly from posttest to delayed posttest. The significant effects of year level only confirmed that the higher performance of fifth graders was kept over time. On the whole, the results indicate that the treatment was similarly beneficial for both year levels.

Table 117.

Learners' development of L2 listening skills over time by year level

Source	F	df1	df2	Sig.
Corrected Model	29.964	5	203	.000
Time	63.966	2	190	.000
Level	15.844	1	89	.000
Level * Time	.560	2	190	.572

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20

Table 118.

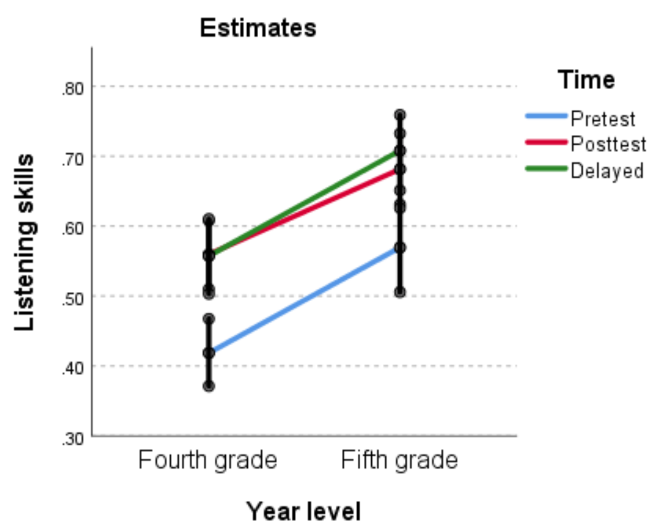
Listening skills: Time pairwise contrasts

Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Pretest - Posttest	-.129	.014	-9.224	241	.000	-.162	-.095
Pretest - Delayed	-.142	.015	-9.785	197	.000	-.175	-.109
Posttest - Delayed	-.014	.015	-.894	148	.373	-.044	.016

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 49.

Listening skills: The trajectory of each year level over time



6.2.1 Comparisons between control and experimental groups

To assess the performance of the control and the experimental groups from pretest to posttest, we ran a compound symmetry structure GLMM (binary logistic regression) with student identification as subjects, and time as repeated measures. The analysis was calculated with learners' listening scores as the target variable by setting 20 (maximum score) as denominator. The fixed effects included in the analysis were class, time, and their interaction. As shown in Table 119, the results revealed significant effects for class (F (5,116)= 3.141, p= .011), time (F (1,103)= 53.426, p< .001) and their interaction (F (5,107)= 6.246, p< .001). Specifically, the Bonferroni adjusted results revealed that all the groups improved significantly from pretest to posttest (p< .001), except for the control group, who were even found to score lower at posttest (see Table 120).

Table 119.

Learners' development of L2 listening skills over time (pretest and posttest) by class

Source	F	df1	df2	Sig.
Corrected Model	10.548	11	179	.000
Class	3.141	5	116	.011
Time	53.456	1	103	.000
Class * Time	6.246	5	107	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20

Table 120.

Listening skills: Time pairwise contrasts by class (pretest and posttest)

Class	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-fourth	Pretest - Posttest	-.119	.033	-3.627	120	.000	-.184	-.054
4-fourth	Pretest - Posttest	-.148	.036	-4.139	110	.000	-.220	-.077
2-fourth	Pretest - Posttest	-.156	.037	-4.200	72	.000	-.230	-.082
2-fifth	Pretest - Posttest	-.092	.026	-3.504	130	.001	-.145	-.040
4-fifth	Pretest - Posttest	-.138	.025	-5.446	180	.000	-.188	-.088
CG1-fifth	Pretest - Posttest	.075	.036	2.105	88	.038	.004	.146

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

6.2.2 Comparisons between experimental groups

A new model was fitted in order to examine the trajectory of the experimental groups from pretest to delayed posttest. To this aim, we ran a compound symmetry structure GLMM (binary logistic regression) with student identification as subjects, and time as repeated measures. The analysis was calculated with learners' listening scores as the target variable by setting 20 (maximum score) as denominator. The fixed effects included in the analysis were class, time, and their interaction. This time, the analyses only yielded significant effects for class (F (4,100)= 4.235, p= .003) and time (F (2,172)= 61.605, p< .001) (see Table 121). The non-significant interaction between class and time was kept in the model to further explore each class' trajectory. As shown in Table 122, the significant effects of class reflected the significantly higher scores obtained by fifth graders over time. Yet, this outcome does not imply that the treatment was more beneficial for this year level. On the whole, the Bonferroni pairwise contrasts in Tables 122 and 123 indicate that all the groups improved significantly from pretest to posttest, and from pretest to delayed posttest. As shown in Table 123, similar patterns were found in all the experimental groups. Moreover, the possible differences that may be observed in Table 123 do not seem to follow a particular pattern to be attributed to lag effects. As for the apparent higher benefits obtained by 3-fourth, they might be associated to their initial lower scores and greater room for learning. Taken together, the results suggest that the treatment was similarly beneficial for all the experimental groups (see Table 124).

Table 121.

Learners' development of L2 listening skills over time by class

Source	F	df1	df2	Sig.
Corrected Model	12.103	14	230	.000
Class	4.235	4	100	.003
Time	61.605	2	172	.000
Class * Time	1.026	8	183	.418

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20

Table 122.

Listening skills: Class and time pairwise contrasts

Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-fourth - 4-fourth	-.016	.058	-.275	146	1.000	-.136	.104
3-fourth - 2-fourth	-.065	.055	-1.187	163	.949	-.203	.073
3-fourth - 2-fifth	-.187	.060	-3.113	63	.028	-.362	-.012
3-fourth - 4-fifth	-.154	.053	-2.920	129	.037	-.303	-.005
4-fourth - 2-fourth	-.049	.053	-.928	207	1.000	-.176	.078
4-fourth - 2-fifth	-.171	.058	-2.932	68	.037	-.337	-.006
4-fourth - 4-fifth	-.138	.051	-2.720	158	.051	-.277	.000
2-fourth - 2-fifth	-.122	.055	-2.212	68	.182	-.273	.028
2-fourth - 4-fifth	-.089	.047	-1.890	188	.301	-.212	.034
2-fifth - 4-fifth	.033	.054	.620	56	1.000	-.089	.155
Pretest - Posttest	-.132	.014	-9.181	202	.000	-.167	-.097
Pretest - Delayed	-.142	.015	-9.773	187	.000	-.175	-.109
Posttest - Delayed	-.010	.015	-.672	137	.503	-.040	.020

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 123.

Listening skills: Time pairwise contrasts by class

Class	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-fourth	Pretest - Posttest	-.119	.033	-3.627	234	.001	-.193	-.045
	Pretest - Delayed	-.184	.037	-4.996	154	.000	-.274	-.095
	Posttest - Delayed	-.066	.032	-2.026	254	.044	-.129	-.002
4-fourth	Pretest - Posttest	-.148	.035	-4.172	189	.000	-.233	-.062
	Pretest - Delayed	-.098	.035	-2.783	201	.012	-.177	-.018
	Posttest - Delayed	.050	.043	1.150	95	.253	-.036	.136
2-fourth	Pretest - Posttest	-.156	.037	-4.200	149	.000	-.246	-.066
	Pretest - Delayed	-.131	.034	-3.904	227	.000	-.207	-.055
	Posttest - Delayed	.025	.035	.713	179	.477	-.044	.094
2-fifth	Pretest - Posttest	-.093	.026	-3.529	257	.001	-.152	-.033
	Pretest - Delayed	-.132	.027	-4.911	240	.000	-.197	-.067
	Posttest - Delayed	-.040	.022	-1.796	268	.074	-.083	.004
4-fifth	Pretest - Posttest	-.132	.025	-5.219	268	.000	-.193	-.071
	Pretest - Delayed	-.146	.029	-5.094	152	.000	-.211	-.081
	Posttest - Delayed	-.014	.032	-.437	85	.663	-.078	.050

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 124.

Summary of findings: Listening skills scores over time

Analysis	Outcome
Significant improvement from pretest to posttest (p<.05).	Both year levels. All the groups, except for the control group. 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth.
Significant improvement from pretest to delayed posttest (p<.05).	Both year levels. All the groups, except for the control group. 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth.

6.3 Listening skills: The influence of treatment-related factors

6.3.1 After-viewing activity type

In order to measure the influence of after-viewing activity type (see descriptive statistics in Table 125) on learners' scores over time, we performed a compound symmetry structure GLMM (binary logistic regression) with student identification as subjects, and time as repeated measures. The analyses were calculated with learners' listening scores as the target variable by setting 20 (maximum score) as denominator. The fixed factors included in the analysis were year level, time, activity type, and their interaction. The interaction between activity type and time was not found to be significant, therefore it was eliminated from the model. As shown in Table 126, the results did not yield significant effects for activity type (F (1,90)= .683, p= .411), suggesting that learners benefitted from the treatment regardless of the type of activity they had to complete after watching each episode. Then, a new model was fitted by adding vocabulary knowledge as covariate to determine whether the non-significant differences between activity-type groups could be attributed to differences in proficiency. Given that activity type remained as a non-significant factor, it may thus be concluded that learners' outcomes did not differ as a function of activity type.

Table 125.

Listening skills: Descriptive statistics per activity type

	Activity type			
	Meaning focused		Construction focused	
	Mean	(SD)	Mean	(SD)
Listening pretest	10.53	(4.25)	9.22	(4.09)
Listening posttest	12.60	(4.13)	12.29	(3.54)
Listening delayed	12.66	(4.62)	12.71	(3.56)

Table 126.

Listening skills: The influence of activity type on learners' outcomes

Source	F	df1	df2	Sig.
Corrected Model	37.712	4	124	.000
Activity type	.683	1	90	.411
Year level	16.070	1	89	.000
Time	63.093	2	186	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20

6.3.2 Viewing distribution

To study the influence of viewing distribution on learners' outcomes, we ran a series of compound-symmetry structure GLMMs (Binary logistic regression) with student identification as subjects and time as repeated measures. Learners' listening scores over time were set as outcome variable with 20 (maximum test score) as denominator. In these analyses, only the scores of the participants that watched either two or four episodes a week were included since 3-fourth did not have a counterpart in year 5. The fixed factors entered into the model were as follows: viewing distribution, year level, time and all possible two-way and three-way interactions. Then, by following a step back procedure, the non-significant interactions were removed one by one until the best fitted model was obtained. As shown in Table 127, viewing distribution did not predict learners' listening scores (F (1,70)= 1.094, p= .299). Thus, a new model was built by entering vocabulary knowledge as covariate to rule out the possibility that viewing distribution did not emerge as significant predictor due to differences in proficiency. The analyses corroborated that viewing distribution did not influence the extent to which learners benefitted from the treatment as concerns L2 listening skills.

Table 127.

Listening skills: The influence of viewing distribution on learners' outcomes

Source	F	df1	df2	Sig.
Corrected Model	29.973	4	106	.000
Viewing distribution	1.094	1	70	.299
Year level	11.618	1	78	.001
Time	52.391	2	155	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20

6.3.3 The influence of treatment-related factors: Summary of findings

This section explored the influence of treatment-related factors (i.e. after-viewing activity type and viewing distribution) on the development of listening skills from captioned-video viewing. The main findings are enlisted as follows:

- The results did not yield significant effect for activity type nor viewing distribution.
- Thus, learners benefitted from the treatment regardless of the number of episodes they had to watch a week and the type of activity they were asked to complete.

6.4 Listening skills: The influence of cognitive and language-related factors

This section focuses on the analyses that examined the influence of cognitive and language-related factors on the development of listening skills. To this aim, we ran a series of compound-symmetry structure GLMMs (Binary logistic regression) with student identification as subjects and time as repeated measures. Learners' listening scores over time were set as outcome variable with 20 (maximum test score) as denominator. As regards the continuous variables that were entered into the models, it is important to mention that collinearity tests were performed before running the analyses to ensure that all the independent variables could be entered (Pallant, 2016).

6.4.1 Cognitive factors

In order to assess the influence of the cognitive factors on the development of listening skills, the following factors were entered into the model: class, time, PSTM, complex working memory, visual processing speed (high vs. low), and all possible two-way interactions. A backward elimination procedure was used to determine the best fitted model. Thus, the non-significant interactions and main factors were removed from the model one by one. In this case, the analyses indicated that visual processing speed did not contribute to the learning process significantly ($p > .05$), therefore, this factor was removed from the best fitted model (see Table 128). As summarized in Table 129, the analyses yielded significant effects for class (F (4,82)= 3.129, p= .019), PSTM (F (1,69)= 10.920, p= .002) and time (F (2,178)= 63.737, p< .001), while complex working memory was only found to approach significance (F (1,67)= 3.395, p= .070). The exponential coefficient in Table 128 indicates that when PSTM scores increased by one, the odds of a correct response increased by 17%. As expected, the contribution of complex working memory was much lower (8%).

Table 128.

Listening skills: Best fitted model obtained to assess the influence of cognitive factors on learners' scores.

Model Term	Coef	Std. Error	t	Sig.	95% CI Lower	95% CI Upper	Exp (Coef)	95% CI for Exp(Coef) Lower	95% CI for Exp(Coef) Upper
Intercept	-1.390	.4839	-2.871	.005	-2.349	-.430	.249	.095	.651
3-fourth	-.475	.2011	-2.362	.020	-.874	-.076	.622	.417	.926
4-fourth	-.280	.1817	-1.543	.124	-.638	.078	.756	.528	1.081
2-fourth	-.176	.1793	-.982	.327	-.530	.178	.839	.589	1.194
2-fifth	.324	.2206	1.469	.149	-.120	.768	1.383	.887	2.156
4-fifth	0^b	.	.	.	.	.	.	.	.
PSTM	.156	.0473	3.305	.002	.062	.251	1.169	1.064	1.285
Complex WM	.076	.0414	1.843	.070	-.006	.159	1.079	.994	1.172
Pretest	-.593	.0623	-9.512	.000	-.716	-.470	.553	.489	.625
Posttest	-.041	.0682	-.597	.552	-.176	.094	.960	.839	1.099
Delayed posttest	0^b	.	.	.	.	.	.	.	.

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20 b. This coefficient is set to zero because it is redundant.

Table 129.

Listening skills: The influence of cognitive factors on learners' scores

Source	F	df1	df2	Sig.
Corrected Model	21.852	8	107	.000
Class	3.129	4	82	.019
PSTM	10.920	1	69	.002
Complex working memory	3.395	1	67	.070
Time	63.737	2	178	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20

6.4.2 Language-related factors

In order to assess the influence of language-related factors on the development of listening skills, a series of Generalized Linear Mixed Model (binary logistic regression) with repeated measures (time) compound-symmetry structure were calculated. The first model was built with L2-related factors and all possible two-way interactions. Specifically, the following variables were entered into the model: class, time, ER efficacy, vocabulary knowledge and English text segmentation. The best fitted model was determined by a backward elimination procedure (see Table 130). As shown in Table 131, the results revealed significant main effects for vocabulary knowledge (F (1,114)= 34.953, p< .001), English text segmentation (F (1,60)= 21.028, p= .002) and time (F (2,171)= 54.912, p< .001). The exponential coefficients in Table 130 indicate that when learners' scores in the EFL picture vocabulary test increased by one, the odds of a correct response in the listening test increased by 48%. Likewise, the odds of an accurate response in the listening test increased by 22% per each additional word identified in the English segmentation test.

Table 130.

Listening skills: Best fitted model obtained to assess the influence of L2-related factors on learners' scores.

Model Term	Coef	Std. Error	t	Sig.	95% CI Lower	95% CI Upper	Exp (Coef)	95% CI for Exp(Coef) Lower	95% CI for Exp(Coef) Upper
Intercept	-2.070	.2803	-7.385	.000	-2.633	-1.507	.126	.072	.222
Vocabulary knowledge	.395	.0668	5.912	.000	.263	.528	1.485	1.301	1.695
English segmentation	.202	.0440	4.586	.000	.114	.290	1.223	1.120	1.336
Pretest	-.591	.0682	-8.676	.000	-.726	-.457	.553	.484	.633
Posttest	-.014	.0751	-.182	.856	-.162	.135	.986	.850	1.144
Delayed posttest	0[b]	.	.	.	.	.	.	.	.

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20
b. This coefficient is set to zero because it is redundant.

Table 131.

Listening skills: The influence of L2-related factors on learners' scores

Source	F	df1	df2	Sig.
Corrected Model	53.218	4	97	.000
Vocabulary knowledge	34.953	1	114	.000
English text segmentation	21.028	1	60	.000
Time	54.912	2	171	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20

Then, a new model was built to explore the influence of L1-related factors on learners' listening scores. To this aim, class, time, Spanish text segmentation, SR efficacy, and all possible two-way interactions were entered as fixed factors. The non-significant interactions and main effects were removed from the model one by one until the best fitted model was obtained (see Table 132). The results showed significant main effects for time (F (2,171)= 57.914, p< .001), Spanish text segmentation (F (1,166)= 10.099, p= .002) and SR reading efficacy (F (1,257)= 22.088, p< .001) (see Table 133). In addition, the exponential coefficients in Table 132 indicate that the odds of obtaining a correct response in the listening test increased by 19% when segmenting a word correctly, and by 13% per each additional point in SR efficacy. However, when fitting a model with the significant L1 and L2 related factors, Spanish text segmentation and SR efficacy were no longer found to be significant. Their influence must have been overshadowed by the strong significant effects of vocabulary knowledge and English text segmentation. Thus, the same model in Table 132 was obtained. By the same token, PSTM was not shown to predict learners' performance at the listening test when building a model with vocabulary knowledge and English text segmentation. Consequently, the model that includes time, vocabulary knowledge and English text segmentation is the one that best explains the extent to which learner-related factors influenced learners' scores in listening over time.

Table 132.

Listening skills: Best fitted model obtained to assess the influence of L1-related factors on learners' scores.

Model Term	Coef	Std. Error	t	Sig.	95% CI Lower	95% CI Upper	Exp (Coef)	95% CI for Exp(Coef) Lower	95% CI for Exp(Coef) Upper
Intercept	-1.508	.2860	-5.272	.000	-2.072	-.944	.221	.126	.389
Pretest	-.573	.0635	-9.020	.000	-.698	-.447	.564	.498	.639
Posttest	-.025	.0695	-.361	.719	-.163	.112	.975	.850	1.119
Delayed	0[b]	.	.	.	.	.	.	.	.
Spanish text segmentation	.172	.0541	3.178	.002	.065	.279	1.187	1.067	1.321
SR efficacy	.124	.0264	4.700	.000	.072	.176	1.132	1.075	1.193

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20
b. This coefficient is set to zero because it is redundant.

Table 133.

Listening skills: The influence of L1-related factors on learners' scores

Source	F	df1	df2	Sig.
Corrected Model	44.740	4	188	.000
Time	57.914	2	171	.000
Spanish segmentation	10.099	1	166	.002
SR efficacy	22.088	1	257	.000

Probability distribution: Binomial
Link function: Logit
a. Target: Listening skills/20

6.4.3 The influence of cognitive and language-related factors: Summary of findings

Overall, the results reported in this section indicate that:

- Among the cognitive factors, PSTM was the only variable that played a more significant role in the development of L2 listening skills. However, it may be considered to be a weak predictor when compared to L2-related factors.
- Within the group of language-related factors, the analyses indicated that vocabulary knowledge and English text segmentation were the strongest predictors of learners' progress over time.
- Although SR efficacy and Spanish text segmentation reached significance levels in the model that only included L1-related factors, these variables seemed to have a weak influence on learners' listening scores since their significant effects disappeared when compared to L2-related factors.

The results concerning the influence of treatment and learner-related factors are summarized in Table 134.

Table 134.

Summary: Predictors of listening skills

Factors	Outcomes
Viewing distribution	***Non-significant.***
Year level	***Significant***. Both year levels improved significantly from pretest to posttest, and from pretest to delayed posttest. However, fifth graders scored significantly higher at the three testing times.
After-viewing activity type	***Non-significant.***
Time	***Significant.***
Vocabulary knowledge	***Significant.*** One of the strongest predictors of learners' listening scores.
English segmentation	***Significant.*** One of the strongest predictors of learners' listening scores.
Spanish segmentation	***Statistically significant*** but a weaker predictor of learners' listening scores. When fitting a model with L2-related factors, Spanish text segmentation was no longer significant.
English reading efficacy	***Non-significant.***

Spanish reading efficacy	***Statistically significant*** but a weaker predictor of learners' listening scores. When fitting a model with L2-related factors, SR efficacy was no longer significant.
Phonological short-term memory	***Statistically significant*** but a weaker predictor of learners' listening scores. Its effects were overridden by L2-related factors.
Working memory	***Non-significant.***
Visual processing speed	***Non-significant.***

6.5 Listening skills: Discussion

The analyses in this section aimed to determine the extent to which captioned-video viewing enhanced the development of L2 listening skills. In addition, it assessed the influence of treatment and learner-related factors on learners' performance over time. To start with, the results revealed that all the groups showed significant improvement from pretest to posttest, and from pretest to delayed posttest, except for the control group. The significant effects of year level indicated that fifth graders consistently outperformed fourth graders, keeping a similar distance between the two groups over time, which may be attributed to their higher L2 proficiency level. It is important to note that the results did not demonstrate that the treatment was particularly more beneficial for one of the year levels; rather, learners' sustained exposure to captioned-video viewing resulted in the development of listening skills regardless of their year level.

Surprisingly, the control group was shown to score slightly lower at posttest. This outcome might be attributed to the characteristics of the instrument, which assesses learners' capacity to listen for words, names and detailed information (Movers, Cambridge Assessment English, 2018), and the specific vocabulary involved in each form (A and B). As mentioned in the methodology section, learners' exposure to the target language seemed to be constrained to the English class and the materials used in the program. Therefore, it might be possible that Movers A (administered at pretest) included a higher number of the words/phrases taught at school. All in all, this result is not conflicting, since despite the potential differences between the two forms, they were able to detect learners' progress over time.

On the whole, the positive findings obtained in this investigation seem to be consistent with previous studies that have demonstrated that the use of captions supports listening comprehension (Baltova, 1999; Montero Perez et al., 2013, Teng, 2019b) and enhances the development of bottom-up processing skills (Bird & Williams, 2002; Birulés-Muntané & Soto-Faraco, 2016; Charles & Trenkic, 2015). Thus, considering that L2 listening activities may be quite challenging for lower proficiency learners due to the online processing pressure (Kormos et al., 2019; Newton & Nation, 2021), the outcomes of this study support the use of bimodal verbal input to enhance the development of L2 listening skills (Chang, 2011). In addition, the findings of this investigation complement those of earlier studies by corroborating that captioned-video viewing may also be beneficial for primary school learners. In fact, the comparable gains obtained by both year levels suggest that captioned-video viewing may be particularly beneficial at early stages of L2 learning.

6.5.1 Treatment-related factors

Regarding the influence of after-viewing activity type and viewing distribution, the results revealed that none of these factors played a significant role in learners' outcomes over time. The non-significant effects of activity type suggest that learners' intention while processing the input (comprehension or committing target language constructions to memory) did not influence learners' performance at the listening tests. As in Montero Perez et al.'s (2018) study, it could also be the case that both groups mainly focused on comprehension regardless of the type of activity they were asked to complete. That being the case, no differences should be expected between the participants that completed comprehension-focused activities and the ones that were tested on comprehension and target language constructions. As reported in the literature, being forewarned about an upcoming comprehension task may be enough to take the viewing experience more seriously (Rodgers & Webb, 2011; Vanderplank, 2016, Webb, 2015) and allocate enough attentional resources on the multimodal input (Montero Perez et al., 2018).

As for viewing distribution, the non-significant effects of the interaction between class and time already anticipated that the participants benefitted from the treatment regardless of the distance between sessions. In addition, the group comparisons only confirmed that fifth graders scored consistently higher over time. This finding is contrary to previous studies which have suggested that more intensive L2 teaching programs enhance the development of L2 listening skills (Collins & White, 2011; Lightbown & Spada, 2020;

Serrano, 2011; Serrano & Muñoz, 2007). Nonetheless, the total amount of time the participants in this study were exposed to the target language was 110 minutes, while the investigations that compared time distributions in language learning programs measured learners' outcomes after a significantly higher number of hours (e.g. 400 hours in Collins & White's (2011) study). Thus, it is possible that differences between groups may emerge after sustained exposure to captioned videos for a longer period of time.

6.5.2 Cognitive and language-related factors

In regard to the cognitive factors assessed in this study, the results revealed that PSTM was a significant albeit weak predictor of learners' development of listening skills. This finding corroborated that PSTM has a stronger influence at early L2 learning stages (Wen & Jackson, 2022; Wright, 2015), and yields evidence that in L2 comprehension, PSTM plays a significant role in word decoding and the storage of phonological information for further consultation (Grabe, 2009; Wen, 2015). However, given that the analyses also indicated that the effects of PSTM were no longer significant when fitting a model with L2-related factors, the results appear to support the idea that the use of aural and written representations (i.e. bimodal verbal input) may compensate for learners' lower PSTM (Porter, 2017). Previous studies with foreign language learners have shown that the use of captions may neutralize the effects of cognitive factors on L2 learning (Gass et al., 2019; Muñoz, 2022; Pattemore & Muñoz, 2020). This may also explain why complex working memory and visual processing speed were not found to predict learners' performance over time. In addition, these two factors seem to be more relevant in the learning of language aspects that require the integration of verbal and non-verbal input to be learned, such as vocabulary learning at level of meaning recognition and recall (Montero Perez, 2022; Suárez & Gesa, 2019). Thus, given that the development of bottom-up processing skills appears to be enhanced by the synergy between audio and text, that is bimodal verbal input (Bird & Williams, 2002; Birulés-Muntané & Soto-Faraco, 2016; Charles & Trenkic, 2015), the influence of complex working memory and visual processing speed should not be expected.

As regards the L2-related factors, the results indicated that vocabulary knowledge and English text segmentation were the strongest predictors of learners' scores in listening skills over time. In relation to vocabulary knowledge, this finding confirms its strong association with L2 comprehension in different modalities (Montero Perez, 2020; Miralpeix & Muñoz, 2018; Proctor et al., 2005; Stæhr, 2008), and learners' capacity to process the captions with greater ease (Teng, 2019b). Considering that the development of listening skills may be attributed to the support of captions throughout the viewing experience (Bird & Williams, 2002; Birulés-Muntané & Soto-Faraco, 2016; Charles & Trenkic, 2015), it is unsurprising that English text segmentation, a measure of lower-level reading skills, played a significant role in the outcomes. It may be hypothesized that the participants that scored lower in English text segmentation struggled to follow the captions and link aural and written representations. A potential explanation for the non-significant effects of ER efficacy, which is a measure that integrates lower and higher-level reading skills, may be that its effects were overridden by the factors that were more closely associated to reading fluency, that is vocabulary knowledge and English text segmentation.

In regard to the L1-related factors, the significant effects of Spanish text segmentation and SR efficacy seem to confirm that, at initial stages, learners assimilate and accommodate their linguistic infrastructure to the patterns of the L2 (Birch & Fulop, 2021; Jiang et al., 2019; Perfetti et al., 2007); therefore, they may rely on L1 reading skills to compensate, to a certain extent, for L2 knowledge gaps (Birch, 2015; Yamashita, 2002). However, the fact that these factors did not reach significance levels when entered into a model with L2-related relators seem to be in accord with the findings of studies on L2 reading that have demonstrated that L2 factors are stronger predictors of comprehension. Thus, in comparison with L2-related factors, L1-reading skills seemed to play a weaker role in the outcomes (Alderson et al., 2016; Jeon & Yamashita; 2014; Sparks, 2021).

Overall, these findings corroborate the bidirectional relationship of listening and reading skills (Birch & Fulop, 2021; Sparks, 2021; Tunmer & Chapman, 2012; Verhoeven & van Leeuwe, 2012), and highlights the importance of implementing activities that foster the development of both receptive skills (Verhoeven & van Leeuwe, 2012). It is important to acknowledge that the findings of this study differ from those of Tragant et al. (2019) who found that the young learners' exposure to 21 graded readers (with and without audio support) did not lead to significant gains in listening nor reading skills when compared to the control group. As Tragant et al. (2019) explained, the length of the intervention was insufficient to observe differences between the groups, which is congruent with the literature that suggests that the development of receptive skills requires plenty of practice (Grabe & Stoller, 2020). Thus, the positive outcomes obtained in the present study, after a relatively short intervention (11 episodes), may be attributed to the participants' limited (or even non-existent) contact with the L2 outside school, and greater room for learning. Another possible explanation may be related to differences in the materials due to the dynamic nature of images and captions in viewing (Tragant & Pellicer-Sánchez, 2019). Together, the significant gains in listening skills obtained in this study confirm that the use of captioned videos are a suitable alternative to increase learners' exposure to the target language and enhance the development of receptive skills. Indeed, the study by Lindgren and Muñoz (2013) demonstrated that fourth graders' sustained exposure to audiovisual input was a strong predictor of their performance in listening and reading.

VII. English and Spanish reading efficacy

This section focuses on the development of English and Spanish reading efficacy from captioned-video viewing in five groups of primary school learners. As mentioned in the methodology section, reading efficacy integrates the measurement of lower-level (silent reading speed) and higher-level reading skills (comprehension). The analyses reported in this section respond to the following research questions:

1) To what extent does viewing distribution (i.e. shorter vs. longer lags) influence young L2 learners' gains from captioned video viewing?

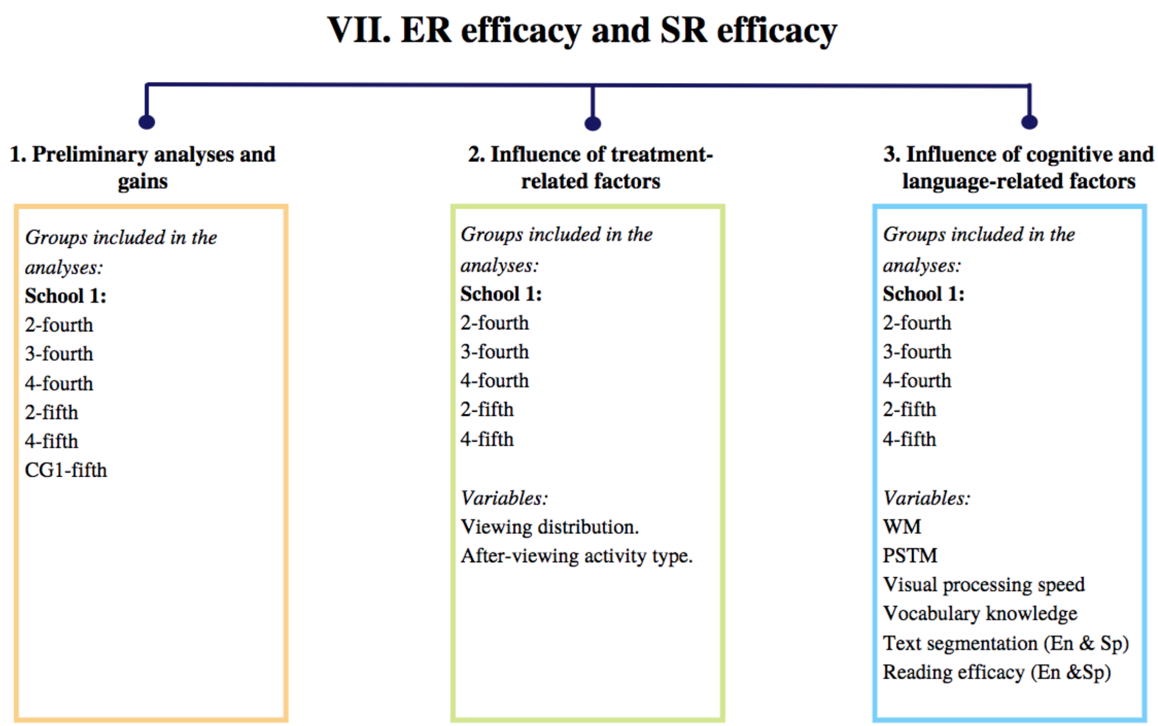
2) In comparison with meaning-focused activities, what are the effects of construction-focused after-viewing activities on L2 learning through captioned-video viewing?

3) To what extent do learner characteristics influence young L2 learners' gains from captioned-video viewing? (i.e. age, vocabulary knowledge, cognitive abilities [phonological short-term memory, complex working memory, and visual processing speed], L2 listening skills, English and Spanish text segmentation), L1 reading habits and attitude towards reading).

The overview of this section is displayed in Figure 50.

Figure 50.

Section 7 overview



7.1 ER efficacy: Preliminary analyses

Given that the target variable (ER efficacy) was not normally distributed, it was square root (SQRT) transformed to reach appropriate normality values ($p > .05$). Then, a set of between-groups comparisons were run with the aim of establishing whether the groups were comparable in ER efficacy at pretest. To start with, the Independent-samples T-test calculated to compare the two year levels indicated that fifth graders outperformed fourth graders at pretest (t (107)= 4.320, p<.001, r=.38) (see descriptive statistics in Table 135). Next, a One-Way ANOVA was run to determine whether the classes from each year level were comparable at pretest. The results revealed that the overall difference between groups (fourth and fifth graders) was statistically significant (F (5)= 4.996, p<.001, η^2 = .195). Specifically, the Tukey pairwise contrasts indicated that 2-fifth scored significantly higher than 3-fourth (p=.024), while 4-fifth obtained a significant higher score than 3-fourth (p <.001), 4-fourth (p=.017), and the control group CG2-fourth (p=.051). In short, these results indicate that the groups in each year level were comparable in terms of ER efficacy at pretest (see Table 136 and Figure 51).

Table 135.

ER efficacy: Descriptive statistics

		Pretest		Posttest		Delayed posttest	
		Mean	(SD)	Mean	(SD)	Mean	(SD)
Class	3-fourth	36.42	26.61	78.21	63.39	88.47	37.14
	4-fourth	42.93	20.46	58.13	30.43	78.66	33.06
	2-fourth	55.41	38.68	69.24	37.06	72.42	26.39
	2-fifth	66.59	35.62	100.61	56.77	112.83	49.93
	4-fifth	79.41	33.45	119.86	45.54	136.30	44.91
	CG2-fourth	50.06	31.24	42.53	21.18	.	.
Year level	Year 4	45.00	30.36	68.98	46.17	80.81	33.03
	Year 5	72.72	34.82	109.59	52.17	124.29	48.46

*Without control group CG2-fourth

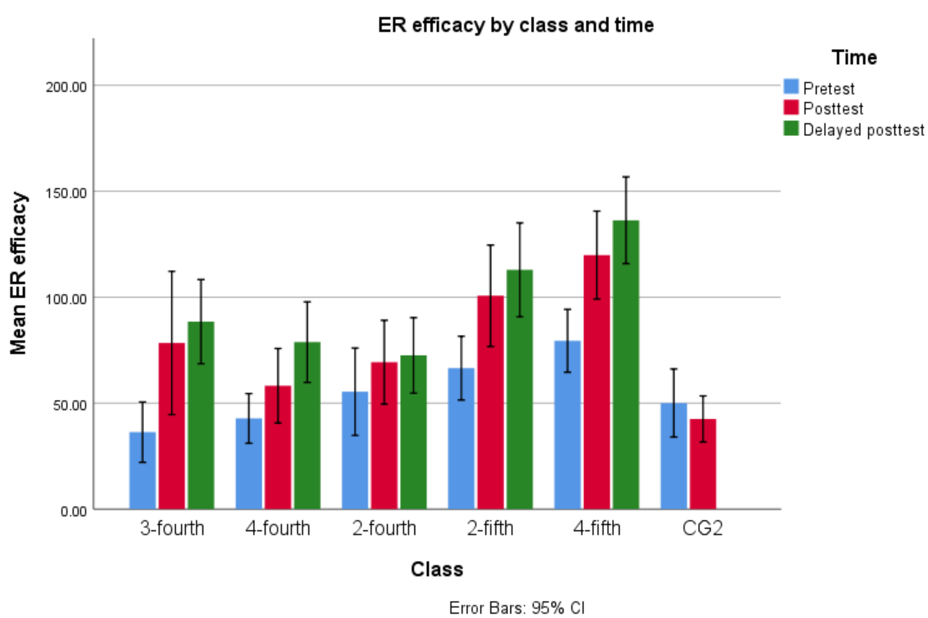
Table 136.

Summary: Between-groups comparisons in terms of ER efficacy pretest scores

Factor	Statistically sig. differences between year levels	Statistically sig. differences between classes
ER efficacy pretest scores	Year 5 > Year 4	2-fifth > 3-fourth 4-fifth > 3-fourth, 4-fourth, CG2-fourth

Figure 51.

ER efficacy: Groups' performance over time



Additionally, Pearson correlations were run in order to explore the relationships between learners' scores at the three testing times and the continuous variables assessed in this study (cognitive and language-related factors) (see Table 137). The results revealed stronger relationships between the ER efficacy scores and L2-related factors, namely vocabulary knowledge, listening skills and English text segmentation, with a large effect size (R^2 > .25) (Larson-Hall, 2010). Although the relationships between ER efficacy and L1-related factors reached statistical significance, their shared variance accounted for a medium effect size in most of the correlations (Larson-Hall, 2010). Yet, ER efficacy was not found to be associated to L1-reading habits and attitude towards reading over time. As for the cognitive factors, they were found to have either a weak or moderate correlation with ER efficacy. Thus, the results obtained from these analyses suggest that ER efficacy was mainly explained by the L2-related factors.

Table 137.

Correlations between ER efficacy scores and learner-related factors

		ER efficacy pretest	ER efficacy posttest	ER efficacy delayed
ER efficacy pretest	Pearson Correlation	1	**.750** (R^2=56)**	**.596** (R^2=35)**
	Sig. (2-tailed)		.000	.000
	N	92	91	82
ER efficacy posttest	Pearson Correlation	**.750** (R^2=56)**	1	**.707** (R^2=49)**
	Sig. (2-tailed)	.000		.000
	N	91	91	82
ER efficacy delayed	Pearson Correlation	**.596** (R^2=35)**	**.707**(R^2=49)**	1
	Sig. (2-tailed)	.000	.000	
	N	82	82	84
PSTM	Pearson Correlation	.231* (R^2=05)	.244* (R^2=05)	.362** (R^2=13)
	Sig. (2-tailed)	.027	.020	.001
	N	92	91	83
Complex WM	Pearson Correlation	.233* (R^2=05)	.272** (R^2=07)	.356** (R^2=12)
	Sig. (2-tailed)	.025	.009	.001
	N	92	91	83
Visual processing speed	Pearson Correlation	.051 (R^2=002)	.215* (R^2=04)	.264* (R^2=06)
	Sig. (2-tailed)	.632	.040	.016
	N	92	91	83
Vocabulary knowledge	Pearson Correlation	**.616** (R^2=37)**	**.724** (R^2=52)**	**.665** (R^2=44)**
	Sig. (2-tailed)	.000	.000	.000
	N	88	88	80
Listening skills	Pearson Correlation	**.503** (R^2=25)**	**.643** (R^2=41)**	**.595** (R^2=35)**
	Sig. (2-tailed)	.000	.000	.000
	N	90	89	81
Segmentation in English	Pearson Correlation	**.534** (R^2=28)**	**.626** (R^2=39)**	**.605** (R^2=36)**
	Sig. (2-tailed)	.000	.000	.000
	N	88	88	80
Segmentation in Spanish	Pearson Correlation	.281** (R^2=07)	.378** (R^2=14)	**.549** (R^2=30)**
	Sig. (2-tailed)	.008	.000	.000
	N	89	88	80
SR efficacy	Pearson Correlation	**.519** (R^2=26)**	.461** (R^2=21)	.487** (R^2=23)
	Sig. (2-tailed)	.000	.000	.000
	N	92	91	82
L1 reading habits and attitude towards reading	Pearson Correlation	-.013 (R^2=0)	.101 (R^2=01)	.078 (R^2=006)
	Sig. (2-tailed)	.901	.348	.485
	N	90	89	82

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

7.2 ER efficacy: Progress over time

To compare the trajectories of both year levels over time, we ran a compound symmetry structure GLMM (linear model) with student identification as subjects, and time as repeated measures. The model was built with learners' scores at the three testing times, and the following fixed effects: time, year level and their interaction. In this model, the scores obtained by the control group were not included. The results yielded significant main effects for year level (F (1,94)= 27.711, p< .001), and time (F (2,166)= 72.697, p< .001), whereas the interaction between these two factors did not reach significance (see Table 138). However, this interaction was kept in the model to further explore the trajectory of each year level (see Figure 52). As shown in Table 139, the experimental groups showed significant improvement from pretest to posttest, and from pretest to delayed posttest, regardless of their year level. In addition, the results indicated that learners' scores increased significantly from posttest to delayed posttest. The significant effects of year level confirmed that fifth graders consistently outperformed fourth graders over time (see Figure 52). Likewise, the time pairwise contrasts per group in Table 140 imply that fifth graders obtained slightly higher gains from the treatment. On the whole, the results indicate that while both year levels benefitted from the treatment, fifth graders obtained marginally higher gains.

Table 138.

Learners' development of ER efficacy over time by year level

Source	F	df1	df2	Sig.
Corrected Model	34.113	5	176	.000
Year level	27.711	1	94	.000
Time	72.697	2	166	.000
Year level * Time	.383	2	166	.682

Probability distribution: Normal
Link function: Identity
a. Target: ER efficacy

Table 139.

ER efficacy: Time pairwise contrasts

Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Year 4 – Year 5	-2.076	.394	-5.264	94	.000	-2.859	-1.293
Pretest - Posttest	-1.739	.188	-9.263	237	.000	-2.191	-1.286
Pretest - Delayed	-2.592	.230	-11.265	123	.000	-3.114	-2.070
Posttest - Delayed	-.854	.214	-3.984	161	.000	-1.277	-.431

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 52.

ER efficacy: The trajectory of each year level over time

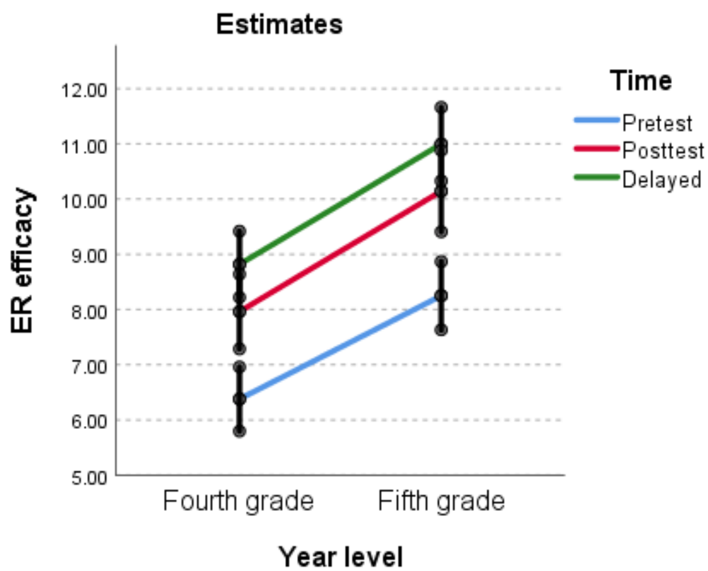


Table 140.

ER efficacy: Time pairwise contrasts by year level

Year level	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Fourth grade	Pretest - Posttest	-1.584	.256	-6.199	261	.000	-2.160	-1.008
	Pretest - Delayed	-2.440	.300	-8.145	185	.000	-3.164	-1.716
	Posttest - Delayed	-.856	.302	-2.831	178	.005	-1.453	-.259
Fifth grade	Pretest - Posttest	-1.893	.275	-6.884	201	.000	-2.514	-1.272
	Pretest - Delayed	-2.744	.349	-7.855	96	.000	-3.596	-1.893
	Posttest - Delayed	-.851	.304	-2.804	146	.006	-1.451	-.251

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

7.2.1 Comparisons between control and experimental groups

In order to assess the performance of the control and the experimental groups from pretest to posttest, we ran a compound symmetry structure GLMM (linear model) with student identification as subjects, and time as repeated measures. The analysis was calculated with learners' listening scores as the target variable, while class, time, and their interaction were entered into the model as fixed factors. As shown in Table 141, the results revealed significant effects for class (F (5,113)= 9.057, p= .011), time (F (1,108)= 70.901, p< .001) and their interaction (F (5,112)= 7.351, p< .001). Specifically, the Bonferroni adjusted results revealed that all the groups improved significantly from pretest to posttest (p< .001), except for the control group, who obtained similar scores at the two testing times (see Table 142).

Table 141.

Learners' ER efficacy over time (pretest and posttest) by class

Source	F	df1	df2	Sig.
Corrected Model	16.162	11	194	.000
Class	9.057	5	113	.000
Time	70.901	1	108	.000
Class * Time	7.351	5	112	.000

Probability distribution: Normal
Link function: Identity
a. Target: ER efficacy

Table 142.

ER efficacy: Time pairwise contrasts by class (pretest and posttest)

Class	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-fourth	Pretest - Posttest	-2.711	.432	-6.282	97	.000	-3.567	-1.854
4-fourth	Pretest - Posttest	-1.016	.401	-2.533	184	.012	-1.807	-.225
2-fourth	Pretest - Posttest	-.954	.329	-2.897	205	.004	-1.604	-.305
2-fifth	Pretest - Posttest	-1.794	.439	-4.084	43	.000	-2.680	-.909
4-fifth	Pretest - Posttest	-2.009	.308	-6.530	205	.000	-2.615	-1.402
CG2-fourth	Pretest - Posttest	.443	.411	1.079	113	.283	-.370	1.257

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

7.2.2 Comparisons between experimental groups

A new model was built in order to examine the trajectory of the experimental groups from pretest to delayed posttest. To this aim, we ran a compound symmetry structure GLMM (linear model) with student identification as subjects, and time as repeated measures. The analysis was calculated with learners' ER efficacy as the target variable, whereas the fixed effects included in the model were class, time, and their interaction. The analyses yielded significant effects for class (F (4, 97)= 10.207, p< .001), time (F (2,164)= 84.153, p< .001), and their interaction (F (8, 176)= 2.608, p= .010) (see Table 143). As shown in Table 144, the Bonferroni pairwise contrasts indicated that all the groups improved significantly from pretest to posttest, and from pretest to delayed posttest. In addition, while all the groups showed improvement from posttest to delayed posttest, it only reached significance levels in the case of 3-fourth and 2-fifth. On the whole, fifth graders (2-fifth and 4-fifth) showed similar gains from pretest to delayed posttest; whereas in the case of fourth graders, 3-fourth appeared to obtain greater benefits from the treatment, particularly when compared to 2-fourth.

Table 143.

Learners' ER efficacy over time by class

Source	F	df1	df2	Sig.
Corrected Model	18.748	14	196	.000
Class	10.207	4	97	.000
Time	84.153	2	164	.000
Class * Time	2.608	8	176	.010

Probability distribution: Normal
Link function: Identity
a. Target: ER efficacy

Table 144.

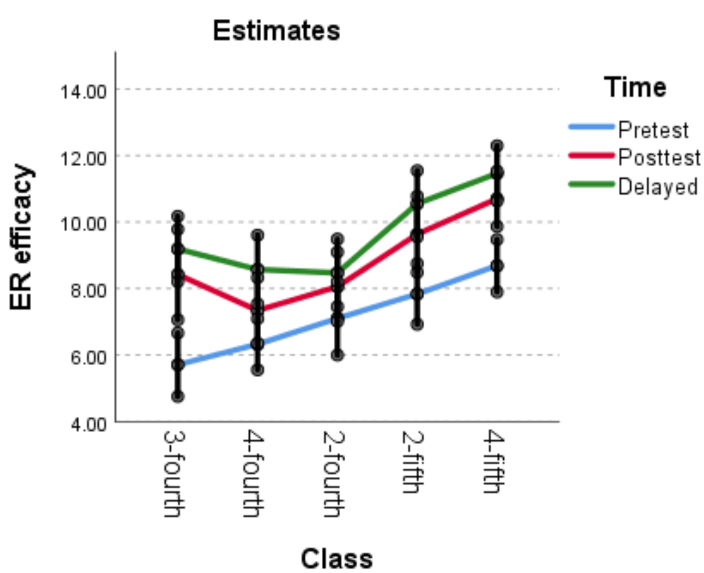
ER efficacy: Time pairwise contrasts by class

Class	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-fourth	Pretest - Posttest	-2.711	.432	-6.282	181	.000	-3.686	-1.736
	Pretest - Delayed	-3.481	.466	-7.470	142	.000	-4.611	-2.352
	Posttest - Delayed	-.771	.587	-1.313	76	.193	-1.940	.399
4-fourth	Pretest - Posttest	-1.016	.401	-2.533	252	.012	-1.805	-.226
	Pretest - Delayed	-2.248	.459	-4.897	252	.000	-3.354	-1.142
	Posttest - Delayed	-1.232	.413	-2.987	252	.006	-2.162	-.302
2-fourth	Pretest - Posttest	-.954	.329	-2.897	252	.012	-1.748	-.161
	Pretest - Delayed	-1.370	.477	-2.874	252	.012	-2.505	-.234
	Posttest - Delayed	-.415	.481	-.864	252	.389	-1.362	.532
2-fifth	Pretest - Posttest	-1.794	.439	-4.084	97	.000	-2.795	-.794
	Pretest - Delayed	-2.713	.498	-5.445	75	.000	-3.934	-1.493
	Posttest - Delayed	-.919	.412	-2.229	152	.027	-1.734	-.105
4-fifth	Pretest - Posttest	-2.014	.308	-6.549	252	.000	-2.755	-1.273
	Pretest - Delayed	-2.776	.487	-5.694	91	.000	-3.887	-1.665
	Posttest - Delayed	-.762	.447	-1.704	107	.091	-1.648	.124

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Figure 53.

ER efficacy: Interaction between class and time



Additionally, we built a model to explore learners' trajectory in silent reading speed (number of words read per minute) over time. Specifically, we ran a series of repeated measures (time) compound-symmetry structure GLMMs (linear model) with learners' scores in silent reading speed as the target factor, and the following independent variables: class, time, and their interaction (see descriptive statistics in Table 145). As shown in Table 146, the results revealed significant effects for class (F (4, 91)= 7.413, p< .001), Time (F (2, 138)= 20.065, p< .001) and their interaction (F (8, 165)= 2.060, p= .043). The Bonferroni pairwise comparisons indicated that in year 4, only 3-fourth improved significantly from pretest to posttest (p<.05), and from pretest to delayed posttest (p<.05), whereas 2-fourth's progress from pretest to posttest only approached statistical significance (p=. 073). As for fifth graders, both groups improved significantly from pretest to posttest (p<.05) and from pretest to delayed posttest (p<.05) (see Table 147 and Figure 54). Thus, learners' progress in reading efficacy was not necessarily associated to their improvement in reading speed but in comprehension.

Table 145.

ER efficacy: Experimental groups' scores in terms of WPM at each testing time

		Pretest		Posttest		Delayed posttest	
		Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*
Class	3-fourth	85.19	*(28.51)*	106.19	*(56.59)*	112.56	*(28.76)*
	4-fourth	110.50	*(36.31)*	106.36	*(19.81)*	112.21	*(27.78)*
	2-fourth	94.62	*(31.09)*	112.38	*(48.71)*	104.00	*(27.16)*
	2-fifth	116.33	*(35.91)*	133.29	*(54.67)*	140.23	*(50.51)*
	4-fifth	131.73	*(39.09)*	152.29	*(36.12)*	158.19	*(40.17)*

Table 146.

ER efficacy: Experimental groups' progress in terms of WPM over time

Source	F	df1	df2	Sig.
Corrected Model	7,565	14	242	,000
Class	7,413	4	91	,000
Time	20,065	2	138	,000
Class * Time	2,060	8	165	,043

Probability distribution: Normal
Link function: Identity
a. Target: WPM

Figure 54.

ER efficacy: Groups' progress over time in terms of WPM

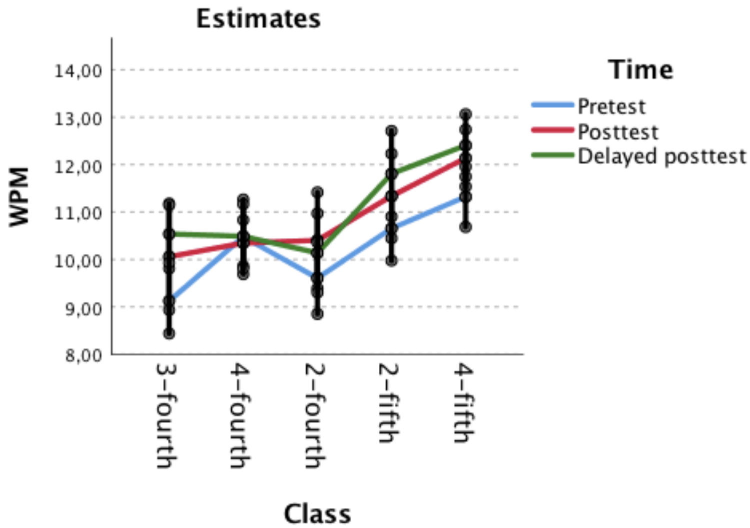


Table 147.

ER efficacy: Pairwise contrasts of the outcomes obtained by each group (WPM) over time

Class	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-fourth	Pretest - Posttest	-,934	,370	-2,526	81	,027	-1,778	-,090
	Pretest - Delayed posttest	-1,413	,340	-4,157	127	,000	-2,237	-,588
	Posttest - Delayed posttest	-,479	,469	-1,021	45	,313	-1,424	,466
4-fourth	Pretest - Posttest	,132	,346	,380	202	1,000	-,603	,866
	Pretest - Delayed posttest	-,010	,344	-,029	229	1,000	-,691	,671
	Posttest - Delayed posttest	-,141	,303	-,467	252	1,000	-,872	,589
2-fourth	Pretest - Posttest	-,796	,350	-2,275	148	,073	-1,642	,051
	Pretest - Delayed posttest	-,530	,259	-2,046	252	,084	-1,114	,054
	Posttest – Delayed posttest	,266	,242	1.097	252	,274	-,211	,743
2-fifth	Pretest - Posttest	-,683	,297	-2,299	126	,046	-1,357	-,009
	Pretest - Delayed posttest	-1,151	,319	-3,608	111	,001	-1,927	-,376
	Posttest - Delayed posttest	-,469	,303	-1,546	140	,124	-1,068	,131
4-fifth	Pretest - Posttest	-,816	,231	-3,534	252	,001	-1,336	-,295
	Pretest - Delayed posttest	-1,083	,262	-4,132	252	,000	-1,715	-,451
	Posttest - Delayed posttest	-,267	,283	-,944	222	,346	-,825	,291

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

The results obtained in terms of ER efficacy over time are summarized in Table 148.

Table 148.

Summary of findings: ER efficacy scores over time

Analysis	Outcome
Significant improvement from pretest to posttest (p<.05).	Both year levels. All the groups, except for the control group. 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth.
Significant improvement from pretest to delayed posttest (p<.05).	Both year levels. All the groups, except for the control group. 2-fourth, 3-fourth, 4-fourth, 2-fifth, 4-fifth.

7.3 ER efficacy: The influence of treatment-related factors

7.3.1 After-viewing activity type

In order to measure the influence of after-viewing activity type (see descriptive statistics in Table 149) on learners' scores over time, we performed a compound symmetry structure GLMM (linear model) with student identification as subjects, and time as repeated measures. The analyses were calculated with learners' ER efficacy as the target variable, and the following fixed factors: year level, time, activity type, and their interaction. The results did not yield significant effects for activity type (F (1,96)= .208, p= .649), suggesting that learners' scores did not vary as a function of the type of activity the participants were asked to complete after watching each episode (see Table 150). Next, in order to determine whether the non-significant effects of activity type may be associated to differences in proficiency, a new model was built by adding vocabulary knowledge as a fixed factor. However, the same results were obtained, corroborating that activity type was not a significant predictor of ER efficacy over time.

Table 149.

ER efficacy: Descriptive statistics per activity type

		ER efficacy pretest		ER efficacy posttest		ER efficacy delayed	
		Mean	(SD)	Mean	(SD)	Mean	(SD)
Activity type	Meaning focused	60.04	(36.81)	90.84	(47.62)	102.72	(48.14)
	Construction focused	57.60	(34.26)	87.46	(58.32)	103.54	(46.00)

Table 150.

ER efficacy: The influence of activity type on learners' outcomes

Source	F	df1	df2	Sig.
Corrected Model	42.582	4	123	.000
Year level	27.842	1	96	.000
Activity type	.208	1	96	.649
Time	72.007	2	169	.000

Probability distribution: Normal
Link function: Identity
a. Target: ER efficacy

7.3.2 Viewing distribution

To study the influence of viewing distribution on learners' outcomes, we ran a series of compound-symmetry structure GLMMs (linear model) with student identification as subjects and time as repeated measures. Learners' ER efficacy scores over time were set as outcome variable. In these analyses, only the scores of the participants that watched either two or four episodes a week were included since 3-fourth did not have a counterpart in year 5. The fixed factors entered into the model were as follows: viewing distribution, year level, time and all possible two-way and three-way interactions. Then, by following a step back procedure, the non-significant interactions were removed one by one until the best fitted model was obtained. As shown in Table 151, viewing distribution did not predict learners' ER efficacy (F (1, 73)= .755, p= .388). Thus, a new model was built by entering vocabulary knowledge as covariate to rule out the possibility that viewing distribution did not emerge as significant predictor due to differences in proficiency. The analyses corroborated that the distance between sessions did not predict learners' scores in ER efficacy.

As for the significant interaction between year level and time (see Table 151), it confirms that fifth graders obtained greater benefits from the treatment (see Figure 55). Yet, this interaction was only obtained when 3-fourth was removed from the sample.

Table 151.

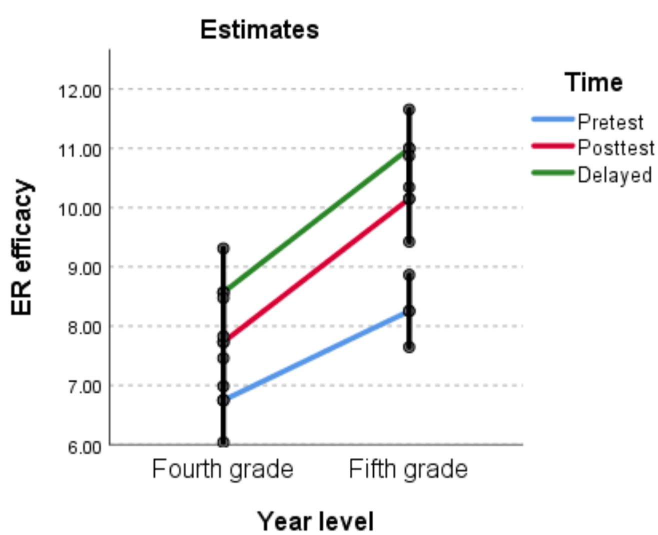
ER efficacy: The influence of viewing distribution on learners' outcomes

Source	F	df1	df2	Sig.
Corrected Model	20.266	6	134	.000
Viewing distribution	.755	1	73	.388
Year level	24.027	1	84	.000
Time	50.199	2	153	.000
Year level * Time	3.286	2	153	.040

Probability distribution: Normal
Link function: Identity
a. Target: ER efficacy

Figure 55.

ER efficacy: Interaction between year level and time (without 3-fourth)



7.3.3 The influence of treatment-related factors: Summary of findings

The analyses reported in this section assessed the influence of treatment-related factors (i.e. after-viewing activity type and viewing distribution) on the development of ER efficacy from captioned-video viewing over time. In summary, the results indicated that:

- Neither activity type nor viewing distribution predicted learners' performance as regards ER efficacy.

- Therefore, learners equally benefitted from the treatment, regardless of the number of episodes they had to watch a week and the type of activity they were asked to complete.

7.4 ER efficacy: The influence of cognitive and language-related factors

This section focuses on the analyses that examined the influence of cognitive and language-related factors on the development of ER efficacy. Before running the analyses, collinearity tests were calculated between the predictor variables to prevent this factor from affecting the results obtained in the analyses (Pallant, 2016). To start with, we ran a series of compound-symmetry structure GLMMs (linear models) with student identification as subjects and time as repeated measures. Learners' ER efficacy scores at the three testing times were set as outcome variable. Additionally, multiple linear regressions were performed with posttest and delayed posttest ER efficacy scores to calculate the exact contribution of the cognitive and language-related variables to learners' performance.

7.4.1 Cognitive factors

In order to assess the influence of the cognitive factors on the development of ER efficacy, the following variables were entered into the model: class, time, PSTM, complex working memory, visual processing speed, and all possible two-way interactions. A backward elimination procedure was used to determine the best fitted model. Thus, the non-significant interactions and main factors were removed from the model one by one. In this case, the analyses indicated that complex working memory did not contribute to the learning process significantly (p >.05), therefore, this factor was removed from the best fitted model. As summarized in Table 152, the analyses yielded significant main effects for class (F (4,90)= 8.642, p< .001) and PSTM (F (1,188)= 13.538, p< .001), suggesting that learners' performance as regards ER efficacy was significantly influenced by PSTM. The results also showed significant interactions between class and time (F (1,155)= 2.328, p= .022) and visual processing speed and time (F (2, 113)= 5.176, p= .007). Figure 56 indicates that the influence of visual processing speed increased with time.

Table 152.

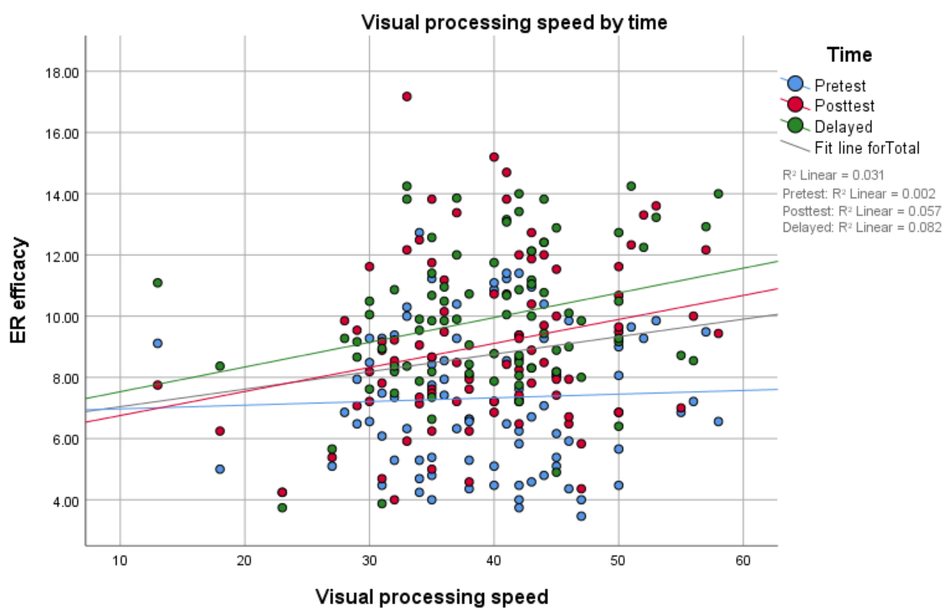
ER efficacy: The influence of cognitive factors on learners' scores

Source	F	df1	df2	Sig.
Corrected Model	20.752	18	190	.000
Class	8.642	4	90	.000
PSTM	13.538	1	188	.000
Time	.739	2	109	.480
Visual processing speed	.164	1	77	.687
Class * Time	2.328	8	155	.022
Visual processing speed * Time	5.176	2	113	.007

Probability distribution: Normal
Link function: Identity
a. Target: ER efficacy

Figure 56.

ER efficacy: The interaction between visual processing speed and time



7.4.2 Language-related factors

In order to assess the influence of language-related factors on the development of ER efficacy, a series of Generalized Linear Mixed Model (linear model) with repeated measures (time) compound-symmetry structure were calculated. The first model was built with L2-related factors and all possible two-way interactions. Specifically, the following variables were entered into the model: class, time, vocabulary knowledge, listening skills and English text segmentation. The best fitted model was determined by a backward elimination procedure. As shown in Table 153, the results revealed significant main effects of class (F (4,100)= 5.965, p< .001), listening skills (F (1,67)= 8.261, p= .005), vocabulary knowledge (F (1,51)= 30.384, p< .001), and time (F (2,88)= 3.232, p= .044). In addition, the analyses yielded significant interactions between class and English segmentation (F (4,97)= 3.301, p= .014), class and time (F (8,140)= 4.173, p< .001), and class and time (F (2,101)= 6.228, p= .003). The significant main effects of vocabulary knowledge suggested that this factor significantly increased the odds of scoring higher in ER efficacy. As for the significant interaction between class and English text segmentation, the results suggest that the extent to which the participants relied on English segmentation varied among the groups. Thus, as displayed in Figure 57, 2-fourth was the group that relied the most on English segmentation. In regard to the significant interaction between listening skills and time, Figure 58 indicates that the positive relationship between ER efficacy and listening skills strengthened over time, especially between pretest and posttest. Concerning the significant interaction between class and time, the outcomes obtained by 3-fourth and 4-fourth were magnified since they made great progress considering their significantly lower proficiency level (see Figure 59).

Table 153.

ER efficacy: The influence of L2-related factors on learners' scores

Source	F	df1	df2	Sig.
Corrected Model	30.031	23	166	.000
Class	5.965	4	100	.000
Listening skills	8.261	1	67	.005
Vocabulary knowledge	30.384	1	51	.000
English text segmentation	1.890	1	79	.173
Time	3.232	2	88	.044
Class * English text segmentation	3.301	4	97	.014
Class * Time	4.173	8	140	.000
Listening skills * Time	6.228	2	101	.003

Probability distribution: Normal
Link function: Identity. a. Target: ER efficacy.

Figure 57.

Interaction between English segmentation and class

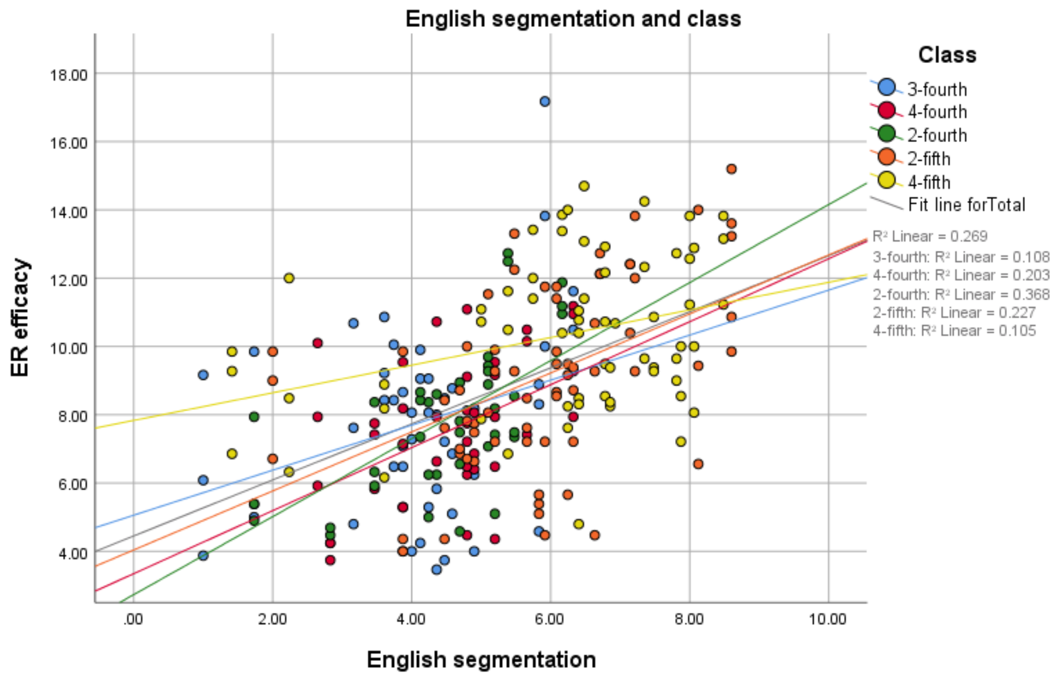


Figure 58.

Interaction between listening skills and time

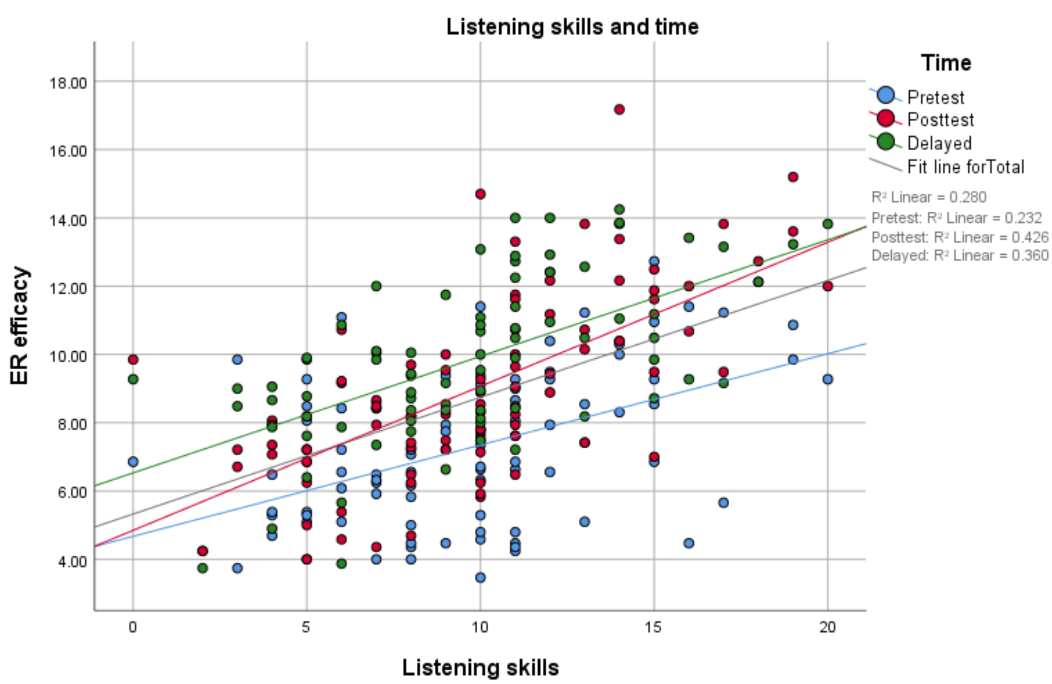
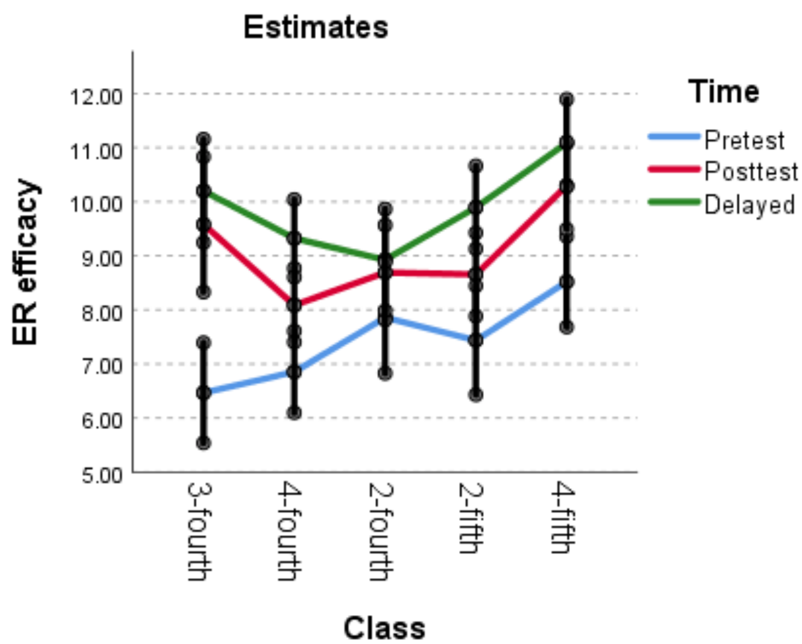


Figure 59.

Interaction between class and time



Then, a new model was built to explore the influence of L1-related factors on learners' ER efficacy scores. To this aim, class, time, Spanish text segmentation, SR efficacy, and L1 reading habits and attitudes towards reading were entered into the model as fixed factors. In addition, all possible two-way interactions were included. Prior to the analyses, the visual binning tool in SPSS was used to transform the reading habits variable into categorical (higher vs. lower scores) since its relationship with the target variable was not linear. The non-significant interactions and main effects were removed from the model one by one until the best fitted model was obtained. This was the case of L1 reading habits and attitudes, which was not found to contribute to the model significantly (p >.05). As displayed in Table 154, the results showed significant main effects for class (F (4,74)= 4.808, p= .002) and SR efficacy (F (1,42)= 26.072, p< .001). As for the significant interaction between SR efficacy and class (F (4,61)= 2.674, p= .040), Figure 60 indicates that fourth graders relied more on SR efficacy, particularly 2-fourth. Concerning the significant interaction between Spanish segmentation and time (F (2,79)= 6.146, p= .003), the positive relationship between ER efficacy and Spanish segmentation increased with time (see Figure 61), showing a medium effect at posttest ($R^2 > .09$) and a large effect at delayed posttest ($R^2 > .25$; Larson-Hall, 2010).

Table 154.

ER efficacy: The influence of L1-related factors on learners' scores

Source	F	df1	df2	Sig.
Corrected Model	23.819	22	158	.000
Class	4.808	4	74	.002
SR efficacy	26.072	1	42	.000
Spanish segmentation	3.319	1	139	.071
Time	.547	2	76	.581
Class * SR efficacy	2.674	4	61	.040
Class * Time	3.563	8	157	.001
Spanish segmentation * Time	6.146	2	79	.003

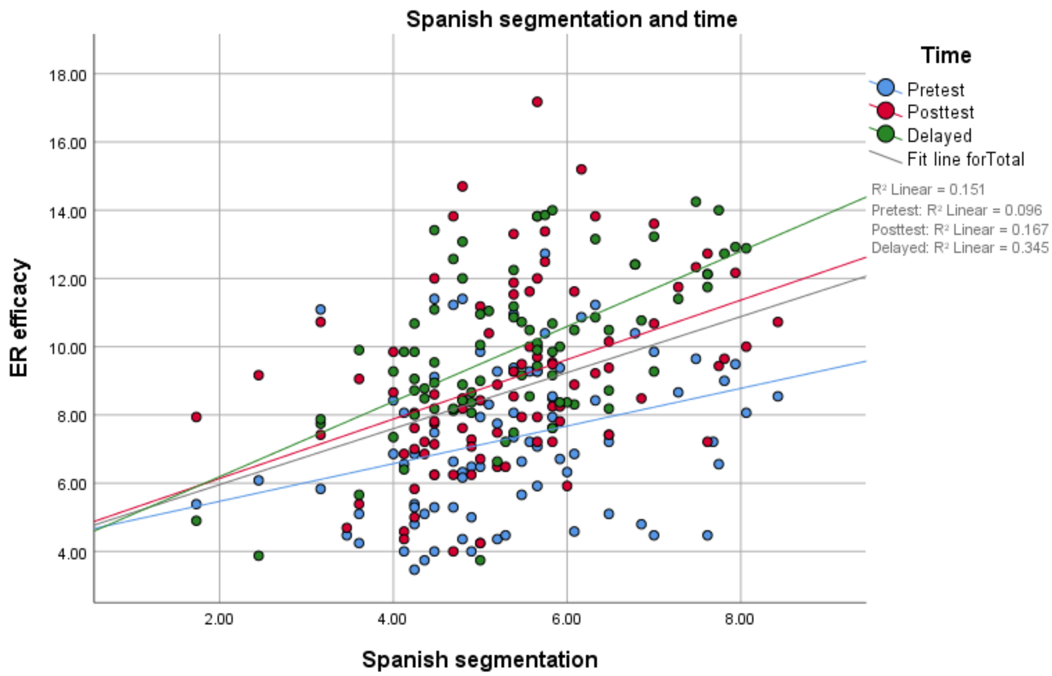
Probability distribution: Normal
Link function: Identity
a. Target: ER efficacy

Figure 60.

Interaction between SR efficacy and class

Figure 61.

Interaction between Spanish segmentation and time



Next, a new model was fitted to compare the effects of the L1 and L2-related factors that reached significance levels in the analyses above. To this aim, the fixed factors entered to the model were as follows: class, time, listening skills, vocabulary knowledge, English segmentation, Spanish segmentation, SR efficacy, and all possible two-way interactions. By following a backward elimination procedure, the non-significant interactions and main effects were removed from the model one by one until the best fitted model was obtained. This was the case for English text segmentation, whose effects were overshadowed by the presence of the L1-related factors. The results revealed significant main effects for class (F (4,83)= 4.212, p= .002), listening skills (F (1,59)= 9.218, p= .004), vocabulary knowledge (F (1,64)= 27.721, p< .001), and SR efficacy (F (1,83)= 12.301, p= .001). In addition, the analyses yielded significant interactions between listening skills and time (F (2,99)= 5.942, p= .004), Spanish segmentation and time (F (2,73)= 3.696, p= .030), Class and SR efficacy (F (4,81)= 2.478, p= .050), as well as class and time (F (8,130)= 3.935, p< .001) (see Table 155).

Table 155.

ER efficacy: The influence of L1 and L2-related factors on learners' scores

Source	F	df1	df2	Sig.
Corrected Model	28.093	26	180	.000
Class	4.212	4	83	.004
Time	.567	2	64	.570
Listening skills	9.218	1	59	.004
Vocabulary knowledge	27.721	1	64	.000
SR efficacy	12.301	1	83	.001
Spanish segmentation	.021	1	99	.885
Listening skills * Time	5.942	2	99	.004
Spanish segmentation * Time	3.696	2	73	.030
Class * SR efficacy	2.478	4	81	.050
Class * Time	3.935	8	130	.000

Probability distribution: Normal
Link function: Identity
a. Target: ER efficacy

Finally, a new model was built with the statistically significant L1 and L2-related factors in the model above, and the cognitive factors that were found to influence learners' performance in ER efficacy, that is PSTM and visual processing speed. Thus, the predictors entered into the model were as follows: class, time, listening skills, vocabulary knowledge, SR efficacy, Spanish segmentation, PSTM, visual processing speed, and all possible two-way interactions. The non-significant effects and interactions were removed from the model one by one until the best fitted model was obtained. The analyses indicated that PSTM was no longer significant, since its effects appeared to be overridden by the language-related factors. On the whole, the results replicated the ones obtained above and added the significant interaction between visual processing speed and time (F (3,102)= 3.663, p= .015) (see Table 156).

Table 156.

ER efficacy: The influence of cognitive, L1 and L2-related factors on learners' scores

Source	F	df1	df2	Sig.
Corrected Model	27.468	29	178	.000
Class	4.311	4	81	.003
Time	2.845	2	77	.064
Listening skills	10.760	1	64	.002
Vocabulary knowledge	32.814	1	67	.000
SR efficacy	12.928	1	86	.001
Spanish segmentation	.095	1	100	.759
Listening skills * time	4.958	2	90	.009
Spanish segmentation * Time	3.400	2	72	.039
Class * SR efficacy	2.174	4	81	.079
Class * Time	3.980	8	122	.000
Visual processing speed * Time	3.663	3	102	.015

Probability distribution: Normal
Link function: Identity
a. Target: ER efficacy

In order to calculate the contribution of each factor on learners' ER efficacy scores at posttest and delayed posttest, we performed multiple linear regressions for each testing time. The predictor variables included were as follows: listening skills, vocabulary knowledge, SR efficacy, Spanish segmentation and visual processing speed. As for ER efficacy posttest score, Spanish segmentation and visual processing speed were not found to contribute significantly (p >.05), so they were removed from the analyses. The results indicated that listening skills, vocabulary knowledge and SR efficacy predicted 57% of the variance at posttest ($F(3, 83) = 39.437, p < .001, R^2 = .573$). The standard coefficients indicated that vocabulary knowledge was the strongest predictor (β=44%, p<.001), followed by listening skills (β=27%, p=.007) and SR efficacy (β=17%, p=.042). As for ER efficacy at delayed posttest, the results revealed that neither SR efficacy nor visual processing speed contributed to learners' ER reading scores significantly, so they were eliminated from the analyses. The results indicated that listening skills, vocabulary knowledge, and Spanish text segmentation predicted 54% of the variance at delayed posttest ($F(3, 72) = 31.433, p < .001, R^2 = .549$). The standard coefficients showed that again, vocabulary knowledge was the strongest predictor of learners' performance in ER efficacy (β=39%, p<.001), followed by Spanish segmentation (β=32%, p<.001). In the case of listening skills, this factor was only found to be marginally significant (β=19%, p=.073).

Considering the relationship between listening skills and the development of ER efficacy, and the fact that listening skills were also tested at three testing times, we additionally ran correlations between the three scores obtained in ER efficacy and listening skills. As shown in Table 157, the relationship between ER efficacy and listening skills strengthened over time (large effect size, R^2=38), suggesting that learners' progress encompassed the development of both receptive skills.

Table 157.

Relationship between ER efficacy and listening skills over time

		Listening pretest	Listening posttest	Listening delayed
ER efficacy pretest	Pearson Correlation	.486** (R^2=23)	.494**	.462**
	Sig. (2-tailed)	.000	.000	.000
	N	90	92	92
ER efficacy posttest	Pearson Correlation	.656**	.618** (R^2=38)	.645**
	Sig. (2-tailed)	.000	.000	.000
	N	89	91	91
ER efficacy delayed	Pearson Correlation	.589**	.664**	.618** (R^2=38)
	Sig. (2-tailed)	.000	.000	.000
	N	81	84	84

**. Correlation is significant at the 0.01 level (2-tailed).

7.4.3 The influence of cognitive and language-related factors: Summary of findings

Overall, the results reported in this section indicate that:

- Among the cognitive factors, PSTM and visual processing speed played a significant albeit weak role in the development of ER efficacy. Still, visual processing was the only cognitive variable that remained in the model that included cognitive and language-related factors.
- As for the L2-related factors, the results indicated that vocabulary knowledge was the strongest predictor of ER efficacy over time. In addition, listening skills seemed to play a greater role at posttest than delayed posttest. As for English text segmentation, it was found to be a significant but weak predictor of learners' scores since its effects were overshadowed by the L1-related factors. In addition, not all the groups relied on

English segmentation to the same extent. More precisely, 2-fourth was found to rely the most on this factor.

- In regard to the L1 related factors, the influence of Spanish text segmentation was found to be stronger at delayed posttest. As for SR efficacy, this factor seemed to play a more prominent role at posttest. Besides, fourth graders appeared to rely more on SR efficacy when compared to fifth graders, particularly in the case of 2-fourth. Finally, learners' scores concerning L1 reading habits and attitude towards reading were not shown to explain their outcomes over time.
- The analyses that compared the contribution of cognitive, L1 and L2-related factors to the development of ER efficacy clearly indicated that vocabulary knowledge was the strongest predictor of learners' performance over time.

The results concerning the influence of treatment and learner-related factors are summarized in Table 158.

Table 158.

Summary: Predictors of ER efficacy

Factors	Outcomes
Viewing distribution	***Non-significant.*** However, the results suggest that the fourth graders that watched fewer episodes a week (2-fourth) relied more on English text segmentation and SR reading efficacy, which might be associated to a more effortful processing of input and their lower gains from the treatment.
Year level	***Significant***. Both year levels improved significantly from pretest to posttest, and from pretest to delayed posttest. However, the results also suggested that fifth graders benefited more from the treatment.
After-viewing activity type	***Non-significant.***
Time	***Significant.*** However, learners' progress is not necessarily associated to an increase in reading speed but comprehension, particularly in 4-fourth and 2-fourth.
Vocabulary knowledge	***Significant.*** The strongest predictor of learners' performance in ER efficacy over time.
English segmentation	***Significant*** but weaker predictor of ER efficacy. 2-fourth appeared to rely the most on this factor.
Spanish segmentation	***Significant.*** This factor was found to be a stronger predictor at delayed posttest.

SR efficacy	***Significant.*** This factor was found to be a stronger predictor at posttest. In addition, its role seemed to play a more prominent role in fourth graders, especially 2-fourth.
Listening skills	***Significant.*** It appeared to play a more important role at posttest. When considering learners' listening scores at the three testing times, the results indicated that the relationship between listening and ER efficacy strengthened over time.
Phonological short-term memory	***Statistically significant*** but a weaker predictor of ER efficacy. Its effects were overridden by the language-related factors.
Working memory	***Non-significant.***
Visual processing speed	***Significant.*** Its relationship with ER efficacy strengthened over time.

7.5 SR efficacy: Preliminary analyses

The analyses in the following sub-sections were performed in order to determine the extent to which the treatment also supported the development of SR efficacy over time. In addition, the analyses assessed the extent to which learners' outcomes were influenced by the L1 and L2-related factors selected for this purpose. Given that the target variable (SR efficacy) was not normally distributed, it was square root (SQRT) transformed to reach appropriate normality values ($p > .05$).

First of all, a set of between-groups comparisons were run with the aim of establishing whether the groups were comparable in SR efficacy at pretest. To start with, the Independent-samples T-test calculated to compare the two year levels indicated that their difference at pretest approached statistical significance (t (106)= 1.923, p=.057, r=.18) (see descriptive statistics in Table 159). Next, a One-Way ANOVA was run to determine whether the classes from each year level were comparable at pretest. The results revealed that the overall difference between groups (fourth and fifth graders) was marginally significant (F (5)= 2.179, p=.062, η^2 = .097). Specifically, the Tukey pairwise contrasts indicated that the difference that approached statistical significance was between 4-fifth and 3-fourth (p=.055), since these were the groups that scored the highest and lowest, respectively. In sum, these results indicate that the groups in each year level were comparable in SR efficacy at pretest (see Table 159 and Figure 62).

Table 159.

SR efficacy: Descriptive statistics

		SR efficacy pretest		SR efficacy posttest		SR efficacy delayed	
		Mean	(SD)	Mean	(SD)	Mean	(SD)
Class	3-fourth	71.06	23.83	85.50	43.37	85.50	43.37
	4-fourth	77.71	36.03	92.71	47.42	91.21	36.28
	2-fourth	88.50	54.62	109.94	58.18	87.36	44.65
	2-fifth	84.92	37.24	127.54	46.97	125.91	58.07
	4-fifth	115.82	58.85	169.52	62.97	150.38	69.03
	CG1-fifth	80.94	27.29	87.13	43.66	.	.
Year level	Year 4	79.15	40.17	96.20	50.16	87.95	40.47
	Year 5*	99.70	50.68	147.13	58.34	137.86	64.09

*Without the control group.

Figure 62.

SR efficacy: Groups' trajectory over time

In order to assess whether the variability in learners' scores was influenced by L1 and/or L2-related factors, the following variables were considered: L2 vocabulary knowledge, ER efficacy, Spanish text segmentation and L1 reading habits and attitude towards reading. Therefore, we calculated Pearson correlations to explore the relationships between SR efficacy over time and the abovementioned variables. The results indicated that the strongest correlations were found at posttest with vocabulary knowledge (R^2=42), ER efficacy (R^2=42) and Spanish text segmentation (R^2=39). The relationship with L1 reading habits and attitudes towards reading appeared to strengthened over time (see Table 160).

Table 160.

Correlations between SR efficacy scores and learner-related factors

		Pretest	Posttest	Delayed
SR efficacy pretest	Pearson Correlation	1	.744** (R^2=55)	.567** (R^2=32)
	Sig. (2-tailed)		.000	.000
	N	108	107	82
SR efficacy posttest	Pearson Correlation	.744** (R^2=55)	1	.754** (R^2=56)
	Sig. (2-tailed)	.000		.000
	N	107	107	82
SR efficacy delayed	Pearson Correlation	.567** (R^2=32)	.754** (R^2=56)	1
	Sig. (2-tailed)	.000	.000	
	N	82	82	84
L1 reading habits and attitude towards reading	Pearson Correlation	.078 (R^2=006)	.204 (R^2=04)	.223* (R^2=04)
	Sig. (2-tailed)	.465	.055	.044
	N	90	89	82
Vocabulary knowledge	Pearson Correlation	.467** (R^2=21)	**.653** (R^2=42)**	**.579** (R^2=33)**
	Sig. (2-tailed)	.000	.000	.000
	N	88	88	80
ER efficacy	Pearson Correlation	**.523** (R^2=27)**	**.662** (R^2=43)**	**.558** (R^2=31)**
	Sig. (2-tailed)	.000	.000	.000
	N	92	91	82
Spanish segmentation	Pearson Correlation	**.515** (R^2=26)**	**.628** (R^2=39)**	**.580** (R^2=33)**
	Sig. (2-tailed)	.000	.000	.000
	N	89	88	80

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

7.6 SR efficacy: Progress over time

To compare the trajectories of both year levels over time, we ran a compound symmetry structure GLMM (linear model) with student identification as subjects, and time as repeated measures. The model was built with learners' scores at the three testing times, and the following fixed effects: time, year level and their interaction. In this model, the scores obtained by the control group were not included. The results yielded significant main effects for year level (F (1,94)= 16.986, p< .001), and time (F (2,144)= 35.383, p< .001), and a significant interaction between year level and time (F (2,144)= 6.590, p= .002) (see Table

161). As displayed in Figure 63 and Table 161, both year levels improved significantly from pretest to posttest, but only fifth graders showed significant gains from pretest to delayed posttest. In addition, the benefits were evidently higher for fifth graders.

Table 161.

Learners' development of SR efficacy over time by year level

Source	F	df1	df2	Sig.
Corrected Model	19.034	5	163	.000
Year level	16.986	1	94	.000
Time	35.383	2	144	.000
Year level * Time	6.590	2	144	.002

Probability distribution: Normal
Link function: Identity
a. Target: SR efficacy

Figure 63.

SR efficacy: Interaction between year level and time

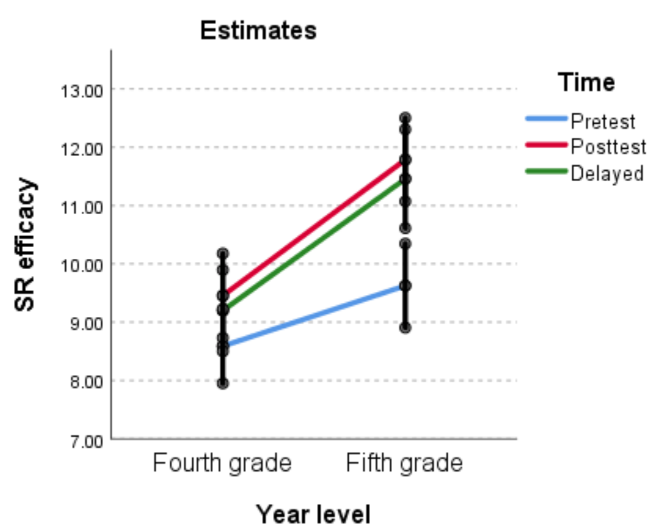


Table 162.

SR efficacy: Time pairwise contrasts by year level

Year level	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
Fourth grade	Pretest - Posttest	-.866	.232	-3.733	261	.001	-1.425	-.307
	Pretest - Delayed	-.610	.313	-1.946	168	.107	-1.318	.099
	Posttest - Delayed	.256	.271	.947	254	.345	-.277	.790
Fifth grade	Pretest - Posttest	-2.162	.275	-7.855	185	.000	-2.827	-1.497
	Pretest - Delayed	-1.833	.394	-4.655	63	.000	-2.738	-.929
	Posttest - Delayed	.328	.323	1.016	101	.312	-.313	.969

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

7.6.1 Comparisons between control and experimental groups

In order to assess the performance of the control and the experimental groups from pretest to posttest, we ran a compound symmetry structure GLMM (linear model) with student identification as subjects, and time as repeated measures. The analysis was calculated with SR efficacy scores as the target variable, while class, time, and their interaction were entered into the model as fixed factors. As shown in Table 163, the results revealed significant effects for class (F (5,103)= 4.732, p= .001), time (F (1,117)= 56.909, p< .001) and their interaction (F (5,118)= 4.549, p= .001). Specifically, the Bonferroni adjusted comparisons revealed that all the groups improved significantly from pretest to posttest, except for the control group and 3-fourth (see Table 164).

Table 163.

Learners' development of SR efficacy at pretest and posttest by class

Source	F	df1	df2	Sig.
Corrected Model	10.592	11	184	.000
Class	4.732	5	103	.001
Time	56.909	1	117	.000
Class * Time	4.549	5	118	.001

Probability distribution: Normal
Link function: Identity
a. Target: SR efficacy

Table 164.

SR efficacy: Time pairwise contrasts by control and experimental groups

Class	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	Upper
3-fourth	Pretest - Posttest	-.616	.427	-1.442	95	.153	-1.465	.232
4-fourth	Pretest - Posttest	-.797	.398	-2.005	177	.046	-1.582	-.013
2-fourth	Pretest - Posttest	-1.177	.362	-3.249	197	.001	-1.891	-.462
2-fifth	Pretest - Posttest	-2.116	.375	-5.640	76	.000	-2.863	-1.369
4-fifth	Pretest - Posttest	-2.217	.405	-5.469	73	.000	-3.025	-1.409
CG1-fifth	Pretest - Posttest	-.233	.350	-.664	203	.507	-.923	.458

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

7.6.2 Comparisons between experimental groups

A new model was built in order to examine the trajectory of the experimental groups from pretest to delayed posttest. To this aim, we ran a compound symmetry structure GLMM (linear model) with student identification as subjects, and time as repeated measures. The analysis was calculated with learners' SR efficacy as the target variable, whereas the fixed effects included in the model were class, time, and their interaction. The analyses yielded significant effects for class (F (4, 91)= 5.538, p< .001), time (F (2,123)= 30.916, p< .001), and their interaction (F (8, 131)= 2.242, p= .028) (see Table 165). As shown in Table 166, the Bonferroni pairwise contrasts indicated that fifth graders improved significantly from pretest to posttest, and from pretest to delayed posttest; whereas in year 4, only 2-fourth improved significantly from pretest to posttest. On the whole, the treatment appeared to benefit fifth graders rather than fourth graders (see Table 167).

Table 165.

Learners' development of SR efficacy by class

Source	F	df1	df2	Sig.
Corrected Model	8.898	13	179	.000
Class	5.538	4	91	.000
Time	30.916	2	123	.000
Class * Time	2.242	8	131	.028

Probability distribution: Normal
Link function: Identity
a. Target: SR efficacy

Table 166.

SR efficacy: Time pairwise contrasts by experimental group

Class	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
3-fourth	Pretest - Posttest	-.616	.427	-1.442	138	.455	-1.652	.420
	Pretest - Delayed	-.616	.427	-1.442	138	.455	-1.652	.420
	Posttest - Delayed	.000	.000	.000	252	1.000	.000	.000
4-fourth	Pretest - Posttest	-.797	.398	-2.005	252	.138	-1.755	.161
	Pretest - Delayed	-.870	.548	-1.589	130	.229	-2.112	.372
	Posttest - Delayed	-.073	.613	-.119	104	.906	-1.288	1.142
2-fourth	Pretest - Posttest	-1.177	.362	-3.249	252	.004	-2.049	-.304
	Pretest - Delayed	-.266	.696	-.382	79	.704	-1.651	1.120
	Posttest - Delayed	.911	.626	1.454	94	.298	-.516	2.338
2-fifth	Pretest - Posttest	-2.116	.375	-5.640	209	.000	-3.021	-1.210
	Pretest - Delayed	-1.976	.560	-3.527	54	.002	-3.267	-.684
	Posttest - Delayed	.140	.489	.287	74	.775	-.833	1.113
4-fifth	Pretest - Posttest	-2.227	.406	-5.486	159	.000	-3.210	-1.245
	Pretest - Delayed	-1.671	.554	-3.019	73	.007	-2.938	-.404
	Posttest - Delayed	.556	.413	1.347	159	.180	-.260	1.372

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Table 167.

Summary of findings: ER efficacy scores over time

Analysis	Outcome
Significant improvement from pretest to posttest (p<.05).	Both year levels. Only 2-fourth, 2-fifth and 4-fifth.
Significant improvement from pretest to delayed posttest (p<.05).	Fifth graders. Only 2-fifth and 4-fifth.

7.7 SR efficacy: The influence of L1 and L2-related factors

In order to assess the influence of L1 and L2-related factors on the development of SR efficacy, a series of Generalized Linear Mixed Model (linear model) with repeated measures (time) compound-symmetry structure were calculated. The following fixed factors were entered into the model: class, time, vocabulary knowledge, ER efficacy, Spanish segmentation, L1 reading habits and attitude towards reading (categorical), and all possible two-way interactions. The visual binning tool in SPSS was used to transform L1 reading habits and attitudes into categorical variable (i.e. two categories: higher vs. lower score). The best fitted model was determined by a backward elimination procedure.

As shown in Table 168, the results revealed significant main effects of class (F (4,85)= 3.415, p= .012), vocabulary knowledge (F (1,104)= 7.750, p= .006), ER efficacy (F (1,50)= 19.649, p< .001), and Spanish segmentation (F (1,47)= 35.525, p< .001). In addition, the analyses yielded a significant interaction between vocabulary knowledge and time (F (2,226)= 6.410, p= .002), as well as a marginally significant interaction between time and L1 reading habits (F (2,128)= 3.007, p= .053). As depicted in Figure 64, the strength of the relationship between vocabulary knowledge and time was shown to increase at posttest with a large effect size (R^2=42%; Larson-Hall, 2010). As for the interaction between time and L1 reading habits and attitude towards reading, the Bonferroni pairwise contrasts indicated that the participants improved significantly from pretest to posttest, regardless of their L1 reading habits. Yet, the ones that showed a better attitude towards reading and were more frequent readers appeared to obtain greater gains. Likewise, they were the only group that showed significant gains from pretest to delayed posttest (see Table 169 and Figure 65).

Table 168.

SR efficacy: The influence of L1 and L2-related factors

Source	F	df1	df2	Sig.
Corrected Model	24.511	14	137	.000
Class	3.415	4	85	.012
Vocabulary knowledge	7.750	1	104	.006
ER efficacy	19.649	1	50	.000
Spanish segmentation	35.525	1	47	.000
Time	.983	2	226	.376
L1 reading habits	.976	1	90	.326
Vocabulary knowledge * Time	6.410	2	226	.002
Time * L1 reading habits	3.007	2	128	.053

Probability distribution: Normal
Link function: Identity
a. Target: SR efficacy

Figure 64.

SR efficacy: Interaction between vocabulary knowledge and time

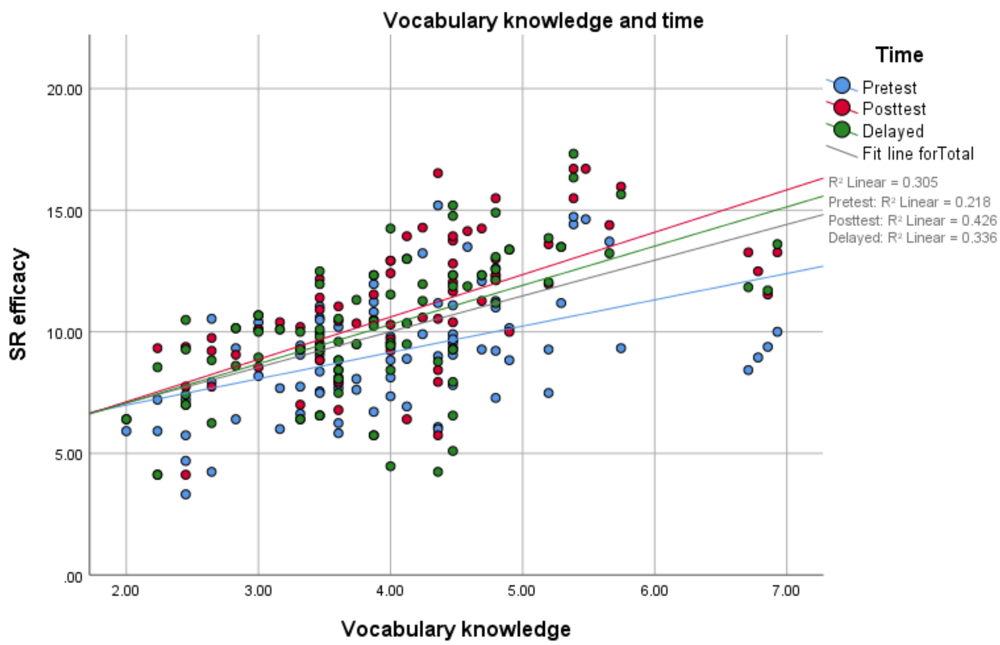


Figure 65.

SR efficacy: Interaction between time and L1 reading habits and attitude towards reading

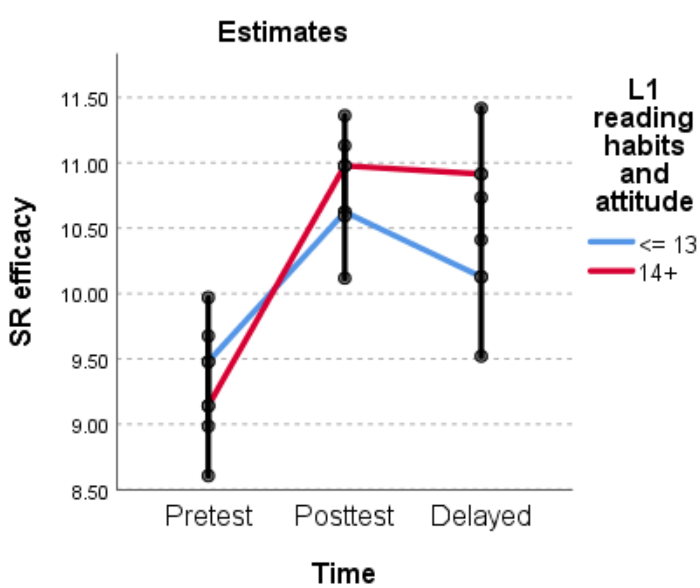


Table 169.

SR efficacy: Time pairwise contrasts by reading habits group

L1 reading habits and attitude towards reading	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
<= 13	Pretest - Posttest	-1.146	.241	-4.750	221	.000	-1.729	-.564
(Lower scores)	Pretest - Delayed	-.650	.338	-1.921	78	.117	-1.423	.123
	Posttest - Delayed	.496	.328	1.515	86	.133	-.155	1.148
14+	Pretest - Posttest	-1.837	.282	-6.507	148	.000	-2.520	-1.153
(Higher scores)	Pretest - Delayed	-1.774	.330	-5.371	115	.000	-2.523	-1.024
	Posttest - Delayed	.063	.230	.276	226	.783	-.390	.516

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Considering that both, L1 and L2-related factors played a significant role in the outcomes, we performed multiple linear regressions in order to calculate the contribution of each factor on learners' SR efficacy scores at pretest, posttest and delayed posttest. The predictor variables included were as follows: vocabulary knowledge, ER efficacy, Spanish segmentation and L1 reading habits and attitude towards reading. At pretest, vocabulary knowledge and L1 reading habits were not found to contribute significantly (p >.05), so they were removed from the analyses. The results indicated that ER efficacy and Spanish segmentation predicted 39% of the variance at pretest ($F(2, 83) = 29.461, p < .001, R^2 = .393$). The standard coefficients indicated that ER efficacy (β=39%, p<.001) and Spanish segmentation (β=39%, p<.001) showed a similar predictive value. As for SR efficacy at posttest, the results revealed that L1 reading habits did not contribute to learners' SR reading scores significantly, so this variable was eliminated from the analyses. The results indicated that vocabulary knowledge, ER efficacy and Spanish segmentation predicted 65% of the variance at posttest ($F(3,81)=54.643, p < .001, R^2 = .657$). Although the standard coefficients showed that Spanish segmentation (β=42%, p<.001) and ER efficacy (β=38%, p<.001) were the strongest predictors of SR efficacy at posttest, vocabulary knowledge was also found to explain a large extent of the variance (β=22%, p<.001). In regard to the SR efficacy scores at delayed posttest, all the predictor variables reached significance levels. Specifically, the results indicated that the L1 and L2-related factors altogether predicted 54% of the variance at delayed posttest ($F(4, 70) = 23.403, p < .001, R^2 = .548$). The strongest predictor was

Spanish text segmentation (*β*=33%, $p<.001$), followed by ER efficacy (*β*=30%, $p=.003$), vocabulary knowledge (*β*=25%, $p=.015$), and L1 reading habits and attitude towards reading (*β*=19%, $p=.017$)

7.8 SR efficacy: Summary of findings

Overall, the results reported as regards SR efficacy indicated that:

- Both year levels improved significantly from pretest to posttest, but only fifth graders showed significant progress from pretest to delayed posttest. When exploring each group's trajectory, the results indicated that in year 4, only 2-fourth obtained significant gains from pretest to posttest.
- 3-fourth, 4-fourth and the control group's scores did not change significantly over time.
- The results indicated that the treatment was clearly more beneficial for fifth graders, which is a finding that may be attributed to their significantly higher proficiency in both languages, English and Spanish.
- As expected, Spanish segmentation was found to be a strong significant predictor of SR efficacy scores.
- ER efficacy remained as a significant predictor of SR efficacy at the three testing times.
- The shared variance between vocabulary knowledge and SR efficacy increased significantly at posttest. Vocabulary knowledge was only found to predict SR efficacy scores at posttest and delayed posttest, implying a potential relationship between the treatment and learners' gains.
- The relationship between L1 reading habits and SR efficacy reached significance at posttest and delayed posttest, being stronger at the last testing time. Specifically, the results indicated that the participants that showed a more positive attitude towards reading and were more frequent readers obtained greater gains from pretest to posttest. Furthermore, their progress was less prone to decay, this is why this was the only group that showed significant gains from pretest to delayed posttest. Indeed, the multiple linear regression indicated that this factor predicted learners' SR efficacy scores at delayed posttest.

7.9 ER and SR efficacy: Discussion

This section reported learners' gains as regards the development of English and Spanish reading efficacy from captioned-video viewing. In addition, the analyses assessed the influence of treatment and learner-related factors on ER efficacy over time. On the whole, the results indicated that the treatment enhanced the development of ER efficacy in both, fourth and fifth graders; whereas learners' gains in SR efficacy were only shown to be significant in fifth graders and 2-fourth, which were the groups that scored higher at pretest. These outcomes are further explained in the sections below.

7.9.1 ER efficacy: Gains

Overall, the findings obtained in the present study lend support to the use of captioned videos to foster the development of L2 reading skills in primary school learners. The positive outcomes are in line with those of previous studies conducted in L1 contexts, which found that the use of L1 captions supported the development of L1 reading skills (e.g. Kothari et al., 2002; Linebarger, 2001; Linebarger et al., 2010). Yet, the analyses also indicated that learners' progress was not necessarily associated to their increase in silent reading speed (number of words read per minute) but their improvement in comprehension. Thus, a possible explanation for this might be that a higher amount of exposure to print and practice allowed the participants to devote less attention to text decoding, and allocate more attentional resources on the comprehension process (Nassaji, 2014; Sadoski & Paivio, 2013). The literature suggests that the automatization of lower-level reading skills requires plenty of exposure to print (Grabe, 2009; Grabe & Jiang, 2018; Grabe & Stoller, 2020), therefore, it might be possible that a higher number of episodes would have resulted in higher reading fluency in the younger groups (i.e. 2-fourth and 4-fourth). This result concurs with the ones obtained by Linebarger et al. (2010), where the participants (second and third graders) benefitted from captioned-video viewing in terms of word recognition and non-word reading, but not reading fluency since six episodes were not enough to improve in this regard.

It is important to note that the yields in this investigation were higher than in previous studies with primary school learners, where the use of bimodal verbal input has not necessarily been found to foster the development of L2 reading skills after a relatively short intervention. The study by Tragant et al. (2019) with fifth graders indicated that the participants exposed to 18 sessions with graded readers (with and without audio support) did not obtain greater gains in reading comprehension nor reading speed when compared to the control group. There are different possible explanations for the conflicting results. To start with, the participants in Tragant and colleagues' (2019) study were already familiar with the use of graded readers in English, and were consistently encouraged to read in their L1 (extensive reading). By contrast, the participants in this study were not familiar with the use of graded readers in English and reported little (or non-existent) exposure to captioned videos. Moreover, only 24% of the participants read books in Spanish every day or almost every day. Thus, the sudden increase in their exposure to onscreen text may explain their significant gains in ER efficacy. In addition, the participants in this study were not able to control the viewing process as in Tragant et al.'s (2019) investigation, where each child could manipulate the audio (e.g. pause) and read the books at their own pace. Thus, the viewers might have made a greater effort to process the captions while they were available on screen.

Another important finding in the present study was that, overall, the treatment was conducive to greater gains in fifth graders, who showed significant improvement in both, reading speed and ER efficacy. This result may be accounted by their significantly higher proficiency level in English and Spanish, as well as their cognitive maturation, implying that fifth graders may be more efficient learners than fourth graders (Holmes & Myles, 2019). Yet, there was great variability among the participants, while in fourth grade the experimental groups did not seem to benefit from the treatment to the same extent. Thus, the following sections provide important insights into the influence of treatment and learner-related factors on the outcomes.

7.9.2 ER efficacy: Treatment-related factors

As regards the influence of treatment-related factors, the results did not yield significant effects for activity type nor viewing distribution. In terms of activity type, the outcomes suggested that learners focused their attention on captions regardless of the type of activity they were asked to complete after watching each episode. This result is consistent with the data obtained in the investigation by Tragant and Pellicer-Sánchez (2019), which examined fifth graders' eye movements while watching an episode of Charlie and Lola for comprehension purposes. The empirical evidence indicated that the participants spent longer on text than images, suggesting that reading is a key aspect of the viewing process. Thus, the positive outcomes obtained in the present study may be accounted by learners' inherent reliance on captions, irrespective of their intention to commit some target language constructions to memory.

With respect to viewing distribution, the results did not point to a clear advantage of any of the treatment conditions, except for the higher gains obtained by 3-fourth and 4-fourth despite their lower proficiency level. While this finding might be attributed to their greater room for improvement (Raudszus et al., 2021), the analyses that assessed the influence of language-related factors on the outcomes suggested that viewing distribution did play a role in year 4. Specifically, the results indicated that 2-fourth relied more on English text segmentation and SR efficacy compared to the rest of the groups, implying that watching fewer episodes a week made the reading of captions more effortful. This falls in line with the results obtained by Greving and Richter (2021), which indicated that the reading of topic-related texts under short-spaced intervals was perceived as less difficult and enhanced the use of top-down processing skills. In addition, this finding seems to be consistent with that of Collins and White (2012), who showed that the concentration of instructional time moderated, to a certain extent, the influence of individual differences. By the same token, Serfaty and Serrano's (2022a, 2022b) findings on lag effects indicated that in the learning of more difficult language aspects, shorter lags between training sessions facilitated the learning process in slower and less proficient participants. Thus, our findings may further support the idea that shorter gaps between sessions might be recommended for less proficient and younger learners. The fact that the potential effects of viewing distribution emerged in year 4 and not in year 5 may not only be accounted by their level of proficiency but also their age. As mentioned in the literature review, middle childhood (6-11/12 years old) is a stage of big changes in physical, socio-emotional and cognitive development (Delgiudice, 2018; Holmes & Myles, 2019; Myles, 2022), which may explain why a single year made a difference in the results.

7.9.3 ER efficacy: Cognitive and language-related factors

With respect to the influence of cognitive and language-related factors on the development of ER efficacy, the results implied that, overall, the language-related factors were stronger predictors of learners' performance over time. To start with, the results indicated that PSTM and visual processing speed played a significant, albeit weaker, role in the outcomes. As for visual processing speed, its relationship with ER efficacy was found to strengthen over time, which is a finding that might be explained by the fact that learners' visual attention was split between onscreen text and imagery. Thus, learners' capacity to process the visual input with greater ease and speed might have influenced the extent to which they benefitted from the treatment.

Concerning PSTM, the results did not only confirm that this factor has a stronger influence at early L2 learning stages (Wen & Jackson, 2022; Wright, 2015), but also that PSTM plays a role in the reading process, namely in word decoding and the storage of phonological information for further consultation (Grabe, 2009; Wen, 2015, p.50). The fact that PSTM did not reach significance when fitting a model with language-related factors is in agreement with the results obtained by Porter (2017), which showed that PSTM had a weak influence on the development of reading comprehension. By the same token, the empirical evidence suggests that in primary school learners, cognitive factors are not strong predictors of L2 reading in comparison with L2-related factors (Alderson et al., 2016; Pattemore & Serra, 2021). The results are also likely to be related to the processing of bimodal verbal input, since the simultaneous processing of audio and text has been found to facilitate text decoding and moderate the effects of cognitive factors (Kormos et al., 2019; Muñoz, 2022; Pattemore & Muñoz, 2020; Pellicer-Sánchez, 2022; Suárez et al., 2021).

As for the L2-related factors, the analyses yielded significant effects for vocabulary knowledge, listening skills and English text segmentation. Based on the Simple View of Reading model (Gough & Tunmer, 1986; Hoover & Gough, 1990; Tunmer & Chapman, 2012), reading comprehension is mainly explained by word decoding and oral general language comprehension, which involves listening skills and vocabulary knowledge. Thus, considering that the importance of these factors has been detected in L1 and L2 contexts, it is not surprising that vocabulary knowledge, listening skills and English text segmentation emerged as significant predictors of ER efficacy (Alderson et al., 2016; Birch & Fulop, 2021; Jeon & Yamashita, 2014; Proctor et al., 2005; Sparks, 2021; Verhoeven & van Leeuwe, 2012).

The fact that English text segmentation was found to be a weaker predictor of ER efficacy may have different explanations. As mentioned earlier, previous research has demonstrated that the simultaneous processing of aural and written input facilitates text decoding in L2 learners (Kormos et al., 2019; Mayer et al., 2020; Pellicer-Sánchez, 2022). Thus, the presence of audio support might have moderated (to a certain extent) the influence of this factor on the outcomes. In addition, the evidence suggests that in late primary school years, lower-level reading skills may play a weaker role in comparison with listening skills, whose contribution to reading comprehension has been shown to increase over the years (Verhoeven & van Leeuwe, 2012). Indeed, the results confirmed the bidirectional association between listening skills and ER efficacy (Verhoeven & van Leeuwe, 2012), since their progress as a result of the intervention appeared to go hand in hand.

Regarding the L1-related factors, the results revealed significant effects for SR efficacy and Spanish text segmentation, but not for L1 reading habits and attitude towards reading. On the whole, this finding provides evidence of the relationship between L1 and L2 reading (Koda, 2007; Llanes, 2018; Nassaji, 2013; Perfetti et al., 2007; Tragant et al., 2019), and further supports the idea that learners use their first language infrastructure (e.g. L1 reading strategies and mechanisms) to deal with L2 reading (Birch and Fulop, 2021; Perfetti et al., 2007). Therefore, as their knowledge of the target language increases, learners may gradually assimilate and accommodate their linguistic infrastructure to the patterns of the L2 (Birch & Fulop, 2021; Perfetti et al., 2007). Although it is hard to determine the extent to which learners' performance is explained by L2-related factors and L1 reading (Birch & Fulop, 2021), the evidence collected in this investigation confirmed that while both groups of variables played a key role in the development of ER efficacy, the L2-related factors, namely vocabulary knowledge and listening skills were the strongest predictors of ER efficacy over time (Alderson et al., 2016; Jeon & Yamashita, 2014; Proctor et al., 2005; Sparks, 2021; Verhoeven and van Leeuwe, 2012; Yamashita, 2002). Thus, the results seem to confirm that learners' reliance on SR efficacy may compensate to only a certain extent for their knowledge gaps (Birch, 2015; Yamashita, 2002).

The result that SR efficacy had a more prominent role at posttest, while Spanish text segmentation predicted learners' performance at delayed posttest may be difficult to interpret. However, a possible explanation for this might be that along the viewing experience, learners compensated for their knowledge gaps and weak word recognition skills through the use of contextual information and the application of higher-level L1 reading strategies (Grabe, 2009), which may be more associated to SR efficacy. In addition, given that thousands of hours of practice are required for the automatization of lower-level reading skills (Grabe & Stoller, 2020), the relationship between ER efficacy and Spanish text segmentation at delayed posttest might also be accounted by learners' greater engagement with reading. Thus, considering that learners' scores in L1 reading habits and attitude towards reading only correlated significantly with Spanish text segmentation, it may be reasonable to hypothesize that the role of Spanish text segmentation and learners' higher scores in ER efficacy at delayed posttest might be associated to their reading practice, or its possible increase as a result of the treatment.

The high complexity and demands of L2 reading may explain the lower popularity of this activity outside the classroom (De Wilde et al., 2019; Muñoz, 2020b; Peters, 2018). This is why in lower proficiency learners, their struggle and reluctance to read might become a vicious circle that prevents them from practicing harder to improve over time (Birch & Fulop, 2021). With this in mind, the findings on ER efficacy that emerged from this study suggest that captioned-video viewing is an advantageous activity for primary school learners that has the potential of breaking the vicious circle of learners' reluctance to read. As mentioned earlier, learners need plenty of practice in order to make the reading process less effortful and improve their comprehension (Grabe & Stoller, 2020; Nassaji, 2014). Thus, the facilitating effects of bimodal verbal input and the support of imagery may encourage them to watch videos in English and increase the required exposure to print. However, it is important to note that these findings also implied that viewing does not replace reading, which is why these two activities should be complemented to boost learners' gains (see the section below).

7.9.4 SR efficacy: Gains and the influence of language-related factors

One of the most interesting findings in this study was that some of the experimental groups were shown to benefit in terms of SR efficacy through the viewing of captioned videos in English. Specifically, the groups that scored higher at pretest, that is 2-fourth, 2-fifth and 4-fifth showed significant improvement from pretest to posttest, while only fifth graders obtained significant gains from pretest to delayed posttest. In other words, the treatment was particularly beneficial for the older learners, who were evidently more proficient in both languages. The observed increase in SR efficacy might be attributed to interlanguage reading since at earlier stages, learners appear to rely on their L1 linguistic infrastructure to process the input in the target language (Birch & Fulop, 2021; Jiang et al., 2019). The literature suggests that in the case of L1-Spanish learners of English, the assimilation and accommodation of their linguistic infrastructure to the patterns of the L2 may be facilitated by their common alphabetic writing system (Birch & Fulop, 2021; Perfetti et al., 2007). Thus, the findings of this study suggest that the application of L1 reading strategies while processing the videos in the target language may also support the development of L1 reading skills.

The fact that learners' gains were influenced by both, L1 and L2-related factors (i.e. L2 vocabulary knowledge, ER efficacy and Spanish text segmentation) may not be surprising given that the two languages were somehow at play while processing the input. Indeed, these factors may explain why fifth graders obtained greater gains from the treatment. Seeing that along middle childhood learners are still developing their L1 reading skills, fifth graders may have had a more robust linguistic infrastructure to cope with the demands of the target language (Ehri, 2005; Holmes & Myles, 2019). Moreover, their additional year of instruction implied greater knowledge of the L2, which is key to facilitate the processing of the target language (Birch & Fulop, 2021; Grabe, 2009; Nassaji, 2014). Thus, fifth graders' greater resources to succeed in the processing of captions may have had a positive effect on the development of SR efficacy. Additionally, based on young learners' slower learning rate (Holmes & Myles, 2019; Muñoz, 2006), it might also be hypothesized that fourth graders required longer exposure to captioned videos to obtain significant gains in SR efficacy.

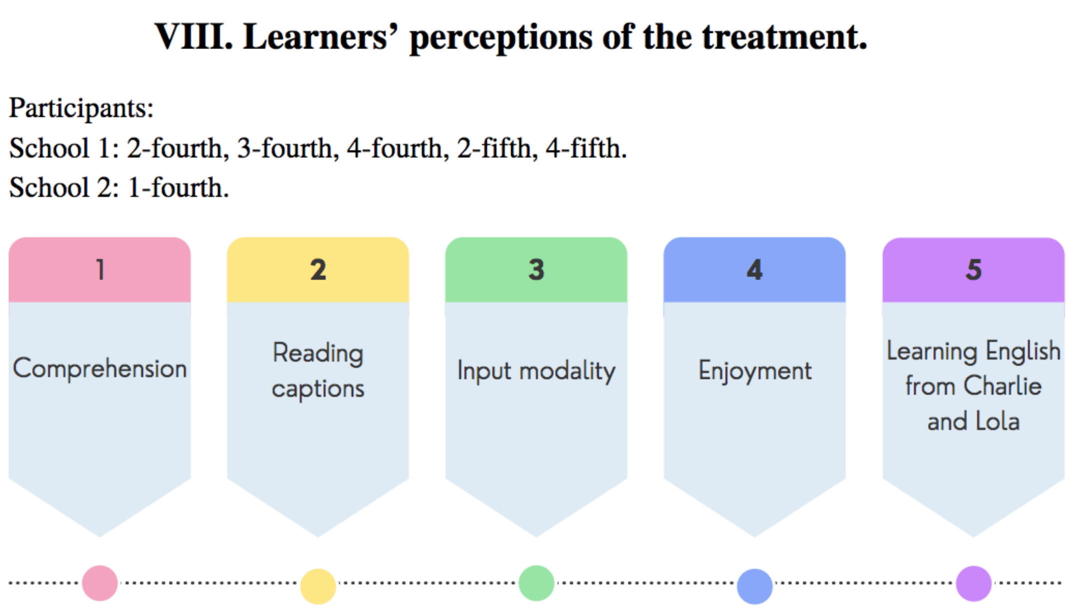
It is important to note that the analyses also indicated that learners' progress in SR efficacy were attributed to the effects of the treatment together with learners' L1 reading habits and attitude towards reading. First, the contribution of captioned-video viewing was visible in the comparison of the experimental groups' performance to that of the control group, as well as in the analyses that showed evidence of the more prominent role of vocabulary knowledge at posttest. As for L1 reading habits and attitude towards reading, the results indicated that the students that read more frequently and showed a more positive attitude towards reading obtained greater gains at posttest, and were able to keep a similar performance at delayed posttest. Hence, according to these data we can infer that it was the synergy between reading and viewing that accounted for learners' gains in SR efficacy over time. To put it in another way, the contribution of captioned-video viewing to the development of SR efficacy may be limited if this task is not complemented by L1 reading activities.

VIII. Learners' perceptions of the treatment

This section focuses on learners' perceptions of the treatment. Specifically, it reports learners' responses to the questionnaire that was administered at the end of treatment (see Appendix 22), which focused on the following areas: comprehension, learners' ability to read captions, their attention to verbal (aural/written) and non-verbal input, levels of enjoyment and students' willingness to watch more videos in class, as well as their learning perceptions from the treatment. The analyses reported in this section respond research question 5: *What are students' perceptions of the viewing experience? How do their answers contribute to the interpretation of the quantitative findings?* With this in mind, the results are interpreted in light of the literature and the findings obtained in relation to the key variables explored in this investigation, such as year level, viewing distribution and activity type. Figure 66 presents an overview of this section.

Figure 66.

Section 8 overview

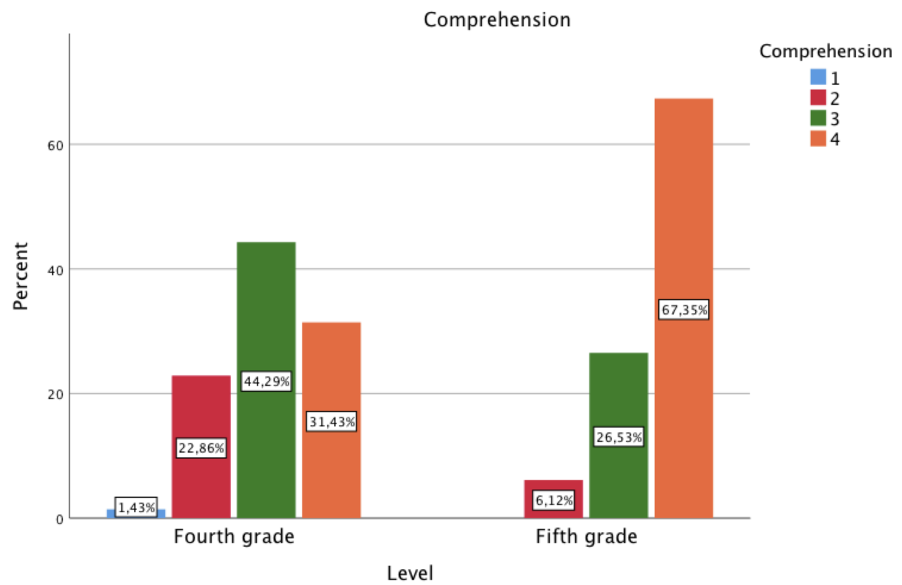


8.1 Comprehension

The first question focused on overall learners' comprehension of the episodes in a Likert scale from 0 (low) to 4 (high). As seen in Figure 67, Year-4 students' responses ranged from 1 to 4 and most responses (around 75%) concentrated on levels 3 and 4. In turn, year-5 students' responses ranged from 2 to 4 and more than two-thirds of the participants selected the highest level of comprehension (67%). To compare both age groups statistically, we computed a Mann-Whitney U Test. The results indicate that fifth-grade students reported a significantly higher level of comprehension than fourth graders (*Mdn*= 4 vs. *Mdn*= 3, respectively; U= 2396. 5, z= 3.9, $p < .001$, $r = .33$).

Figure 67.

Learners' overall comprehension of the episodes per year level

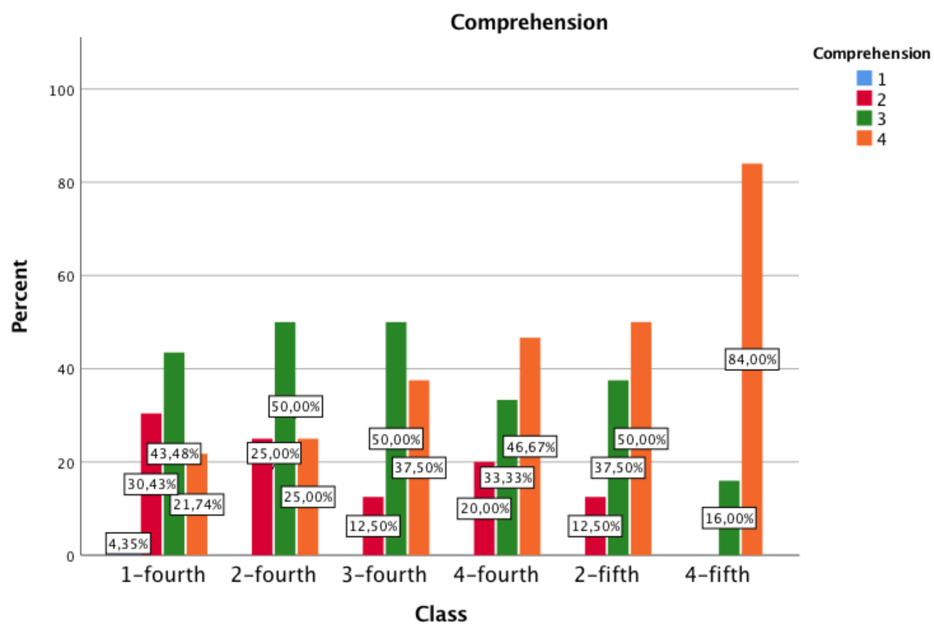


When contrasting learners' responses per class (see Figure 68), the percentages suggest that 1-fourth achieved lower levels of comprehension during the viewing process, whereas the last two groups (i.e. 2-fifth and 4-fifth) reported better outcomes (M= 3.38, SD= .711; M= 3.84, SD= .374, respectively). A Kruskal-Wallis test revealed that the groups significantly differed in terms of comprehension (H (5)= 24.974, $p < .01$, $\eta^2 = .17$). Specifically, the Bonferroni pairwise comparisons indicated that the differences were statistically significant between 4-fifth (M= 3.84, *Mdn*= 4, SD= .374) and 1-fourth (M= 2.83, *Mdn*= 3, SD= .834, $p < .001$), and between 4-fifth and 2-fourth (M= 3, *Mdn*= 3, SD= .730, p= .004) (see Table 169). The examination of the distribution of learners' responses indicates that 3-fourth and 4-fourth differed from the groups that watched fewer episodes a week (i.e.

1-fourth and 2-fourth) as regards the number of participants that selected the highest rating, and the ones that reported lower levels of comprehension (1-2) (see Figure 68). A similar picture was observed in year 5. On the whole, the results suggest that watching more episodes a week enhanced learners' comprehension (see Table 169).

Figure 68.

Students' overall comprehension of the episodes per class



Learners' overall comprehension ratings were also examined per activity type (meaning-focused vs. comprehension-focused activity). The descriptive statistics in Table 170 suggests that the participants' levels of comprehension changed as a function of the type of activity they had to complete after watching each episode. A Mann-Whitney U Test corroborated that the students that completed meaning-focused activities reported higher levels of comprehension in comparison with the ones that did construction-focused activities (Mdn= 4 vs. Mdn= 3, respectively; U= 1251, z= -2.998, p = .003, r = .25).

Table 170.

Viewing comprehension and learners' capacity to follow captions by class, year level and activity type.

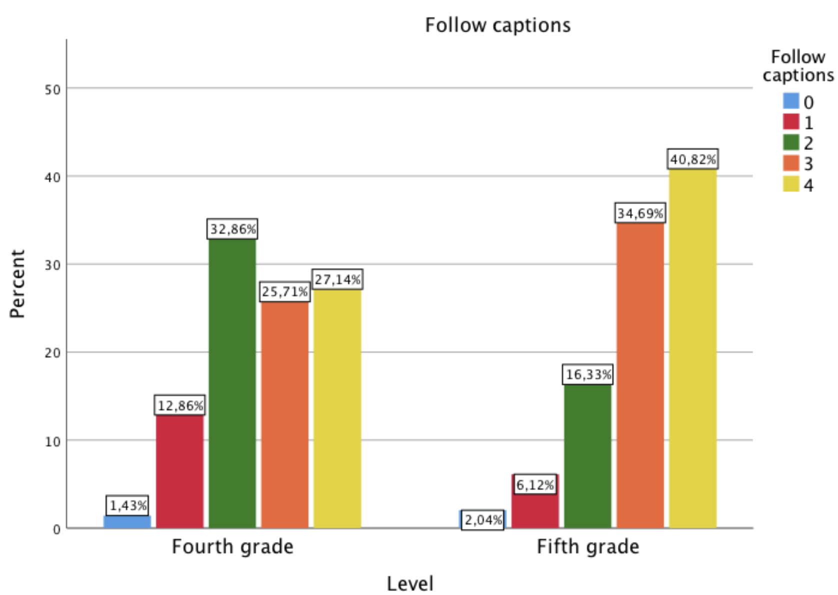
		Comprehension		Follow captions		
		Mean	(SD)	Mean	(SD)	N
Level	Fourth grade	3.06	(.78)	2.64	(1.06)	71
	Fifth grade	3.61	(.61)	3.06	(1.01)	49
Class	1-fourth	2.83	(.83)	2.39	(1.20)	24
	2-fourth	3.00	(.73)	2.75	(.93)	16
	3-fourth	3.25	(.68)	2.56	(1.03)	16
	4-fourth	3.27	(.80)	3.00	(1.00)	15
	2-fifth	3.37	(.71)	2.96	(1.00)	24
	4-fifth	3.84	(.37)	3.16	(1.03)	25
Type of activity	Meaning	3.47	(.75)	2.97	(1.08)	60
	Form	3.10	(.73)	2.67	(1.02)	60

8.2 Reading captions

The second question asked the students to self-report their capacity to read (follow) the captions in a Likert scale from 0 (low) to 4 (high) (see Figure 69). The percentages suggest that fourth-graders needed to make a greater effort to follow the captions. This was confirmed by a Mann-Whitney test, which showed that fifth graders significantly differed from fourth-graders' ratings (M= 3.06, SD= 1.008 vs. M= 2.64, SD= 1.064, respectively; U= 2113, z= 2.240, $p < .025$, $r = .19$).

Figure 69.

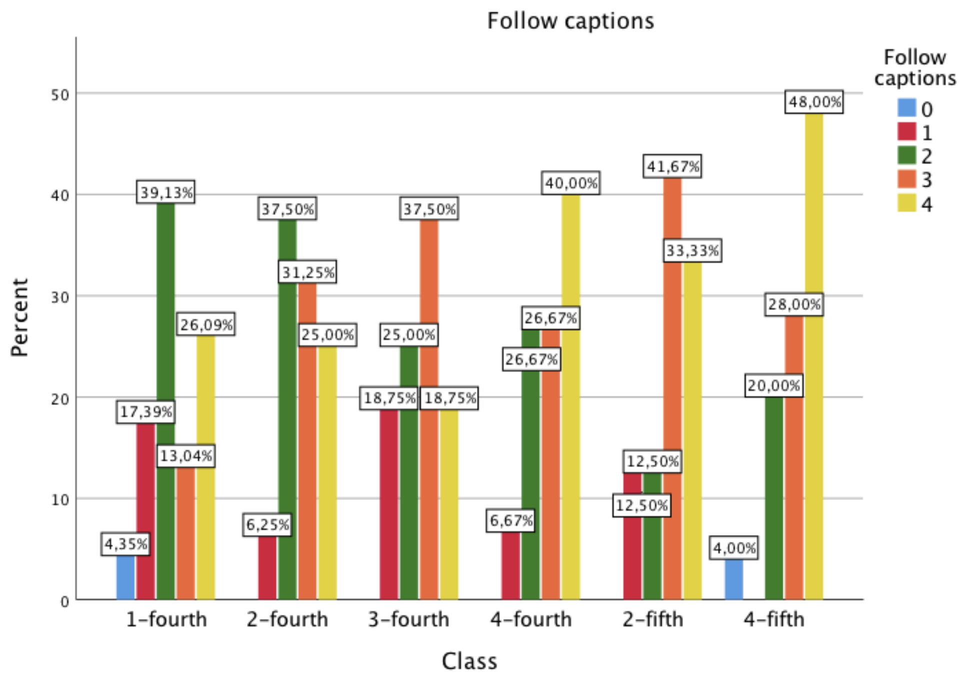
Learners' capacity to follow the captions per year level



As shown in Figure 70, most of the participants in 1-fourth and 2-fourth selected number 2, whereas the learners from the rest of the groups reported a higher capacity to follow the captions (level 3 in 3-fourth and 2-fifth and level 4 in four-fourth and four-fifth). Yet, a Kruskal-Wallis test indicated that the differences between groups did not reach statistical significance (H (5)= 8.421, p= .135, η^2 = .03) (see Table 170).

Figure 70.

Learners' capacity to follow the captions per class



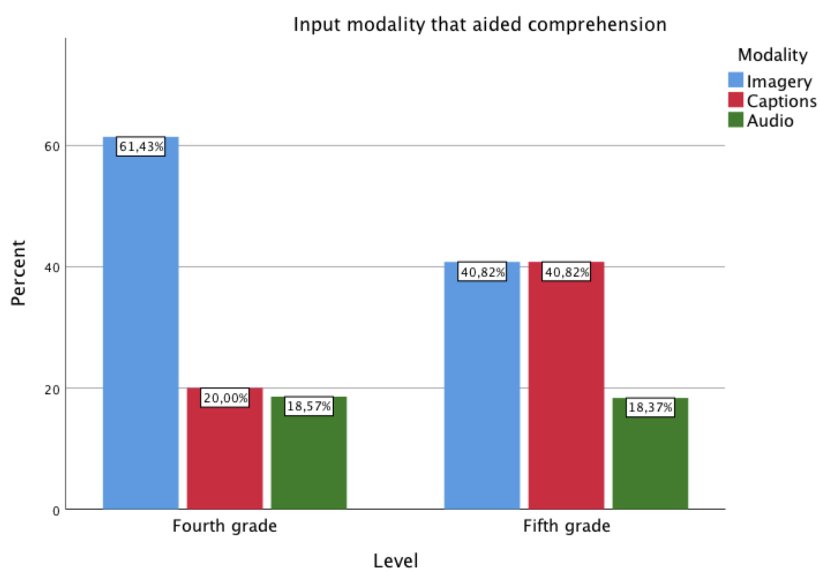
Then, the activity type groups were compared in relation to their capacity to follow the captions. A Mann-Whitney U Test indicated that the students that completed meaning-focused activities reported higher ratings than the ones that did construction-focused activities; however, the comparison was only found to be marginally significant (*Mdn*= 4 vs. *Mdn*= 3, respectively; U= 1446, z= -1.795, p = .073, r = .15) (see Table 170).

8.3 Input modality

Question 3 asked the students to select the input modality that most facilitated comprehension. The distribution of learners' choices shown in Figure 71 suggests that almost two-thirds of year-4 students relied on imagery to comprehend the episodes (61,43%). In contrast, an equal proportion of fifth graders reported to have used either captions (40,82%) or imagery (40,82%) to improve comprehension. Accordingly, a higher number of fifth graders appeared to have taken advantage of captions in comparison with the younger group (40,82% vs. 20%, respectively). With respect to aural input, a low number of participants in both year levels reported to have relied on this input mode to support comprehension (around 18%). Two Independent-samples Mann-Whitney U tests were run to compare both groups as regards the use of imagery and captions. The first analysis revealed that fourth graders relied on imagery significantly more than fifth graders (M= .61, SD= .49 vs. M= .40, SD= .40, respectively; U= 1362.5, z= -2.208, p =.027, r = .17). Conversely, the second analysis confirmed that fifth graders relied significantly more on captions than fourth graders to comprehend the episodes (M= .4082, SD= .496 vs. M= .2000, SD= .402, respectively; U= 2072, z= -2.463, p =.014, r = .17).

Figure 71.

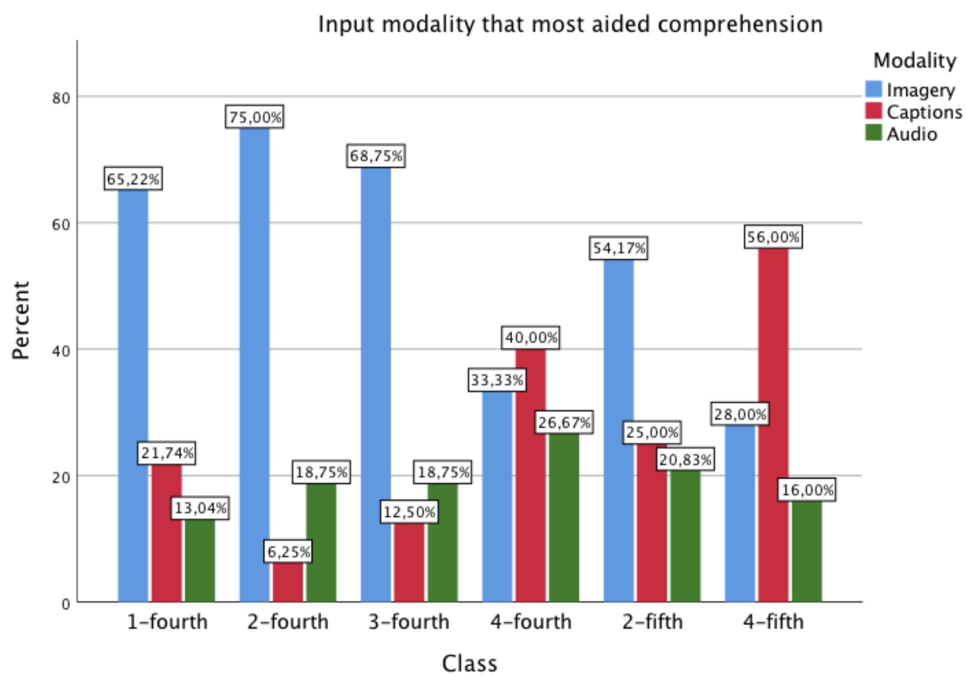
Main input modality that aided comprehension per year level



When comparing students' choices by class, overall, the tendencies seem to be similar to the ones depicted in Figure 71. Nevertheless, as shown in Figure 72, 4-fourth and 4-fifth groups differed from the rest of the conditions in that the number of participants who reported to have relied on captions to improve comprehension was higher than the ones who made use of imagery. A Kruskal-Wallis test was run for each modality so as to assess the differences between groups. The first analysis indicated that the groups significantly differed as regards the use of imagery (*H* (5)= 14.569, *p*= .012, η^2 = .085); however, the Bonferroni pairwise comparisons showed that differences were only significant between 2-fourth and 4-fifth groups (*p*= .05). Likewise, the second analysis revealed that differences between groups with respect to the use of captions to boost comprehension was also significant (*H* (5)= 16,642, *p*= .005, η^2 = 0,103), specifically between 2-fourth and 4-fifth (*p*= .009), and between 3-fourth and 4-fifth (*p*= .041).

Figure 72.

Main input modality that aided comprehension per class



Turning now to question 4, students were asked to select the input modality they paid most attention to while watching the episodes. It is important to mention that the fourth option provided in this question (I did not pay attention because I could not understand the videos and I got tired very easily) was not selected by any participant, therefore, its outcome is not reported in this section. Even though the results displayed in Figures 73 and 74 are similar to the ones obtained in question 3, differences between groups (year levels and classes) and modalities (imagery, captions and audio) appear to be less pronounced, especially in the case of the 2-fourth group. When comparing both year levels per modality, differences were non-significant for imagery (U= 1484 , z= -1.441, p = .150, r = .114), captions (U= 1830.5, z= .767, p = .443, r = .05), nor audio (U= 1830.5, z= .927, p = .354, r = .05). Likewise, differences between conditions (classes) did not reach statistical significance for imagery (H (5)= 10.827, p= .055, η^2 = .052), captions (H (5)= 4.927 p= .425 , η^2 = .001), or audio (H (5)= 8.573, p= .127, η^2 = .032). In addition, the data was assessed to determine whether students' attention to a specific input mode varied as a function of the type of after-viewing activity they had to complete (see Figure 75). The percentages suggest that both groups did not differ as regards the input modality they paid most attention to, which was confirmed through the performance of a Mann-Whitney U test for imagery (U= 1755.5, z= -.089, p = .929, r < .001), captions (U= 1730.5, z= -.258, p= .796, r < .001) and audio (U= 1824, z= .427, p= .669, r = .001).

Figure 73.

Input modality each year level mainly focused on

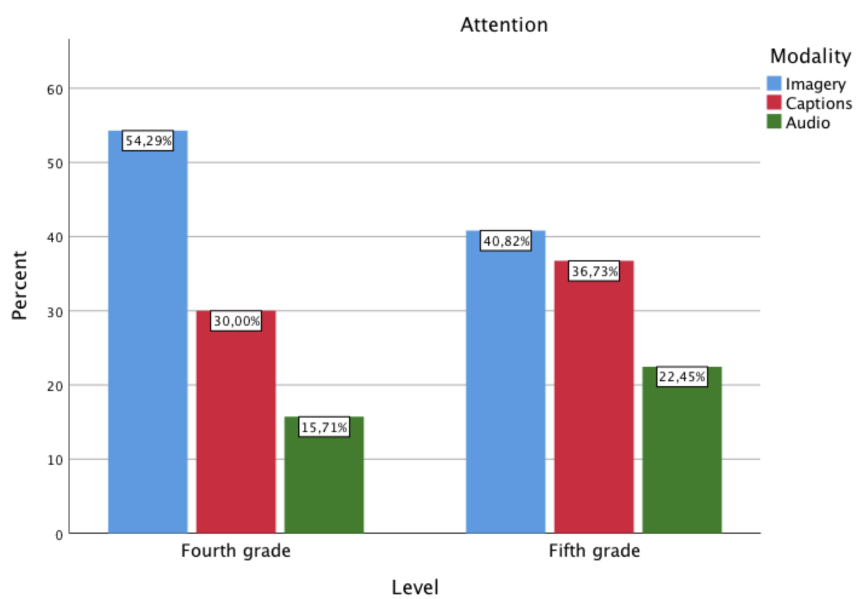


Figure 74.

Input modality each class mainly focused on

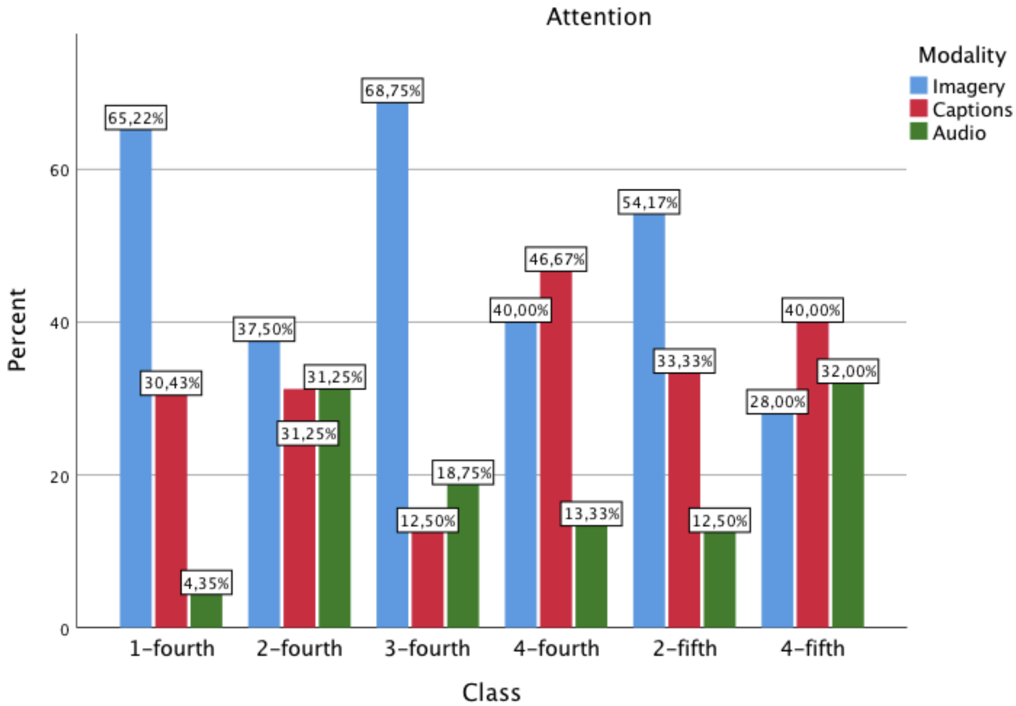
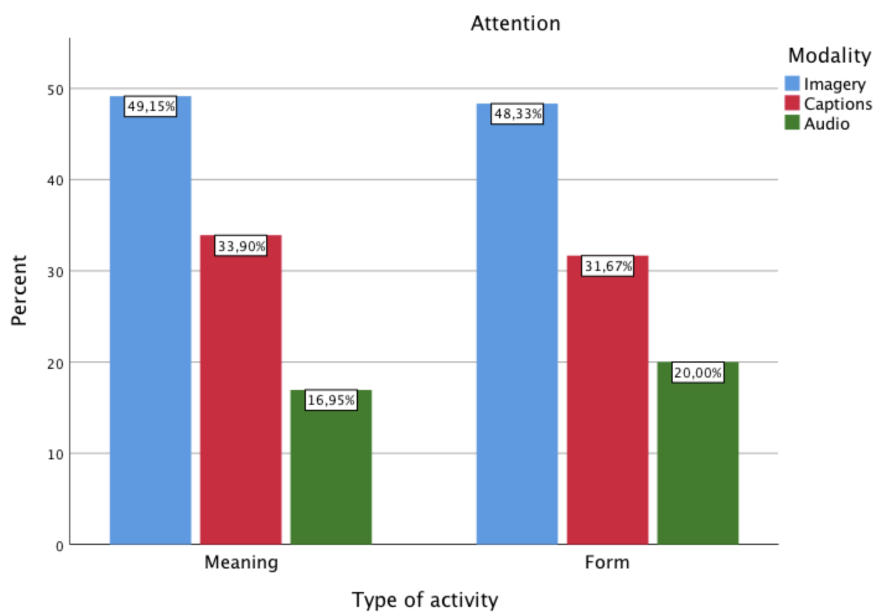


Figure 75.

Input modality each activity group mainly focused on



8.4 Enjoyment

Question 6 asked the students to report the extent to which they liked watching the episodes in a Likert scale that ranged from 0 (did not like it) to 4 (liked it a lot). Figure 76 suggests that both year levels showed high levels of enjoyment. However, a Mann-Whitney U test revealed that fifth graders enjoyed the episodes significantly more than fourth graders (M= *3.92*, SD= .344 vs. M= 3.73, SD= .612, respectively; U= 1931, z= 1.969, p =.049, r = .10). With respect to the treatment conditions, it can be seen from Figure 77 that the scores are similar in the groups that watched between two and four episodes a week, whereas the level of enjoyment appeared to be lower in 1-fourth. When subjecting the data to a Kruskal-Wallis test, the results demonstrated that differences between groups were statistically significant (H (5)= 13.174 p= .022 , η^2 = .072), and the Bonferroni pairwise comparisons indicated that only the 2-fifth group significantly differed from 1-fourth (p= .028).

Figure 76.

How much the participants liked watching Charlie and Lola by level.

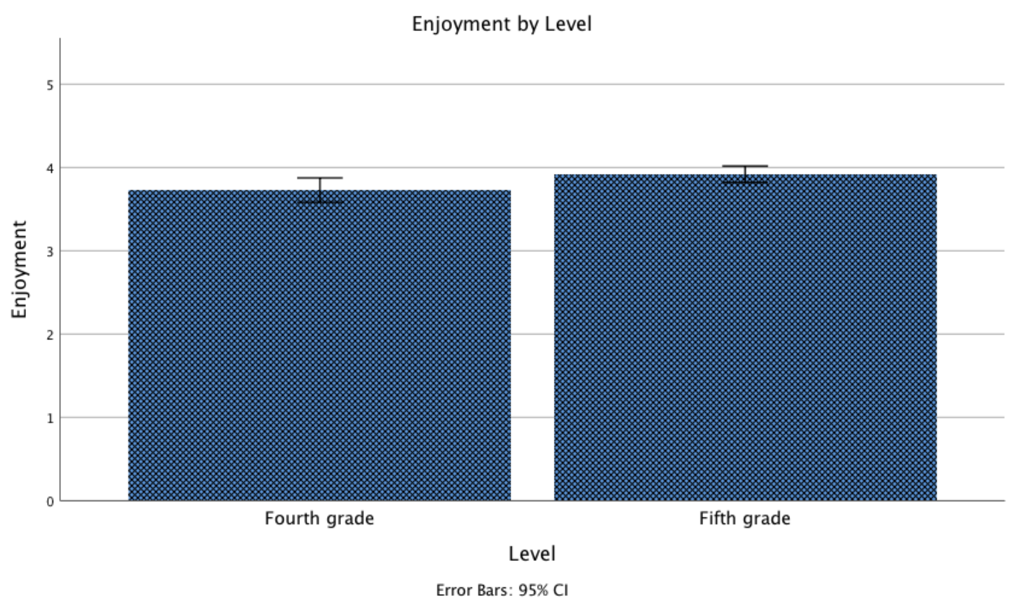
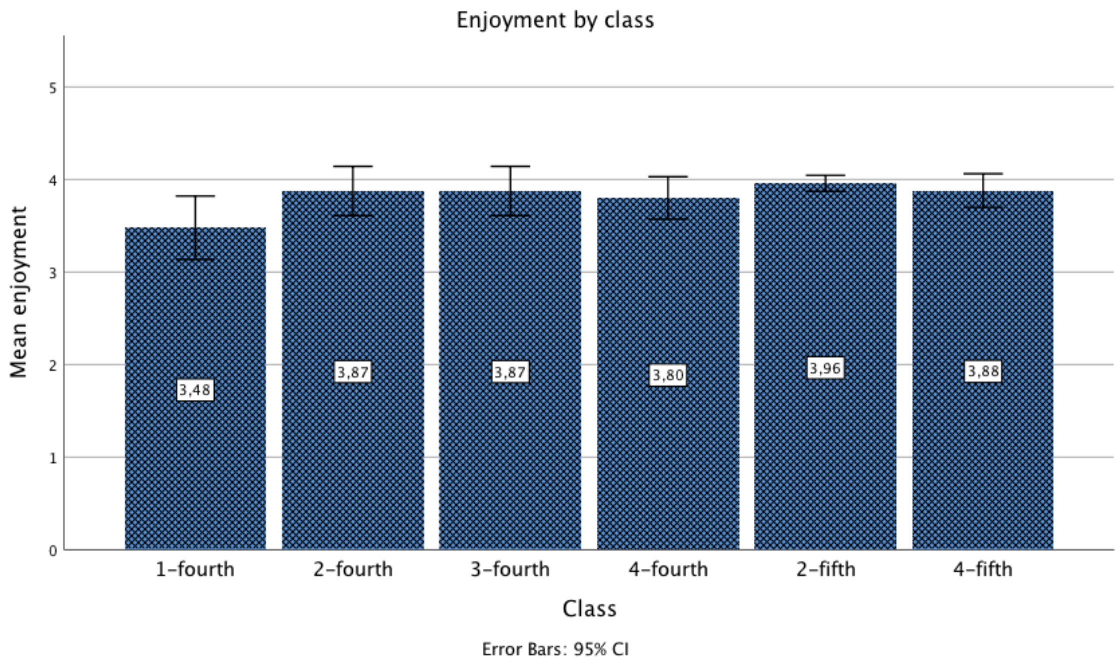


Figure 77.

How much the participants liked watching Charlie and Lola by class.



Given that viewing was not a regular activity in the English class, question 8 asked the students to report whether they would like to watch more videos in the future or not. To this aim, the participants had to select a level from a Likert scale that ranged from 0 (no) to 4 (a lot). As observed in Figure 78, even when both year levels showed to be keen on this idea, fifth graders seemed to be more eager to watch videos in the future. A Mann-Whitney U test confirmed that the difference between fourth and fifth graders was statistically significant (M=3.44, SD=.845 vs. M=3.78, SD=.550, respectively; U= 2095, z= 2.564, p =.010, r = .18). However, it is important to point out the lowest scores were provided by 1-fourth and 4-fourth (see Figure 79). When subjecting the data to a Kruskal-Wallis test, the results revealed that the differences between groups were statistically significant (H (5)= 14.969 p= .010 , η^2 = .088). Specifically, the Bonferroni pairwise tests indicated that 4-fourth significantly differed from 4-fifth (p= .019), and 2-fifth (p= .018).

Figure 78.

Learners' willingness to watch more videos in the future by year level

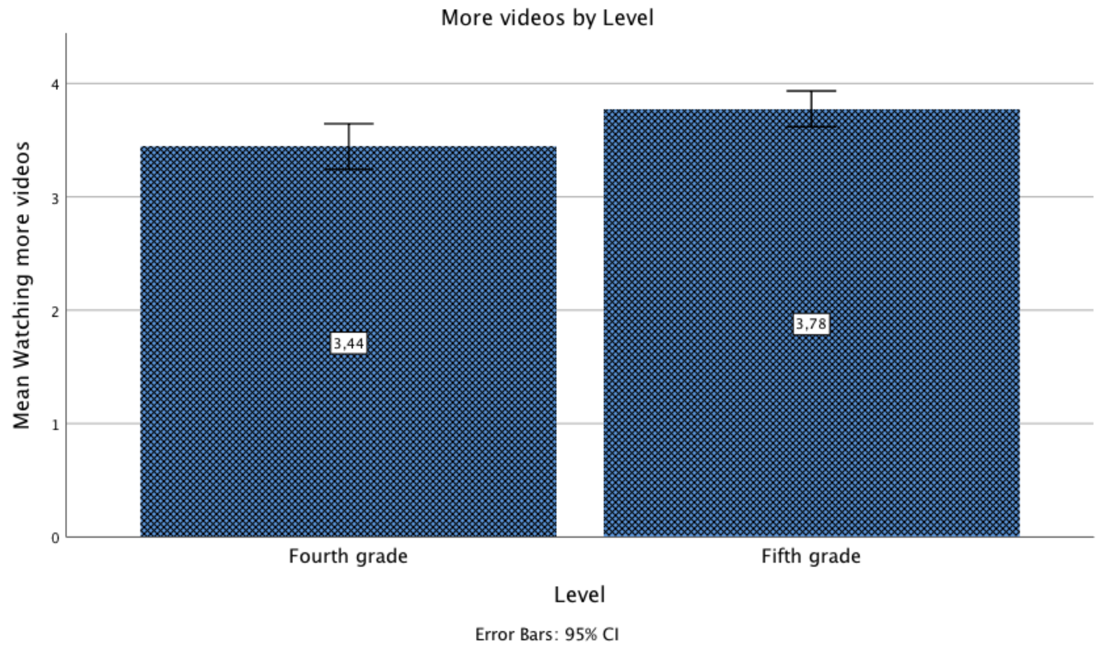
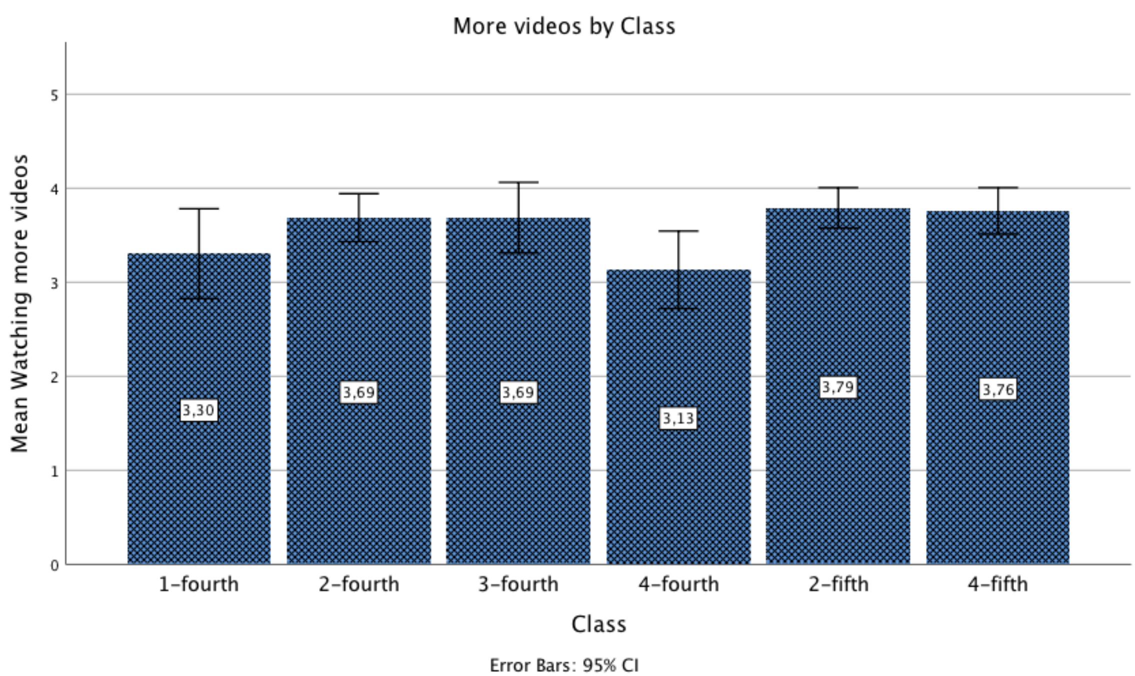


Figure 79.

Learners' willingness to watch more videos in the future by class



8.5 Learning English from Charlie and Lola

Question 5 asked the students to estimate the extent to which students might learn from Charlie and Lola on a Likert scale from 0 to 4. As depicted in Figure 80, both groups reported a high mean score, which suggests that they believed that it is possible to learn English from the animated cartoon. Despite the fact that fifth graders' mean score is slightly higher, the difference between both groups did not reach statistical significance (U= 1982, z= 1.587, p =.113, r = .13). Similarly, the comparisons between treatment conditions (classes) were non-significant either (H (5)= 8.143 p= .149 , η^2 = .028).

As regards their own English learning from viewing Charlie and Lola (question 7), the students were asked to indicate one of the four alternatives (4= Yes, a lot; 3= Yes, a little; 2= Not much; 1= Nothing). The results showed that the participants from both year levels believed that, overall, they learned from the treatment. In addition, a Mann-Whitney U test confirmed that there was no significant difference between fourth and fifth graders (M=3.56, SD=.605 vs. M=3.57, SD=.677, respectively; U= 1773, z= .369, p =.712, r = .031). Similarly, as shown in Figure 81, most responses in each treatment condition ranged between 3 and 4, except for 1-fourth group, where fewer students selected the highest level (4= a lot). However, when subjecting the data to a Kruskal-Wallis test to make group comparisons, the outcome only approached statistical significance (H (5)= 10.656 p= .059 , η^2 = .05).

Figure 80.

Possibility of learning from Charlie and Lola by level

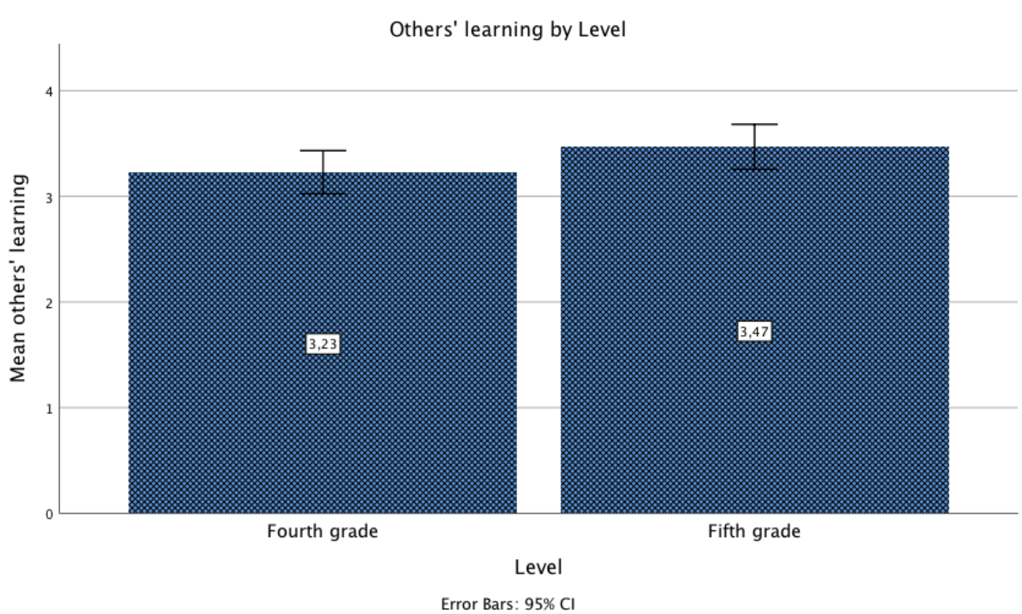
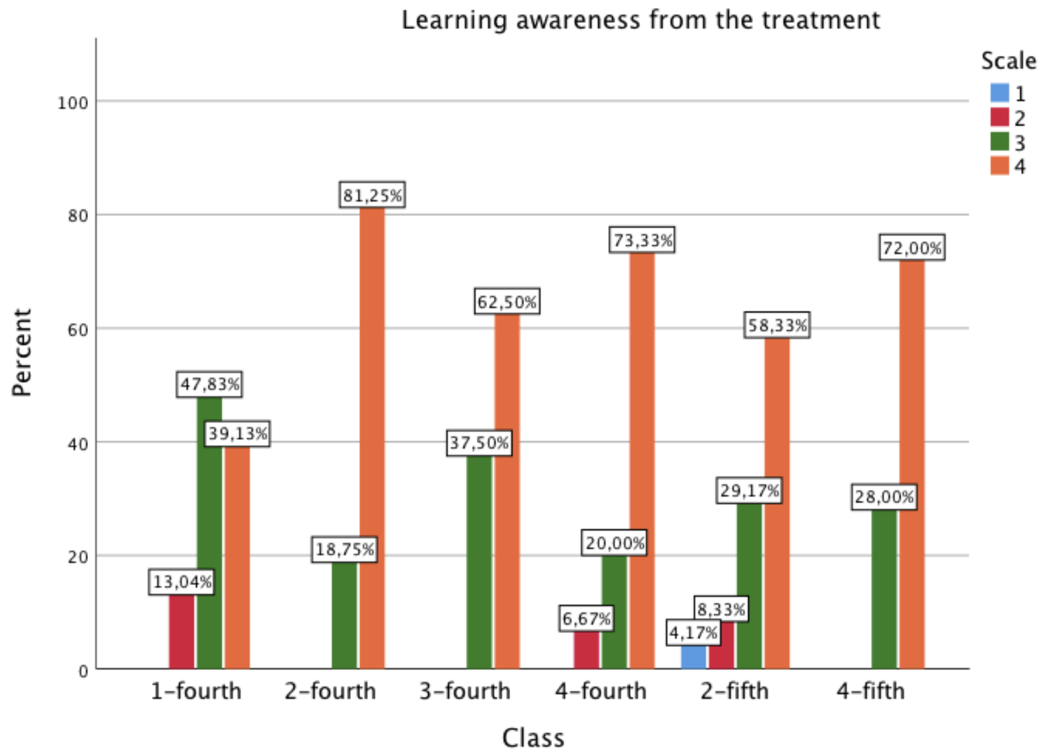


Figure 81.

Learning awareness from the treatment by class



The participants who selected the highest three options in question 7 (4= Yes, a lot; 3= Yes, a little; 2= Not much) were asked to indicate all the language aspects that might have improved as a result of the treatment. Table 171 displays the percentage of students that selected each alternative by contrasting their responses as function of the type of after-viewing activity they had to complete and the year level they belonged to. Undoubtedly, grammar was the least selected option, while the highest difference between groups (year level) was found in pronunciation. In fact, the results of a Mann-Whitney U test confirmed that a significantly higher number of fifth graders reported to have learned pronunciation (U= 2173, z= 2.859, p =.004, r = .22). As regards activity type, the participants that completed construction focused activities did not show higher learning awareness from captioned-video viewing, except for the learning of phrases.

Table 171.

Learners' perspective concerning their L2 gains from the treatment.

Options	Type of activity		Year level	
	Meaning-focused	Construction-focused	Year 4	Year 5
Learning vocabulary	61%	56,7%	61,4%	55,1%
Learning pronunciation	50,8%	48,3%	38,6%	65,3%
Learning grammar	11,9%	3,3%	7,1%	8,2%
Learning phrases	47,5%	55%	57,1%	42,9%
Improved comprehension of videos in English.	62,7%	51,7%	57,1%	57,1%
Learning how words are spelled in English.	52,5%	51,7%	54,3%	49%
Learning to read captions	47,5%	51,7%	50%	49%

8.6 Learners' perceptions of the treatment: Discussion

8.6.1 Comprehension and reading of captions

This section explored learners' perception of the treatment with the aim of further interpreting the findings obtained in previous sections as regards vocabulary learning and receptive skills development, as well as the influence of treatment and learner-related factors on the outcomes. In relation to comprehension, the results revealed that fifth graders self-reported better comprehension than fourth graders, which is expected considering their significantly higher proficiency level in both languages. By the same token, fifth graders were better equipped to cope with the speed of captions, which was clearly confirmed by their superior self-reported capacity to follow the onscreen text. As a result, fifth graders' higher proficiency level may have allowed them to process the input with greater ease in order to leave some attentional resources available to notice unknown target language constructions (Kim & Webb, 2022a; Kormos, 2017; Lin & Siyanova-Chanturia, 2015; Montero Perez, 2020). This may explain why they scored consistently higher in written-word form recall and written-word form and meaning recognition over time. Likewise, throughout the process, their lower cognitive effort might have allowed them to attain a better integration of audio, text and imagery to keep their advantage in receptive language skills (Mayer, 2014, 2022; Pellicer-Sánchez, 2022). It is not to say that the viewing experience was free of challenges for fifth graders, however, they seemed to have greater resources to face them. As the literature suggests, extensive viewing practice is needed to improve learners' processing of audio, text and imagery over time (Bravo, 2008; Pujadas, 2019; Vanderplank, 1988), this is why fourth graders may have required a higher number of episodes to attain higher levels of comprehension and feeling more in control of the reading of onscreen text.

With respect to the comparisons between treatment conditions, the findings suggested that in fourth grade, watching fewer episodes a week resulted in lower levels of comprehension and more difficulty to follow the captions. This outcome ties in with the results of the present investigation that indicated that shorter lags between episodes moderated the influence of individual differences, such as SR efficacy and English text segmentation (Collins & White, 2012). By the same token, this result concurs with those of Greving and Richter (2021), who found that in the reading of related texts, shorter spacing was associated to higher levels of comprehension and lower perceived difficulty. This means that the younger and less proficient participants need to watch the videos with shorter spacing to make stronger connections between the episodes and benefit from narrow viewing (Rodgers & Webb, 2011). Even when 2-fifth's lower scores in comprehension might suggest that in this year level viewing distribution also played a role, its effects might have been neutralized by fifth graders' stronger skills to cope with the greater demands imposed by the implementation of longer lags between episodes. This may explain why 2-fifth and 4-fifth obtained comparable gains in vocabulary and receptive language skills. Taken together, these findings confirm than in less proficient and younger participants, captioned-video viewing should be implemented with shorter time intervals due to the high difficulty of the task (Serfaty & Serrano, 2022a, 2022b; Suzuki et al., 2019).

Regarding the reports obtained by each activity type group, the results indicated that watching the videos for dual purposes, that is comprehension and intentional L2 learning, increased the cognitive load and hindered comprehension. Similarly, the completion of construction-focused activities affected learners' capacity to follow the captions with greater ease. These results reflect those of Pujadas and Muñoz (2020), which showed that the simultaneous allocation of cognitive resources on comprehension and word learning may have resulted in cognitive overload (Mayer, 2014, 2022; Mayer & Fiorella, 2022).

8.6.2 Input modality

On the question about the input mode that facilitated comprehension, the results showed evidence of fourth graders' high reliance on imagery, corroborating that, in lower proficiency learners, imagery works as a compensatory mechanism that fills knowledge gaps (Peters & Muñoz, 2020) and facilitates comprehension (Clark & Paivio, 1991; Durbahn et al., 2020, 2022; Mayer, 2014, 2022; Mayer & Fiorella, 2022; Peters, 2019; Rodgers, 2020). Although a high percentage of fifth graders also reported to have relied on imagery (40%), an equal number of fifth graders was shown to rely on captions to improve comprehension. This finding is congruent with the literature that suggests that the use of captions facilitates speech segmentation and word recognition, making the aural input more comprehensible (Charles & Trenkic, 2015; Kormos et al., 2019; Teng, 2019a, 2019b; Toscano-Fuentes, & Julián de Vega, 2018). Likewise, the results corroborated that L2 listening may be a challenging task for lower proficiency learners (Kormos et al., 2019; Newton & Nation, 2021), this is why the presence of bimodal verbal input and imagery may encourage learners to stay on task and support the development of receptive language skills over time (Bird & Williams, 2002; Birulés-Muntané and Soto-Faraco, 2016; Charles & Trenkic, 2015; Mitterer & McQueen, 2009).

The examination of learners' choice by treatment condition (i.e. class) confirmed the same tendency aforementioned. Yet, in the groups that watched more episodes a week (4-fourth and 4-fifth), a higher number of participants was shown to rely on captions. Considering that younger learners have been found to struggle to cope with the speed of captions due to their still-developing reading skills (Muñoz, 2017; Vanderplank, 2016), this finding might point to the facilitating effects of shorter lags (Serfaty & Serrano, 2022a, 2022b; Suzuki et al., 2019).

On the question about the input mode the participants paid most attention to, the results were similar to the ones reported in the previous question. Yet, it is important to acknowledge that this question was harder to answer for some of the participants, who orally expounded that they were either unaware of this issue or devoted similar levels of attention to audio, text and imagery. Thus, a more accurate answer to this question may be obtained through the study of learners' eye movements (captions vs. imagery). The investigation by Tragant and Pellicer-Sánchez (2019) with fifth graders shed some light on this regard. While watching an episode of Charlie and Lola, the students were found to spend longer time on the reading of captions than on the processing of images, which seemed to be successfully integrated to achieve comprehension. Yet, these results may not necessarily be extrapolated to fourth graders due their differences in proficiency and the great changes experienced along middle childhood (Delgiudice, 2018; Harris & Westermann, 2015; Holmes & Myles, 2019). Yet, it is worth mentioning that the results obtained by Black (2020) with 8-9 year-olds indicated that, on the whole, learners spent a greater amount of time on L1 subtitles but they fixated longer on images, implying that the processing of non-verbal input is also an important part of the viewing process in young learners.

With respect to activity type, similar results were obtained for each group, suggesting that learners paid greater attention to imagery, followed by captions and audio (in this order). This result may explain why activity type was not found to have a significant effect on the development of receptive language skills. Although the implementation of construction-focused activities might have increased the cognitive load, it could be the case that learners showed similar processing patterns along the viewing experience. Based on Montero Perez et al.'s (2018) findings on test announcement, the participants that completed construction-focused activities might have prioritized comprehension. More precisely, the presence of two common comprehension questions in the after-viewing activities might have neutralized any potential difference between the conditions.

8.6.3 Enjoyment

Learners' responses reported high levels of enjoyment in both year levels, corroborating the findings of previous investigations with school learners (Black, 2022; Bravo, 2008; Marzá & Torralba, 2015; Pujadas, 2019; Zabalbeascoa et al., 2015). However, the small difference between groups reached statistical significance, indicating that the viewing experience was significantly more enjoyable for fifth graders. In addition, the examination of each treatment condition showed that the lowest rating was provided by 1-fourth. Next, when the participants were asked whether they would like to continue watching videos in the future, both groups were shown to be keen on this idea. Still, fifth graders' ratings were found to be significantly higher. Specifically, the comparisons between classes indicated that the lowest scores were assigned by 4-fourth and 1-fourth. On the whole, it is reasonable to conclude that in general, the participants showed high levels of enjoyment and were willing to watch more videos in the future. However, there are different possible explanations for the differences between year levels and within fourth graders.

To start with, the literature suggests that throughout middle childhood, there is an increase in learners' awareness of their learning process and their own limitations (Muñoz, 2017b; Myles, 2022). Furthermore, children become more goal oriented and develop their capacity to use a wider range of strategies to overcome the challenges encountered in their learning process (Myles, 2022). Thus, although primary school learners do seem to enjoy the implementation of fun activities, their sense of progress and actual learning achievements appear to be key aspects of their motivation (Muñoz, 2017b, 2017c; Myles, 2022). With this in mind, it is possible to assume that in line with prior studies with young learners (Marzá and Torralba, 2015), fourth graders' lower levels of enjoyment and enthusiasm for extending the viewing experience may be attributed to their significantly lower proficiency level in English and less developed L1 reading skills, which seemed to hinder comprehension and their capacity to follow captions. Along with the challenges associated to the processing of audiovisual input (Black, 2022; Muñoz, 2017a; Teng, 2019b; Vanderplank, 2016), it is also likely that some of the methodological decisions made in this investigation aiming at the quality of the research design disregarded some fundamental learners' needs. More precisely, the lack of feedback, instruction and scaffolding along the viewing experience might have negatively affected the participants' viewing self-efficacy, particularly in the case of fourth graders, who needed greater support to develop more effective viewing strategies. As Graham (2022) suggests, learners will persist as long as they feel in control of the factors and strategies that are required to face the difficulties of a task (p.188).

Alternatively, fourth graders' lower levels of enjoyment and willingness to watch more episodes might also be attributed to their lack of awareness of the extent to which captioned-video viewing might support L2 learning. Some of the responses obtained from fourth graders after the piloting of a sample episode of Charlie and Lola showed evidence of learners' concern about the use of videos in class, since it was seen as a leisure activity that lacked clear L2 learning purposes. This might have been the case of 4-fourth, whose English classes were mainly devoted to the viewing tasks for three consecutive weeks. Thus, even when this group in the following questions reported that the viewing experience was actually conducive to learning, this activity might not have been considered as effective as their regular English sessions. As a result, the use of after-viewing activities may require more explicit explanations concerning the role of viewing inside the L2 classroom. The studies by Black (2022) and Zabalbeascoa et al. (2015) introduce some practical ideas on this regard.

8.6.4 Learning from Charlie and Lola

On the question about the possibility of learning English from Charlie and Lola, both groups reported a high mean score, which was congruent with the positive findings obtained as regards vocabulary learning and the development of receptive language skills. Although fifth graders' average rating was found to be slightly higher, the group comparisons did not reach statistical significance. Furthermore, similar results were obtained when the participants were asked to rate their own learning experience. Still, the examination of the options selected in each group suggested that 1-fourth showed lower sense of achievement. This result was unexpected considering that in comparison with the rest of the groups in year 4, they obtained the highest scores in general vocabulary knowledge and text segmentation in both languages. In other words, they seemed to be better prepared to deal with the viewing process and benefit from it. Once again, this observation might support the hypothesis that under long-spaced conditions, learners may struggle to connect the content from related texts (videos in this case), which is a factor that may increase the amount of effort required to process the input (Greving and Richter, 2021). Thus, the higher perceived difficulty of the task might have resulted in learners' lower sense of achievement.

As for the specific language aspects that benefitted from the treatment, grammar was the least selected option. Previous studies have also shown that the learning of language aspects other than grammar, such as vocabulary, multi-word units, spelling, comprehension, and, to a lesser extent, pronunciation are the most common gains addressed by language learner-viewers (Pattemore et al., 2020; Pujadas, 2019; Vanderplank, 1988). Still, learners' reports may not necessarily be congruent with their actual performance and learning gains (Pattemore & Muñoz, 2020; Pattemore et al., 2020; Pujadas, 2019). While in the investigation by Pattemore et al. (2020) the participants exposed to input enhancement reported higher learning gains in vocabulary, the findings of the present investigation did not find clear differences between activity type groups. On the contrary, the students subjected to construction-focused activities only reported slightly higher gains in the learning of phrases, which is not surprising seeing that in most of the word-focused questions the target words were inserted in phrases. The fact that the completion of construction-focused activities did not lead to higher perceived gains in vocabulary may be associated to their self-reported lower levels of comprehension and ability to follow captions. In other words, the higher cognitive load involved in the completion of construction-focused activities affected learners' perceptions of their actual gains and did not reflect the beneficial effects of this activity type on the recall of written-word forms, and to a lesser extent on the learning of form-meaning links.

Another interesting finding was that fifth graders reported higher gains in terms of pronunciation, which may be associated to their significantly higher proficiency level. As it will be expounded in the following section, when the participants talked about pronunciation, they mainly referred to their awareness of the differences between English and Spanish in relation to language transparency. However, even when both year levels mentioned that this is one of the key sources of difficulty in learning English, the literature suggests that the learning of phoneme and grapheme correspondences takes time and requires great amounts of exposure to the target language to show significant progress in this regard (Muñoz, 2017b). Therefore, fifth graders' greater knowledge of the target language might have increased their perceived gains in pronunciation, which is somehow related to their actual gains in written-word form recall. Yet, considering the relatively low gains obtained in this vocabulary dimension, the integration of viewing and more explicit methodologies (e.g. phonics) may be more effective at boosting learners' outcomes (Marian et al., 2021; Muñoz, 2017b; Pérez Cañado, 2006; Porter, 2020).

IX. Interview on learners' perception of the viewing experience

As explained in the methodology section, three groups of participants (n=18; one of fourth graders and two of fifth graders) were interviewed to gain deeper comprehension of students' perceptions of their viewing experience. The data[17] was elicited in groups of six students through the administration of the same questionnaire reported in section 8. These questions were only used as a starting point, given that the participants provided richer information about their viewing process. Therefore, the data was not coded nor searched for patterns based on those questions. Instead, the patterns were identified across the whole interviews to dig deeper and identify participants' intentions and interpret the meaning behind their words (Braun & Clarke, 2013, 2022). Specifically, the analyses and results reported in this section respond to research question 5: What are students' perceptions of their viewing experience? How do their answers contribute to the interpretation of the quantitative findings?

The theme development was consistently refined to find the best fit of analysis to answer the research questions. This process was supported by the design of a mind map, where the themes were actively generated by using different colors and levels (see Figure 82; Braun & Clarke, 2022). The number of students that participated in the interviews was limited, so the analyses primarily focused on meaning rather than frequency to respond to the main inquiries. Yet, the running of queries and the generation of word clouds also provided useful information to determine whether the topics that were mainly addressed by the students were part of the themes that emerged. The data was also collated by year level and learners' vocabulary knowledge, which was used as a measure of proficiency.

The resulting overarching themes were as follows:

-Learners' attitude towards the viewing experience.

-L2 gains.

-Main challenges encountered during the viewing experience.

-Learners' strategies and processing of audiovisual input.

-The role of after-viewing activities.

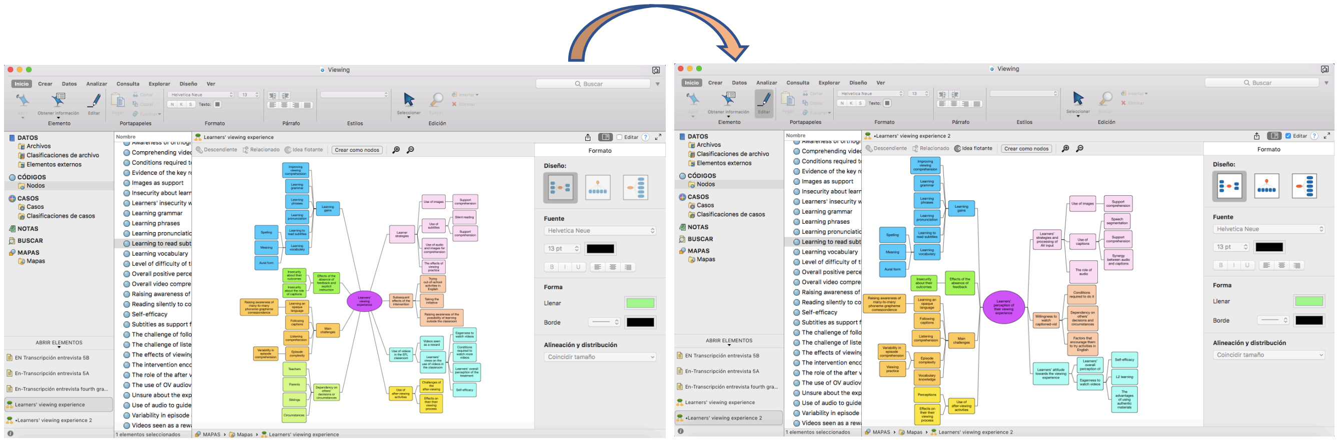
-Lack of feedback.

-Learners' willingness to watch captioned videos in the future.

[17] Interview transcriptions: https://drive.google.com/drive/folders/1t9FsWsK0NnvoyDqVkgn7k287jOMN2LP1?usp=sharing

Figure 82.

Qualitative analyses: Theme development



9.1 Results

The main conventions used in the transcriptions and the analyses reported in this section are displayed in Table 172.

Table 172.

Transcription conventions

	Description
[ ]	Further explanations on students' statements. Description of some of the actions that took place throughout the interview.
CAPITAL LETTERS	Emphasis on words or ideas while speaking.
[…]	Pauses and hesitations.

9.1.1 Learners' attitude towards the viewing experience

Learners' attitude towards the use of captioned videos was, without a doubt, overwhelmingly positive. In line with the results obtained through the questionnaire, learners' responses concerning the extent to which they enjoyed the viewing experience concentrated on the highest score on the Likert scale that ranged from 0 to 4. Apart from reporting their ratings, some of the participants explicitly employed words such as 'liked', 'fun' and 'entertaining' to refer to the viewing experience, which was also connected to the fact that watching videos may additionally lead to learning gains. As CV said "…Charlie and Lola was fun because you learn at the same time." When sharing their views about the intervention, some of the participants also reflected on the value of using authentic materials in the EFL classroom to learn through ‘real’ interaction and the language used by the protagonists. For instance, JC commented “You have to think that this TV program was not made to learn English, it’s a normal program, so you can learn more words…” while CV seemed to focus on the pronunciation used by Charlie and Lola, which was found to be a salient aspect of the treatment, “sometimes they said ‘this is tall and thin’ ‘it’s my pet’, that sounds better in the video…I don’t know how to explain it. It’s the English accent…”.

Overall, learners’ attitude towards the use of captioned videos and their willingness to do this activity was connected to their viewing self-efficacy. Learners’ responses reflected their pride and sense of achievement when comprehending the main ideas and identifying some of the words. This was exemplified in fifth graders’ comments when explaining why they paid attention to the audio: “There were words that I knew and I could understand them” (JP); “The audio…because I could understand many things” (RA); “The audio because you can understand many words when you are in fifth grade” (NM). In spite of the multiple challenges faced throughout their viewing experience, the characteristics of the animated cartoon selected for the purpose of this study seemed to encourage the learners to try their own strategies to compensate for their low proficiency level so as to stay on task. The answers provided by RT and RB illustrate learners’ capacity to take advantage of multimodality to fill knowledge gaps and improve comprehension: “…I looked at the images…because I understand better. For example, if they say an action, I can see what they are doing” (RT); “So, when I didn’t understand what they said, I read the subtitles and then I could understand” (RB).

9.1.2 L2 gains

The interviewees unanimously selected vocabulary learning and pronunciation as the main L2 gains obtained from this intervention. As regards vocabulary learning, they employed different examples to demonstrate their outcomes, such as “I learned to say ‘cabbage’ for ‘repollo’” (JP) and “…I learned more words. For example, while watching Charlie and Lola I learned to say ‘drops’, which I THINK meant ‘gotas’…” (JC). Their examples also showed evidence of the different vocabulary dimensions learned throughout the process. To start with, they referred to the use of imagery to learn the meaning of unknown words or phrases. CV said “in the first episode of Charlie and Lola, they mentioned the word ‘mash’. I didn’t know the meaning but I could see it in the images, so then I said ah! mash…”, and a similar example was provided in the conversation between JJ and TE:

JJ: …once they said fish ‘something’ *[forgot the word fingers]* and I could see they were like ‘nuggets’ *[that’s the word used in Chile for fish fingers and chicken nuggets]*.

TE: Fish fingers *[laughs]*.

JJ: Yes, fish fingers *[laughs]*.

Indeed, learners’ comments demonstrated their high reliance on imagery rather than contextual clues to foster comprehension and form-meaning mapping. In addition, the data revealed that the simultaneous exposure to L2 audio and L2 captions helped the participants fill knowledge gaps and map aural and written-word representations. A recurrent comment along the interviews was the learning of written and aural word forms. This was explicitly stated by JG and RT: “I learned to write words…because I read the captions” (JG); “I learned how to pronounce and write words” (RT). This was also exemplified by MC when explaining why he focused on captions to improve comprehension: “…because they help me understand how words are spelled, and that helps me.” Also, the viewing experience allowed one of the participants to identify differences between her L2 knowledge and the language used by Charlie and Lola. As JG said “Sometimes, I didn’t understand the words, so I paid attention to the captions. For example, I say /təˈmeɪtoʊz/ and Charlie and Lola said /təˈmɑːtəʊz/.”

The learning of pronunciation was explicitly mentioned by most of the interviewees at least once. Interestingly, learners also emphasized that this was one of the most challenging aspects encountered in the viewing process, which may be associated to their limited contact with the target language prior to the intervention. As developed in the following section, the participants relied on captions to make the aural input comprehensible. Therefore, learners’ comments on pronunciation primarily focused on how the processing of bimodal input scaffolded speech segmentation and promoted comprehensibility. This was visible in NM’s comment, “I think the captions helped a lot because of the pronunciation. It was like, they mixed the words […] they mixed the words and it was difficult to understand without captions, so we could understand with the captions.” This interdependency between aural and written input also led some of the participants to consider viewing comprehension and their ability to read captions as part of their learning gains from the treatment.

With respect to grammar learning, only one participant explicitly referred to this language aspect. Although the concept of grammar was explained in a child-friendly manner during the interview and before the administration of the questionnaire, the participants did not consider grammar as part of their learning gains. Still, grammar seemed to be present when one of the participants mentioned the learning of word order, and, perhaps, when other interviewees addressed the learning of phrases through the viewing experience. From a usage-based perspective, what learners alluded as the learning of phrases may be considered as the learning of grammatical constructions. In any case, the participants did not provide further details or examples to fully understand what they specifically meant by the learning of phrases.

9.1.3 Main challenges encountered during the viewing experience

One of the recurrent topics along the interviews was associated to the challenge of learning an opaque language with many-to-many phoneme-grapheme correspondences, as well as their struggle to segment and identify words in the stream of speech. This is what JP addressed as "weird" because, in his view, "…it's like they mixed letters. They mixed a letter with another…". This issue was echoed in the three groups, where the participants provided multiple examples to illustrate their difficulty and the strategies they employed to cope with it. For instance, the students unanimously reported their reliance on captions to identify the words that were unintelligible in aural speech but part of their previous knowledge in written form. As CV explained "…sometimes they said words such as /wɔːl/ and /tɔːl/, and you read them to understand them. So, then you say: 'Ah! It's TALL!'…"; similarly, TA commented "The word 'tall' appeared on the video. I didn't understand it when they pronounced it, but the subtitles say everything. So, I know they are saying 'tall' but it isn't as if they were saying that word."

Despite their reliance on captions to enhance speech comprehension, the participants also alluded to the challenge of following the speed of captions, since sometimes, they were unable to read all the words available on the screen. As JM explained "Yes, sometimes, I missed a word, one word!". Yet, the interviewees also indicated that their difficulty to cope with the speed of captions varied as a function of phrase length and episode characteristics. Furthermore, this challenge may also have been influenced by their low vocabulary knowledge, which was reported by AG when saying "I could follow the captions but not so well…because I didn't understand some words." Based on learners' responses, we may also assume that the less proficient participants that experienced higher levels of difficulty required to make a greater cognitive effort to process captions. This is why some participants struggled when switching their attention between captions and imagery. This is what MJ replied in response to the question that focused on their capacity to follow captions: "…sometimes I got distracted, I don't know, Charlie was doing one thing and Lola was doing another." TA also mentioned "…they distract me and you HAVE TO read them. I looked at the images but I focused more on the captions because they were there and you have to read them. They appeared and disappeared." This greater cognitive effort may also be connected to one of the participants' deliberate intention to ignore the captions when feeling tired. JP said "…when I was, I don't know, like tired, I couldn't pay much attention, so I focused on the audio."

As regards the challenges that emerged from low vocabulary knowledge, they were mainly (albeit not exclusively) addressed by the participants that obtained the lowest scores at the EFL picture vocabulary test. Overall, their vocabulary knowledge seemed to affect the extent to which they comprehended each episode. This is clearly illustrated in CV's comment, "…because sometimes a word may be key to understand an episode and I didn't understand that word." Nonetheless, given the variety of topics addressed over the process, their comprehension along the sessions varied depending on their knowledge of the key words and contents developed in each episode. As MJ explicitly explained, "…it depended on the episode…. sometimes, the vocabulary was easier, some other times, not really easy." Conversely, for some of the participants, their level of comprehension did not necessarily depend on the vocabulary and topics addressed in each episode, but on the amount of practice accumulated over the sessions. In other words, comprehension progressively improved along the treatment. This was the case of MG, "I understood more throughout the process because, at the beginning, there were some episodes that did not really make sense. But then I started to understand more."

9.1.4 Learners' strategies and processing of audiovisual input

As mentioned earlier, learners' responses showed evidence of the strategies they used to face their challenges and take advantage of the synergy between imagery and verbal input (written and aural). Even when the students were not given any kind of instruction apart from paying attention to comprehend each episode and then answer some questions, they found different ways to cope with the viewing task demands. The data indicated that the participants were not fully aware of the modality they paid most attention to. This was visible in their hesitations and their difficulty to report only one mode. Yet, their examples and explanations provided rich evidence of the role that each modality played in the viewing process and how they helped them compensate for their low proficiency level.

Although the audio was reported as a challenging aspect of their viewing experience, learners' responses suggested that this mode was an essential component that guided their viewing process. For example, JJ explained that "…without the audio you can't understand anything", while JA preferred "listening and not watching the images than watching the images and not listening." Some of the participants emphasized that the audio was always present and it had to be attended. As AG said, she may even look through the window and still listen to the audio. This was also exemplified by CV, who referred to the fact that it was impossible to cover her ears and stop attending the audio.

Other multiple illustrations were provided by the students to explain the role that images and captions played in the viewing process. As regards the use of images, learners' responses clearly indicated that they were used to support comprehension and fill knowledge gaps (e.g. word meanings), particularly in the case of fourth graders and lower proficiency learners. Likewise, the participants from both year levels consistently reported relying on captions to support comprehension. One example of this is JM's comment, when a technical problem raised awareness of the role of captions in the viewing process. JM said, "The day the screen was purple I didn't understand because I couldn't see the captions."[18] It is also important to point out that in one of the groups of fifth graders, the participants indicated that reading captions was less demanding than reading out loud in reading-only condition, which may be associated to the simultaneous exposure to aural and written input. As JP explained, "…if you read 'in your mind' you don't get really tired."

---

[18] *[Refers to the beginning of the session, when we had to ask for IT support to solve the problem].*

9.1.5 The role of after-viewing activities

Although learners' views on the use of after-viewing activities were not explicitly elicited through the questionnaire, one of the groups of fifth graders addressed how these activities influenced their viewing process. Only JG referred to the use of comprehension-focused activities when she mentioned that the checking of two comprehension questions altogether (as a class) supported learning. The conversation primarily concentrated on the benefits of construction-focused activities, as well as their learning burden. AU acknowledged that this type of activity facilitated vocabulary learning since she had to focus on key words. This is why she paid special attention to the captions, "…as we had to answer questions, sometimes I needed to see how the words were written and I had to memorize them" (AU). Still, she also indicated that she could only focus her attention on *some* of the words[19]. In the case of MJ, who was less proficient, she tried to learn word meanings while viewing, albeit this seemed to be a difficult task for her. She said "the questions about words were very difficult for me. I tried to see the meaning of the words." Another interviewee alluded to the fact that the level of difficulty was not always the same, which might have been associated to the complexity of each target language construction and the clues provided to figure out their meaning. As RT explained, "they were not so difficult, sometimes they were difficult."

9.1.6 Lack of feedback

Feedback was not provided at any point of the investigation between pretest and delayed posttest. This was done to ensure that their learning gains were the product of the viewing experience and the influence of the factors under study. Nonetheless, despite the methodological issues behind this decision, the lack of feedback was found to have some negative effects on some students' perception of their learning process. While CV was explaining how captions supported comprehension, she stopped to say "maybe, reading the captions was not that effective for me because, the last time I came to the library, I don't think I did very well. I don't know." Likewise, JC's insecurity was evident in his words, "…I learned to say 'drops', which I THINK meant 'gotas'…". This was also visible in AU and MJ's hesitation when saying how much they had learned from the viewing experience.

---

[19] *[The participants were not told which words were going to be tested].*

9.1.7 Learners' willingness to watch captioned videos in the future

Learners' comments concerning their future exposure to captioned videos either inside or outside the classroom showed evidence of their willingness to do this activity. However, their answers demonstrated that their attitude also depends on the extent to which the input characteristics (e.g. speed of captions and language complexity) facilitate comprehension and match their L2 proficiency level to make the experience appropriately challenging. As JA said, "It depends on how they speak. Because sometimes they speak too fast." This idea was also evident when the interviewees from one of the classes mentioned their plan to share a list of TV programs on the classroom board, where the main requirement was that the options had to be "easy to understand." Likewise, learners reported to be open to watch other animated cartoons such as 'Dora The Explorer' given that their slower pace and lower language complexity made the episodes comprehensible and increased the possibility of learning from the experience. Nonetheless, when it comes to more complex audiovisual materials, some of the participants seemed to be reluctant to watch them, whereas others suggested the use of L1 or reversed subtitles to ensure comprehension.

The data also indicated that, at this early age, learners' viewing experience was highly influenced by other people's behaviors and decisions, as well as by specific circumstances and events in their lives. To start with, their exposure to audiovisual input in the EFL classroom appears to depend on their teachers' beliefs and actions. After the intervention, learners' responses suggested that the use of videos remained as a reward rather than a learning tool. This could be observed in JA's comment when he indicated that whenever their teacher promised them to play a video, it would be as an additional task in one of the short sessions they had at the end of the day. In other words, their regular instructional materials (e.g. coursebook and worksheets) were still prioritized over the use of authentic materials, which were not explicitly integrated in the lesson plans. Equally important, the implementation of student-led initiatives on video-viewing also depended on learners' behavior in class. RA alluded to this issue when talking about their shared list of TV series, "but now the teacher said that we are not doing it because we misbehaved yesterday."

With respect to students' viewing habits outside the EFL classroom, their scant exposure appeared to be influenced by their relatives' actions. This was illustrated by NM, whose mother changed the settings in Netflix to encourage him to watch videos in English.

In the case of JA, he did not seem to be happy with his father pushing him to watch difficult videos in English. As for the participants who had older siblings, they reported watching some movies in English with either L1 subtitles or L2 captions when joining the activity. Still, the settings were manipulated by their relatives and these young learners kept a passive role at home. When TE said that she only watched movies in English when her sister was watching, she added, "…my sister always, always, watches movies in English with subtitles in English, she's the worst!". JJ also indicated that he may only watch movies in English when his siblings were doing it. Overall, most of the participants watched videos in English when they did not have other alternatives. For instance, CV admitted that she only watched TV in English when the videos were not available in Spanish. Similarly, AG, was forced to watch videos in English with L1 subtitles because of the circumstances. As she said "we have a small TV in the car and we don't know how to change the language, so I watch TV in English."

Given these points, in dubbing countries like Chile, learners may need greater encouragement to use captioned videos and take ownership of their viewing experience. This data showed evidence of learners' need to get effective support and guidance to improve their exposure to the target language. Sometimes, a single event or successful viewing experience may push their willingness to watch videos in English and explore new informal activities in the target language. In the case of JG, her trip to the USA had encouraged her to watch captioned-TV programs in English. Yet, JG was an exception since fifth graders acknowledged that prior to the intervention their exposure to audiovisual materials was poor or practically nonexistent. For example, JC said, "Before that, I didn't watch TV in English. I watched more programs in Spanish." Nonetheless, in one of the interviews with fifth graders, the participants unanimously agreed that the intervention awakened their interest to explore new activities in the target language. For instance, some of them reported watching some TV series with reversed or L1 subtitles, while NM decided to play some videogames in English in his phone. At the end of the conversation JC and RA commented:

JC: I think we started to watch more things in English.

RA: Yes.

JC: I think most of us.

RA: Yes, most of us.

9.2 Discussion

Learners' responses provided rich information on their viewing experience which contributes to a better understanding of the findings obtained in the previous sections. On the whole, learners' statements confirmed that the viewing experience was highly enjoyable and conducive to learning, which is a finding that concurs with the positive outcomes obtained in the present study and in previous investigations with school learners (Black, 2022; Bravo, 2008; Marzá & Torralba, 2015; Pujadas, 2019; Zabalbeascoa et al., 2015). Still, their participation in this intervention was not free of challenges, and learners' willingness to continue watching captioned videos in class seemed to be influenced by their capacity to take advantage of multimodal input to compensate for their knowledge gaps and cope with the difficulties encountered while viewing. Along middle childhood, learners gradually become more aware of their learning process and their limitations (Muñoz, 2017b; Myles, 2022), this is why the lack of feedback and instruction throughout the intervention may well have affected their viewing self-efficacy and motivation to persist in the face of difficulties (Graham, 2022, p. 188). On the whole, learners' answers did not show unwillingness to watch more videos in the future but their need to be appropriately challenged. This explains why some of the students clearly expounded that their decision was subject to the complexity of the videos that would be used in class.

As regards their self-reported language gains, the students mainly focused on the learning of vocabulary, phrases and pronunciation, which is consistent with the findings that emerged from the questionnaire and the results observed in previous studies (e.g. Pattemore et al., 2020; Pujadas, 2019; Vanderplank, 1988). Although the participants appeared to struggle to identify the mode they mainly focused on (i.e. audio, captions or imagery), they provided several examples that illustrated how each modality supported the viewing process and enhanced learning. To start with, learners' responses corroborated that imagery supported comprehension and filled knowledge gaps (Durbahn et al., 2020, 2022; Peters, 2018; Rodgers, 2020), especially in the case of the less proficient and younger participants (Peters & Muñoz, 2020). Thus, considering that more concrete words tend to be graphically represented on screen (Peters, 2020), this finding may additionally explain why concreteness did not only play a key role in receptive form-meaning mapping, but also in written word-form recall (Clark & Paivio, 1991; Sadoski et al., 2004). As mentioned in the literature review, the simultaneous processing and encoding of verbal and non-verbal information enhances learning and further retrieval (Clark & Paivio, 1991). In relation to the processing of audio and captions, the participants emphasized how the reading of onscreen text supported word recognition and made the audio more comprehensible (Bird & Williams, 2002; Birulés-Muntané and Soto-Faraco, 2016; Charles & Trenkic, 2015; Mitterer & McQueen, 2009). By the same token, in line with the literature, the reading process seemed to be facilitated by the aural support (Pellicer-Sánchez, 2022). In view of these observations, we can infer that learners' capacity to take advantage of imagery and the synergy between aural and written representations fostered comprehension, vocabulary learning and the development of receptive language skills.

With respect to the learning of pronunciation, this was a salient topic in the interviews, which was mainly addressed in relation to the mapping of aural and written representations and the difficulties encountered while processing a language that differs from their L1 in language transparency. In general, the participants underscored the advantages of captioned-video viewing associated to the matching between aural and written representations to enhance comprehension and learning. However, the evidence suggests that learners' awareness of the differences between languages needs the support of explicit training in phonics to make the learning of phoneme-grapheme correspondences more efficient (Marian et al., 2021; Muñoz, 2017b; Pérez Cañado, 2006; Porter, 2020). The relatively low gains of the students in written-word form recall corroborates that despite the beneficial effects of bimodal verbal input, the learning of the words that have more irregular orthographic patterns requires greater effort and time, which was especially true in the case of the younger and less proficient students. Not to mention that the gains obtained by the participants in the learning of less regular written-word forms may not necessarily be associated to the learning of orthographic patterns but rather to their prior knowledge and their capacity to memorize specific word-forms. In fact, this seemed to be demonstrated by their higher gains in the learning of shorter and more concrete words. Together, these findings confirm that, in classroom contexts, the use of captioned videos under more effective intentional learning conditions might boost learners' gains in vocabulary learning and the development of receptive language skills (Holmes & Myles, 2019; Kim & Webb, 2022a;

Nakata & Webb, 2016; Nation & Webb, 2011; Porter, 2020; Webb & Nation, 2017; Webb et al., 2020).

The comments that emerged in one of the groups of fifth graders as regards the use of construction-focused activities confirmed that learners' dual focus on comprehension and the learning of unknown target language constructions increased the cognitive load. While the use of construction-focused activities led to greater gains in written-word form recall, this finding also aligns with the significant lower ratings reported by this group in comprehension and their capacity to cope with the speed of captions. Thus, these results match those observed in Pujadas and Muñoz's (2020) study, which indicated that school learners' effort to commit target words to memory hindered comprehension. Perhaps some of the main drawbacks of the activities designed for the purpose of this study lie in the forewarning of a vocabulary-focused activity that did not anticipate the specific target items that were going to be tested, as well as the absence of repeated viewing. Previous studies with primary school learners have demonstrated that repeated viewing results in higher levels of comprehension (Teng, 2019b) and may successfully lead to vocabulary learning (Alexiou & Yfouli, 2019).

Another valuable finding concerns the place of viewing in the L2 classroom and the role of students' relatives on their exposure to subtitled videos at home. To start with, learners' assertions suggested that viewing was seen by their teachers as a filler or a reward, rather than a tool that was effectively integrated in the English class to attain specific learning objectives. This may partially explain the participants' lack of familiarity with subtitled videos outside the classroom. The literature suggests that a principled-approach to viewing, which raises awareness of its benefits and the strategies that may be used to improve the experience, is key to encourage learners to do this activity at home and increase their exposure to the L2 (Webb, 2015; Webb, 2020). While previous investigations with school learners have lent support to these claims (e.g. Black, 2022; Pujadas, 2019; Zabalbeascoa et al., 2015), it is likely that in the present study, the absence of feedback and explicit instruction may be responsible for the differences observed in the interviews among the participants. More precisely, learners' increase in their interest in watching videos with different types of subtitles at home was more evident in fifth graders, which might be associated to their socio-emotional and cognitive development, as well as their higher L2 proficiency level and stronger L1 literacy skills (Andringa, 2022; Holmes & Myles, 2019; Muñoz, 2006; Muñoz

& Spada, 2018; Myles, 2022; Singleton & Pfenninger, 2019). It can thus be suggested that fifth graders had higher levels of autonomy and were better equipped to face up to the challenges that may arise while experimenting with this activity at home. Conversely, the lower-achievers and younger participants still seemed to favor the watching of dubbed videos at home over the original versions, unless they were pushed by the circumstances (e.g. availability) or their relatives. In view of this finding, it is possible to conclude that these are the groups that require more explicit support to take ownership of their learning experience outside the classroom and develop different strategies to experiment with viewing and persist in the task (Graham, 2022). Equally important, seeing that young learners' family may play a fundamental role in this regard, a principled-approach to viewing should also inform and train their parents on the actions that should be taken to provide appropriate support.

X. General discussion and conclusions

10.1 Gains

The present study was designed to determine the effects of captioned-video viewing on L2 vocabulary learning and the development of receptive language skills in six groups of primary school learners from Chile. In addition, the aim of this study was to investigate the extent to which learners' outcomes were influenced by treatment, input and learner-related factors. Equally important, this investigation also elicited information from the participants to further explore their perceptions of their viewing experience. In general, the results revealed that the students showed significant improvement as regards vocabulary learning (i.e. written-word form recall and receptive form-meaning mapping) and the development of receptive language skills (i.e. L2 listening, as well as English and Spanish reading efficacy), supporting previous research with participants of different characteristics (e.g. Birulés-Muntané & Soto-Faraco, 2016; Linebarger et al., 2010; Pujadas & Muñoz, 2019; Teng, 2019a).

In relation to vocabulary learning, students' gains were significantly higher in written-word form and meaning recognition, which was associated to the higher cognitive demands involved in recall compared to recognition (González-Fernández & Schmitt, 2020; Montero Perez, 2022; Teng, 2019a). In addition, the differences between English and Spanish in terms of language transparency seemed to increase the difficulty of written-word form recall (Birch & Fulop, 2021; Sun-Alperin & Wang, 2008). Nonetheless, despite learners' superior performance in written-word form and meaning recognition, the results also showed evidence of the efforts associated to the processing and integration of verbal and non-verbal input to figure out the meaning of unknown items and enhance comprehension (Montero Perez, 2022; Muñoz, 2022; Suárez & Gesa, 2019).

As for the development of receptive language skills, the results showed significant gains in listening comprehension and the development of ER efficacy, whose progress appeared to go hand in hand (Hoover & Gough, 1990; Sparks, 2021; Verhoeven & van Leeuwe, 2012). As the literature suggests, there is a bidirectional association between listening and reading, which explains why their relationship was found to strengthen over time (Verhoeven & van Leeuwe, 2012). On the whole, these positive outcomes may be attributed to the beneficial effects of the processing of bimodal verbal input (audio and text), which has been found to enhance aural word recognition (Birulés-Muntané and Soto-Faraco, 2016; Charles & Trenkic, 2016) and facilitate text decoding (Kormos et al., 2019; Pellicer-Sánchez, 2022). With respect to ER efficacy, the results indicated that learners' progress was not necessarily associated to their increase in reading speed, particularly in the case of fourth graders, but to their capacity to devote less attentional resources on text decoding to attain higher levels of comprehension (Grabe & Stoller, 2020; Nassaji, 2014; Sadoski & Paivio, 2013). Additionally, the results showed that the participants also benefitted in terms of SR efficacy despite the fact that the audiovisual materials were fully in English (audio and captions). As it will be explained in the following sections, the analyses indicated that learners' L1 reading skills played a significant role throughout the process to compensate to a certain extent for their knowledge gaps and low exposure to L2 print (Yamashita, 2002). Thus, in the case of fifth graders, who had more robust L1 reading skills and greater knowledge of the L2, their more efficient use of their linguistic infrastructure while processing the videos resulted in the development of SR efficacy. Yet, their progress was not fully accounted by captioned-video viewing alone but rather by the combination of their L1 reading habits and exposure to L2 captions along the intervention.

10.2 Answers to the research questions

RQ1. To what extent does viewing distribution (i.e. shorter vs. longer lags) influence young L2 learners' gains from captioned video viewing?

The first research question focused on the influence of viewing distribution on the gains obtained from the treatment. Overall, the results revealed that fourth graders were more sensitive to the effects of viewing distribution, which might be explained by their significantly lower proficiency level in both languages (Serfaty & Serrano, 2022a, 2022b) and, possibly, to their less developed cognitive skills (Delgiudice, 2018; Kim & Webb, 2022b). Specifically, the analyses consistently suggested that shorter lags between episodes moderated the influence of learners' reading skills, confirming previous findings on the association between short-spaced conditions and the weaker role played by learners' individual differences (Collins & White, 2012), and learners' perception of higher levels of comprehension (Greving & Richter, 2021). Based on the beneficial effects of narrow viewing

(Rodgers & Webb, 2011), it is reasonable to assume that under shorter time intervals, the processing of audiovisual input is supported by the use of higher-level processing skills. In fact, the data collected by means of a questionnaire corroborated that watching more episodes a week led to higher levels of comprehension and enhanced learners' capacity to cope with the speed of captions. This finding might also be supported by 4-fourth and 4-fifth's higher self-reported reliance on captions to achieve comprehension.

In vocabulary learning, the effects of viewing distribution appeared to be more evident in written-word form recall, where 4-fourth obtained significantly higher gains in comparison with the rest of the groups in the same year level. A possible explanation for this result may be associated to the higher difficulty entailed in written-word form recall (Suzuki et al., 2019; Serfaty & Serrano, 2022a, 2022b), which may have been counteracted by the implementation of shorter lags between episodes. For the learning of receptive form-meaning links, the influence of viewing distribution was found to be neither clear-cut nor robust. This result might be explained by the lower complexity encountered at the level of recognition. While the first analyses suggested that watching one episode a week led to higher levels of retention, its effects disappeared when entering vocabulary knowledge to the analyses. Among the possible interpretations, it may be the case that viewing distribution had weak effects on learners' performance or that the advantage of 1-fourth in retention could be attributed to their slightly higher score in vocabulary knowledge rather than viewing distribution. Alternatively, it might also be hypothesized that the potential effects of longer spacing on vocabulary retention might have been overshadowed by the weaker influence of SR efficacy observed in the group that watched more episodes a week (4-fourth).

As for the development of receptive language skills, the results indicated that learners' scores in listening skills did not vary as a function of viewing distribution. In addition, although the analyses indicated that this factor did not play a role in the development of ER efficacy, the results showed evidence of 2-fourth's higher reliance on English text segmentation and SR efficacy, suggesting that longer lags between episodes made the processing of onscreen text more effortful. This finding may explain the lower gains obtained by this group over time. As explained in the literature review, the development of receptive language skills takes time and requires plenty of practice (Grabe & Stoller, 2020;

Tragant et al., 2019), this is why it may be assumed that a higher number of episodes may be required to observe clearer differences between viewing schedules.

RQ2. In comparison with meaning-focused activities, what are the effects of construction-focused after-viewing activities on L2 learning through captioned-video viewing?

With respect to the use of after-viewing activities, the results indicated that both activity types were equally effective in the learning of form-meaning links and the development of receptive language skills. Thus, the implementation of construction-focused activities was only found to be a stronger predictor of learners' gains in written-word form recall. On the grounds of learners' assertions in one of the interviews, the beneficial effects of the construction-focused activities on this regard may be the product of students' intentional effort to commit unknown items to memory (Hulstijn, 2003). However, in line with the literature, learners' self-reported outcomes also corroborated that this type of activity increased the cognitive load and hindered comprehension (Pujadas & Muñoz, 2020). In relation to written-word form and meaning recognition, the significant effects of activity type in favor of the construction-focused activities were found to be overridden by the cognitive and language-related factors, implying that learners' individual differences were stronger predictors of the learning of form-meaning links. Additionally, the weaker effects of construction-focused activities on this vocabulary dimension might also be interpreted in light of the concept of transfer-appropriateness (Brandsford et al., 1979; Lightbown, 2008) given that the participants that completed comprehension-focused activities needed to figure out the meaning of unknown words to fill knowledge gaps and achieve comprehension.

From a methodological perspective, these two types of activities were easy to implement due to the simplicity of the format that allowed the participants to work independently without noticing that each group directed its attention on different aspects. However, the implementation of two types of activities in the same classroom was also found to entail some costs. To start with, both activity types included two comprehension questions in common, which were checked out loud at the end of the session with the aim of resembling the structure of their regular English class. Yet, no further explanations were given to the students on the reasons why they were only given feedback on two questions each class. Thus, the higher cognitive load detected in the group that completed construction-focused activities might be associated to their dual focus while viewing (i.e. comprehension and learning from the input). Previous research with school learners has demonstrated that learners' intention to learn from the input hinders comprehension due to their need to split their attentional resources on both tasks (Pujadas & Muñoz, 2020). Furthermore, the empirical evidence has also shown that learners may inherently prioritize comprehension over learning (Montero Perez et al., 2018), which is a factor that might explain the lack of differences between conditions in most of the measures. In the current study, the participants that completed construction-focused activities were never given feedback on the non-comprehension questions, which is an action that might have possibly influenced the amount of attention devoted to each aspect. The fact that both activity type groups reported similar amounts of attention to each of the input modes (audio, text and imagery) might support this assumption. Finally, the construction-focused activities worked as a sort of test announcement since the students were unable to anticipate which words were going to be tested. This must have been especially overwhelming for lower proficiency learners, who might have encountered a higher number of unknown words in the episodes.

In view of these findings, the use of audiovisual materials should be complemented by intentional activities that may boost learners' gains without interfering with the viewing process, such as the use of vocabulary flashcards (Nakata & Webb, 2016; Webb et al., 2020). Yet, it is important to note than when analyzing the results from a different perspective, a valuable finding emerged. Specifically, the positive outcomes obtained in relation to the use of comprehension-focused activities suggest that viewing did promote incidental L2 learning. It may thus be concluded that in input-limited contexts, primary school learners' informal exposure to captioned videos may well strengthen their learning process.

RQ3. To what extent do learner characteristics influence young L2 learners' gains from captioned-video viewing?

The present study sought to determine the influence of cognitive and language-related factors on learners' outcomes over time. On the whole, the yields of this investigation consistently demonstrated that the language-related factors were stronger predictors of learners' performance in vocabulary learning and the development of receptive language skills. In fact, written-word form and meaning recognition was the only measure where the cognitive factors were found to play a more prominent role. This result corroborated the demands involved in the processing and integration of verbal and non-verbal input (Mayer, 2014; 2022; Montero Perez, 2022; Pellicer-Sánchez, 2022; Suárez & Gesa, 2019). More precisely, the results in this regard indicated that the three factors, PSTM, complex working memory and visual processing speed, influenced the extent to which the participants learned form-meaning links. Yet, PSTM did not reach significance levels when fitting a model with language-related factors. Thus, considering the results obtained in previous investigations, the weaker influence of PSTM may be associated to the facilitating effects of the processing of aural and written representations (Porter, 2017). A similar picture was observed in written-form recall and receptive language skills, where the significant effects of PSTM were overshadowed by the presence of language-related factors. Concerning the role of complex working memory in the learning of form-meaning links, the results showed evidence of the need to integrate verbal and non-verbal input effectively and efficiently (Pellicer-Sánchez, 2022; Sadoski & Paivio, 2013). Finally, in light of the Dual Coding theory (Paivio, 1986), the significant effects of visual processing speed on vocabulary retention might have been associated to learners' capacity to process the visual input with greater ease to build stronger referential connections between verbal and verbal representations (Clark & Paivio, 1991; Sadoski & Paivio, 2013).

As for the influence of the L2-related factors, the results consistently demonstrated that they were the strongest predictors of vocabulary learning and the development of receptive language skills. This result may not be surprising, considering that the empirical evidence has shown that reading comprehension, which is somehow involved in the viewing process, is mainly predicted by L2-related factors (Alderson et al., 2016; Jeon & Yamashita, 2014; Proctor et al., 2005; Sparks, 2021). It is not to say that cognitive and L1 reading skills may not play a role in reading, but they seem to have a lower contribution to learners' outcomes (Alderson et al., 2016; Sparks, 2021). Thus, a higher proficiency level may make the viewing process less effortful, leaving more attentional resources available to show higher levels of comprehension and learn from the input (Kormos, 2017; Lin & Siyanova-Chanturia, 2015; Sadoski & Paivio, 2013).

The results also showed evidence of learners' interlanguage reading (Birch & Fulop, 2021; Jiang et al., 2019). More specifically, the findings appeared to confirm that at early stages, learners rely on their L1 reading skills to process the target language (Birch & Fulop, 2021), to compensate, to a certain extent, for their low exposure to print and limited knowledge about the L2 (Yamashita, 2002). As explained in the literature review, learners gradually assimilate and accommodate their L1 linguistic infrastructure to the patterns of the L2 (Perfetti et al., 2007), which is a process that evolves according to their proficiency level (Birch & Fulop, 2021; Jiang et al., 2019). Consistent with the literature, not all the groups were found to rely on SR efficacy to the same extent. As previously expounded, the strongest relationships between SR efficacy and the target variables were found in year 4, specifically in the groups that watched fewer episodes a week.

RQ4. To what extent do context and word-related factors influence vocabulary learning?

The varying relative gains observed among the items corroborated that context and word-related factors may either facilitate or hamper their learnability (Barclay, 2021; Barclay & Pellicer-Sánchez, 2022; Peters, 2020). To start with, the results indicated that word frequency played a significant role in learners' outcomes when the repetitions were concentrated in a single episode (Uchihara et al., 2019). Thus, the odds of learning the target words increased with the number of encounters, unless the repetitions were distributed across the episodes. In addition, word regularity was found to be a strong predictor of word learning, particularly in written-word form recall, which was associated to the differences between English and Spanish in terms of language transparency (Hamada & Koda, 2008; Ijalba & Obler, 2015; Sun-Alperin & Wang, 2008). Therefore, the words whose orthographic patterns were less consistent with the regular patterns of L1 Spanish were evidently harder to learn and may need additional instruction to show greater improvement in this regard (Marian et al., 2021; Muñoz, 2017b; Pérez Cañado, 2006; Porter, 2020).

As for word concreteness, the results confirmed that higher concreteness ratings enhanced the learning of written forms and form-meaning links (Sadoski et al., 2004). Specifically, its facilitating effects may be accounted by their higher imageability (Peters, 2020) and saliency in the input (Crossley et al., 2016), as well as the tendency to be graphically represented in audiovisual materials (Peters, 2020). In light of the Dual Coding Theory (Paivio, 1986), concreteness mediates the strength of the referential connections between verbal and non-verbal representations, which determine learners' capacity to evoke the words and phrases encoded (Clark & Paivio, 1991). This may explain why concreteness not only did enhance the learning of form-meaning links but also the recall of written-word forms (Sadoski et al., 2004).

For word length, a different picture emerged in each word dimension. On the whole, the results corroborated that shorter words are easier to learn (Ellis & Beaton, 1993a; Barclay & Pellicer-Sánchez, 2022), which was especially true in the case of written-word form recall. As explained in the literature review, their lower learning burden may be attributed to the fact that they are easier to process (Grabe, 2009) and store in PSTM (Birch, 2015). Interestingly, in written-word form and meaning recognition, the results also indicated that word length was not an obstacle in the learning of more concrete words. On the contrary, the synergy between higher length and concreteness appeared to boost learners' gains. This result was found to be partially congruent with the one obtained by Puimège and Peters (2019b), who suggested that the saliency of longer words in the stream of speech enhanced their acquisition.

It is important to acknowledge that these findings may be somewhat limited by the use of authentic materials. In comparison with previous investigations, the target items could not be manipulated to measure the effects of each specific variable and control for the possible interactions between the factors (Barclay, 2021; Barclay & Pellicer-Sánchez, 2022). Still, the insights gained from these analyses may be of assistance for L2 teachers and materials developers to identify the words that may be more easily picked from the input and which ones would require additional intentional efforts to be learnt.

RQ5. What are students' perceptions of the viewing experience?

The question on how suitable and beneficial the viewing experience was for the groups of primary school learners may not be fully answered if learners' perceptions of the treatment were not considered. Learners' views are essential to make sound conclusions and take fully advantage of the learning potential of captioned videos (Pinter, 2017, 2022). On the whole, the information elicited by means of a questionnaire and group interviews clearly demonstrated that the viewing experience was highly enjoyable. However, given that in middle childhood learners become gradually more aware of their learning process, their goals and limitations (Muñoz, 2017c; Myles, 2022), the attention should not only be drawn to their levels of enjoyment but also to their viewing self-efficacy, their perceived outcomes and hitches, along with the strategies used to face the difficulties.

On the whole, the viewing experience was proven to be highly beneficial. Nonetheless, the data also demonstrated that the process was not exempt from challenges. The younger and less proficient participants reported lower levels of comprehension and greater efforts to cope with the input demands. This finding is consistent with the results obtained from the statistical analyses, which showed that the language-related factors were the strongest predictors of learning. Hence, fifth graders were arguably better prepared to face the challenges. Nonetheless, despite the reported difficulties, the interviews also provided rich evidence on the strategies the students used to persist in the task. Actually, they provided multiple examples on how the use of imagery supported comprehension and filled knowledge gaps. Similarly, they illustrated how the simultaneous processing of aural and written representations facilitated input decoding and supported comprehension. Hence, despite the absence of explicit strategy instruction, the participants were capable of taking advantage of the different modalities to somehow face the difficulties encountered throughout the viewing experience.

Overall, the results suggest that despite the challenges and differences in gains, the use of captioned videos may be suitable and beneficial for both year levels. Yet, fourth graders' lower (albeit high) willingness to watch more episodes in the future or experiment with viewing at home suggests that some actions need to be taken to help these students to be more in control of the factors and strategies that may allow them to face the challenges more effectively and persist over time (Graham, 2022). Some of their needs were clearly identified in the interviews. To illustrate, learners' assertions evinced their need to test their hypotheses and assess their progress. In addition, some of them expressed their frustration when being unable to figure out the meaning of the words that were essential to comprehend some episodes. Thus, primary school learners' viewing experience may be improved through the implementation of vocabulary pre-teaching activities (Gesa, 2019; Pujadas & Muñoz, 2019), the use of glossaries or dictionaries (Fievez et al., 2021; Teng, 2022), viewing repetition (Alexiou & Yfouli, 2019; Teng, 2019b), as well as the provision of feedback and strategy instruction (Graham, 2022).

XI. Pedagogical implications

This study has several pedagogical implications. Our findings have demonstrated that the use of captioned videos is appropriate and beneficial for primary school learners since the age of 9-10. However, there are multiple factors that should be considered in order to ensure that learners' viewing experience becomes sufficiently motivating and conducive to learning. As the literature suggests, along middle childhood, children become more goal oriented and aware of their learning process (Myles, 2022). Therefore, learners' eagerness to do an activity may not only depend on their levels of enjoyment but also on their actual capacity to deal with the challenges and clearly benefit from it (Muñoz, 2017c; Myles, 2022). Accordingly, the selection of audiovisual materials that matches learners' characteristics is key to attain appropriate levels of comprehension and foster learning (Lin & Siyanova-Chanturia, 2015). In the present investigation we used eleven episodes of the animated cartoon *Charlie and Lola* (Carrington & Child, 2005-2008) due to its highly supportive imagery and high vocabulary coverage at K1 level. In addition, the episodes did not contain complex storylines and the speech was clearly enunciated (Donaghy, 2019). Although the fact that this animated cartoon aims at toddlers and preschoolers may be a point for concern, this did not seem to be an issue for our participants. Learners' assertions indicated that they were fully aware of their limitations, so they preferred to be appropriately challenged. Furthermore, given that the participants were not familiar with the animated cartoon before this intervention, it was easier to help the children relate to the TV program by raising awareness of the fact that the main characters could somehow resemble their own experiences and relationships with siblings, cousins and friends.

As regards the use of captions, the results indicated that in primary school learners, the support of onscreen text is strongly required to make the aural input more accessible (Montero Perez, 2022). Listening activities may be quite challenging for low proficiency learners due to the online processing pressure (Kormos et al., 2019; Newton & Nation, 2021). Therefore, based on the positive outcomes obtained in this investigation, it may thus be concluded that the use of captioned videos may not only facilitate learners' immediate comprehension but also foster the development of listening skills at early stages of L2 learning. By the same token, the participants' reports also illustrated how the reading of captions was facilitated by the aural support. Thus, in view of learners' significant improvement in ER efficacy, the reading of captions may compensate, to a certain extent, for their lack of exposure to print, which is crucial to automatize lower-level reading skills (Grabe, 2009; Grabe & Stoller, 2020). Given that the use of multimodal input may help less proficient learners compensate for their knowledge gaps (Peters & Muñoz, 2020), primary school students might be encouraged to process texts they would be unable to comprehend in listening-only or reading-only condition (Pellicer-Sánchez et al., 2018). It goes without saying that the use of captioned videos would not replace the implementation of listening or reading programs; however, these multimodal resources might become a more effective tool to break the vicious circle of low-achievers' reluctance to listen or read (Birch & Fulop, 2021). Perhaps one of the key advantages of the processing of captioned videos is the simultaneous development of both receptive language skills through a common task, which is a process that may be enhanced by the bidirectional relationship between reading and listening (Verhoeven & van Leeuwe, 2012).

Although the data demonstrated that the viewing process was not exempt from challenges, learners' capacity to take advantage of multimodality seemed to be key to encourage the participants to persist and benefit from the viewing experience. Hence, due to the effort entailed in the processing of audiovisual input, it may be hypothesized that the relatively short length of the episodes (10-minute long) was optimal to keep high levels of attention. Previous research has shown that when learners struggle with the reading of captions, they may easily get off task after 10 minutes (Marzá & Torralba, 2015; Zabalbeascoa et al., 2015). It is important to acknowledge that at the beginning of every session, the students were reminded that not understanding all the ideas conveyed in the episodes was completely normal, so they had to do their best to understand as much as they could. Hence, it is likely that primary school learners additionally need consistent encouragement to persevere and face the difficulties encountered in the viewing process (Webb, 2015).

The findings of this study also suggested that in the younger and less proficient participants, watching few episodes a week (e.g. 1 or 2) was more effortful and less beneficial. Therefore, the extent to which fourth graders benefitted from narrow viewing (i.e. lighter lexical load and higher comprehensibility; Rodgers & Webb, 2011) might have depended on the distance between the episodes. As a result, seeing that the instructional time devoted to L2 teaching at schools tends to be limited, the younger and less proficient students should also watch videos at home. To this aim, Webb (2015) proposes the implementation of a principled viewing approach at school that raises awareness of the benefits of captioned videos and incorporates strategy instruction to help learners feel more in control of the factors that may allow them to face the challenges encountered while viewing at home. This is key to increase their viewing self-efficacy and ensure that this activity is sustained in time (Graham, 2022). In the present investigation, the data showed evidence of learners' poor exposure to the L2 outside the classroom, and confirmed that neither the teachers nor the families were aware of the beneficial effects of viewing (Black, 2022). Thus, teachers and families should play an active role in the implementation of extensive viewing programs to guarantee that the goal of increasing young learners' exposure to the L2 through audiovisual input is actually fulfilled.

Taken together, the results of this study strengthen the idea that captioned-video viewing should have an important place inside and outside the L2 classroom. Nonetheless, this does not mean that viewing should replace learners' formal instruction as some language aspects do need to be explicitly and effectively taught to maximize learning. To illustrate, it is likely that the students would have obtained greater gains in written-word form recall if viewing had been complemented by more effective intentional vocabulary learning activities (Webb et al., 2020) and phonics instruction (Marian et al., 2021; Muñoz, 2017b; Pérez Cañado, 2006; Porter, 2020). In this respect, the results of this investigation also indicated that the use of construction-focused activities, which only worked as a sort of test announcement, increased the cognitive load and hindered comprehension. Accordingly, this result implies that the activities designed to complement the viewing tasks should facilitate and not interfere with the comprehension process. For instance, learners' outcomes might be improved through repeated viewing (Alexiou & Yfouli, 2019; Muñoz et al., 2022; Teng, 2019b), the pre-teaching of the words that are essential to comprehend the episodes, and the use of intentional learning activities that have proven to be more effective (Webb et al., 2020).

XII. Limitations and further research

The major limitation of this study is the fact that the data was collected during the COVID-19 pandemic. Thus, many of the decisions made throughout the process were constrained by the contingency measures taken by the Chilean government and the schools' authorities. Every week, the Ministry of Health informed which towns were going to be put in quarantine, which meant that the schools were closed and the classes had to go online. This uncertainty shortened the treatment (i.e. only 11 episodes) and moved the administration of delayed posttests ahead (2-3 weeks). Given that primary school students have a slower learning rate (Holmes & Myles, 2019; Muñoz, 2008), future studies should measure the effects of longer interventions, especially in relation to the development of language skills. As regards the use of after-viewing activities, learners were not allowed to work in groups, therefore, the format was designed to ensure learners' capacity to complete the activities independently. By the same token, the fact that the school day had been reduced in a 25% during the pandemic implied that the viewing sessions had to last less than 25 minutes (half of the English class). Thus, the after-viewing activities designed for the purpose of this study were very brief and simple. Further research should explore the effects of longer and different types of activities.

On the whole, the circumstances affected the timing of the data collection due to the number of additional procedures that resulted from the COVID-19 preventive measures. Hence, the administration of each instrument had to be carefully planned, and some variables had to be prioritized. For instance, due to the relatively low gains obtained in written-word form recall, the dictation test was only administered at posttest, which means that we were unable to measure retention. Given that fifth graders showed a better performance in this respect, the assessment of written-word form recall at delayed posttest should not have been cancelled in this year level. This data would have been useful to assess the effects of viewing distribution on the retention of written-word forms (Rogers, 2017). The time constraints also affected the possibility of interviewing a higher number of participants. The small pool of students that participated in the interviews provided rich data on their viewing experience, which contributed to the interpretation of the findings that emerged from the quantitative analyses. Hence, this is likely to be a fruitful area for further work.

With respect to group size, the number of students in each viewing distribution group was limited. In addition, neither 1-fourth nor 3-fourth had a counterpart in year 5. Therefore, the findings obtained in this investigation should be interpreted with due caution. Further research should be undertaken to test the outcomes obtained in this investigation and measure the effects of different viewing schedules in fifth graders. Equally important, the use of eye-tracking methodologies would be of great help to analyze the processing patterns in each viewing distribution group. Likewise, the study of learners' eye movements may be useful to further explore the differences between year levels and proficiency groups.

Another limitation of this study was the use of captions with all the experimental groups. Further work needs to be done to compare the effects of L2 captions and L1 subtitles in primary school learners. Moreover, considering that *Charlie and Lola* is also available in audiobook format, it would also be interesting to compare the results obtained through static (book) and dynamic (video) text. Other fruitful areas for future work may be the use of different captioning conditions (e.g. keyword captions) and enhancement techniques (Teng, 2021), as well as the implementation of vocabulary pre-teaching activities and glossaries to support comprehension (Teng, 2022).

In relation to episode comprehension, the two types of activities implemented in this study had only two comprehension questions in common, which was a factor that prevented the comparisons between groups and the exploration of learners' performance along the intervention. Although learners' self-reported comprehension was key to further interpret some of the findings, more accurate measures would be required to make sound conclusions. It is also important to note that learners' reports focused on the product but not on the process. Thus, we were unable to determine whether learners' levels of comprehension improved over time or whether their results varied according to the complexity of each episode. While the interviews showed evidence of these two possible pictures, previous research with school learners has demonstrated that their perceptions may not necessarily match the actual results (e.g. Pujadas, 2019).

An issue that was not addressed in this study was the extent to this intervention actually changed learners' viewing habits and increased their exposure to the target language. The data collected by means of the questionnaire and the interviews mainly focused on the possibility of watching captioned videos in the future but this information was not corroborated through the administration of a new instrument in the following months. On the whole, the results suggested that the younger participants did need the implementation of a principled-viewing approach (Webb, 2015) to increase their viewing self-efficacy and be encouraged to experiment with viewing at home. This may certainly constitute the object of future studies.

Finally, it important to acknowledge that the analyses of the interviews lacked an interrater. The same teacher from school 1 that provided feedback on the translations was in charge of checking the final results to ensure that the interpretations matched the data. Yet, future research should ensure the participation of a second researcher in the analyses to increase the levels of reliability.

XIII. Concluding remarks

The present investigation has been one of the first attempts to thoroughly examine the benefits of captioned-video viewing with primary school learners from an input-limited context. It has demonstrated that the use of captioned videos may be feasible and fruitful since the age of 9-10 as long as the materials match learners' characteristics and allow their reliance on different modalities to compensate for their knowledge gaps. While late primary school students are more likely to benefit from viewing due to their cognitive maturity and higher proficiency level in both languages, the results of this investigation suggest that this activity does contribute to L2 learning in 9-10 year olds, and even more so when it is done with regularity (e.g. four times a week). We hope that the insights gained from this study may be of assistance to teachers, stakeholders and materials developers to increase the use of captioned videos inside and outside the primary L2 classroom.

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.

Alderson, J. C., Haapakangas, E.–L., Huhta, A., Nieminen, L., & Ullakonoja, R. (2015). *The diagnosis of reading in a second or foreign language*. Routledge.

Alderson, J.C., Huhta, A., Nieminen, L. (2016). Characteristics of weak and strong readers in a foreign language. *The Modern Language Journal, 100*(4), 1-28.

Alexiou, T. (2015). Vocabulary uptake from Peppa Pig: A case study of preschool EFL learners in Greece. In C. Gitsaki & T. Alexiou (Eds.), *Current issues in second/foreign language teaching and teacher development: Research and practice* (pp. 285–301). Cambridge Scholars Publishing.

Alexiou, T., & Yfouli, D. (2019). *Charlie & Lola: An innovative way of promoting young learners' EFL lexical development* [conference abstract]. 4th International Conference for the promotion of educational innovation, Larissa, Greece.

Allan, D. (2004). *Oxford placement test 1*. Oxford University Press.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*(6), 1698–1716.

Andringa, S. (2022). *Does literacy change language learning? Insights from awareness research* [Conference presentation]. EuroSLA 31st conference, University of Fribourg, Switzerland.

Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Anthony, L., & Nation, I.S.P. (2017). Picture Vocabulary Size Test (version 1.2.0) [Computer software and measurement instrument]. Tokyo, Japan: Waseda University. http://www.laurenceanthony.net/software/pvst

Avello, D. (2020). *Captioned video viewing and L2 learning* [Presentation]. TESOL Spain convention, Salamanca, Spain.

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

Baddeley, A. (2015). Working memory in second language learning. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 17-28). Bristol: Multilingual Matters.

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia, 49*, 1393–1400.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47-89). Academic Press.

Baddeley, A., Hitch, G., & Allen, R. (2021). A multicomponent model of working memory. In R. Logie, V. Camos, & N. Cowan (Eds.), *Working memory* (pp. 10-43). Oxford University Press.

Baltova, I. (1999). *The effect of subtitled and staged video input on the learning and retention of content and vocabulary in a second language* [Unpublished doctoral dissertation]. Ontario Institute for Studies in Education of the University of Toronto.

Barclay, S. (2021). *Examining the Learning Burden and Decay of Second Language Vocabulary Knowledge* [Unpublished doctoral dissertation]. University College London, Institute of Education.

Barclay, S., & Pellicer-Sanchez, A. (2022). Exploring the learning burden and decay of foreign language vocabulary knowledge. The effect of part of speech and word length. *International Journal of Applied Linguistics, 172*(2), 259-289. https://doi.org/10.1075/itl.20011.bar

Beal, A.L., Holdnack, J.A., Saklofske, D.H., & Weiss, L.G. (2019). Practical considerations in WISC-V interpretation and intervention. In L.G. Weiss, D.H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V: Practical resources for the mental health professional* (pp. 23-54). Academic Press.

Birch, B. (2015). *English L2 reading: Getting to the bottom* (3rd ed.). Routledge.

Birch, B., & Fulop, S. (2021). *English L2 reading: Getting to the bottom* (4th ed.). Routledge.

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*, 635– 650.

Bird, S.A. and Williams, J.N. (2002) The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509–533.

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PLoS ONE, 11*(6), 1-10. doi:10.1371/journal.pone.0158409

Black, S. (2020). Could integrated subtitles benefit young viewers? Children's reception of standard and integrated subtitles: a mixed methods approach using eye tracking, *Perspectives*, DOI: 10.1080/0907676X.2020.1849324

Black, S. (2021). The potential benefits of subtitles for enhancing language acquisition and literacy in children. An integrative review of experimental research. *Translation, Cognition & Behaviour*. https://doi.org/10.1075/tcb.00051.bla

Black, S. (2022). Subtitles as a tool to boost language learning? Children's views and experiences of watching films and television programmes in other languages with interlingual subtitles. *Journal of Audiovisual Translation, 5*(1), 73–93. https://doi.org/10.47476/jat.v5i1.2022.157

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak, & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 331–54). Erlbaum.

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. SAGE Publication.

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. Sage.

Bravo, M. C. (2008). *Putting the reader in the picture. Screen translation and foreign-language learning* [Unpublished doctoral dissertation]. University Rovira i Virgili and University of Algarve.

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

Brysbaert, M., Warriner, A. & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Method, 46*(3), 904–911.

Cambridge Assessment English. (2018). *Pre A1 starters, A1 Movers and A2 Flyers sample papers. For exams from 2018.* Retrieved from https://www.cambridgeenglish.org/exams-and-tests/movers/preparation/

Cambridge Assessment English. (2020). Key and Key for schools vocabulary list. Retrieved from https://www.cambridgeenglish.org/blog/updating-the-a2-key-and-b1-preliminary-vocabulary-lists

Caravolas, M. (2004). Spelling development in alphabetic writing systems: A cross-language perspective. *European Psychologist, 9*(1), 3–14.

Carpenter, S.K. (2017). Spacing effects on learning and memory. In J.H. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 465-485). Elsevier.

Carrington, M., & Child, L. (Executive producers). (2005-2008). *Charlie and Lola* [Animated TV series]. Tiger Aspects Productions.

Casulleras, M., & Miralpeix, I. (2017). Watching cartoons with L1 or L2 subtitles: A classroom-based study with young learners [conference abstract]. International Conference Media Literacy in Foreign Language Education Digital and Multimodal Perspectives, University of Munich, Germany.

Center, Y., Freeman, L., Robertson, G., & Outhred, L. (1999). The effect of visual imagery training on the reading and listening comprehension of low listening comprehenders in year 2. *Journal of Research in Reading 22*, 241–256. doi.org/10.1111/1467-9817.00088

Cepeda, N.J., Pashler, H., Vul, E., Wixted, J.T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning a temporal ridgeline of optimal retention. *Psychological Science, 19*, 1095– 1102.

Chang, A. (2011). The effect of reading while listening to audiobooks: Listening fluency and vocabulary gain. *Asian Journal of English Language Teaching, 19*, 43–64.

Charles, T., & Trenkic, D. (2015). Speech segmentation in a second language: The role of bi-modal input. In Y. Gambier, A. Caimi & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 173-197). Bern: Peter Lang.

Cho, K., & Krashen, S. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading, 37*(8), 662-667.

Cintrón-Valentín, M., García-Amaya, L., & Ellis, N. (2019). Captioning and grammar learning in the L2 Spanish classroom. *Language Learning Journal, 47*(4),1-21.

Cintrón-Valentín, M., & García-Amaya, L. (2021). Investigating textual enhancement and captions in L2 grammar and vocabulary. *Studies in Second Language Acquisition*, 1–26. https://doi.org/10.1017/S0272263120000492.

Clark, J., & Paivio, A. (1991). Dual Coding theory and education. *Educational Psychology Review, 3*(3), 149-210.

Cobb, T. VocabProfilers [computer program]. Accessed 30 Oct 2019 at https://www.lextutor.ca/vp/eng/

Collins, L., Halter, R., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in second language instruction. *TESOL Quarterly, 33*(4), 655–680.

Collins, L., & White, J. (2011). An intensive look at intensity and language learning. *TESOL Quarterly*, *45*, 106– 133.

Collins, L., & White, J. (2012). Closing the gap: Intensity and proficiency. In C. (Ed.), *Intensive exposure experiences in second language learning* (pp. 45–65). Bristol, England: Multilingual Matters.

Crossley, S., Kyle, K., & Salsbury, T. (2016). A usage-based investigation of L2-lexical acquisition: The role of input and output. *The Modern Language Journal, 100,* 702-715.

Danan, M. (2015). Subtitling as a language learning tool: Past findings, current applications and future paths. In Y. Gambier, A. Caimi & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 41-61). Bern: Peter Lang.

De Groot, A., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning, 50*(1), 1-56.

Dehn, M. (2022). Assessing children's working memory. In J. Schwieter, & Z. Wen (Eds.), *The Cambridge handbook of working memory and language* (pp. 219-246). Cambridge University Press.

Delgiudice, M. (2018). Middle childhood: An evolutionary-developmental synthesis. In N. Halfon, C. Forrest, R. Lerner, & E. Faustman (Eds.), *Handbook of life course health development* (pp. 95-107). Springer.

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2019). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition, 23*(1), 171- 185. https://doi.org/10.1017/S1366 728918001062

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2022). Formal versus informal L2 learning: How do individual differences and word-related variables influence French and English L2 vocabulary learning in Dutch-speaking children? *Studies in Second Language Acquisition, 44*(1), 87-111. doi:10.1017/S0272263121000097

Donaghy, K. (2019). Using film to teach languages in a world of screens. In C. Herrero & I. Vanderchelden (Eds.), *Using film and media in the language classroom: Reflections on research-led teaching* (pp. 3-16). London: Multilingual Matters.

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System, 88*, 1-13. https://doi.org/10.1016/j.system.2019. 102166

Durbahn, M., Macís, M., Rodgers, M., & Peters, E. (2022). *Lexical coverage in L2 viewing comprehension: An extension of van Zeeland and Schmitt (2012)* [Conference presentation]. EuroSLA 31st conference, University of Fribourg, Switzerland.

d'Ydewalle, G., & Pavakanun, U. (1997). Could enjoying a movie lead to language acquisition? In P. Winterhoff-Spurk & T. Van der Voort (Eds.), *New horizons in media psychology* (pp. 145-155). Westdeutscher Verlag GmbH.

d'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research, 28*(3), 227-244. doi: 10.1023/A:1023202130625

Edington, D. (2015). *Statistics for linguists: a step-by-step guide for novices.* Cambridge Scholars Publishing.

EdPuzzle. (n.d.). *Make any video your lesson*. https://edpuzzle.com/

Ehri, L. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167-188. DOI: 10.1207/s1532799xssr0902_4

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143–188.

Ellis, N. C., & Beaton, A. (1993a). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning, 43*(4), 559–617.

Ellis, N. C., & Beaton, A. (1993b). Factors Affecting the Learning of Foreign Language Vocabulary: Imagery Keyword Mediators and Phonological Short-Term Memory. *The Quarterly Journal of Experimental Psychology Section A, 46*(3), 533–558. https://doi.org/10.1080/14640749308401062

Ellis, N. C. & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form, and function. *The Modern Language Journal, 93*(3), 329–335.

Ellis, N. C., & Wulff, S. (2015). Usage based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 75-93). Routledge.

Ellis, R. (2022). Foreword. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. xxiv-xxvii). Routledge.

Enever, J. (2018). *Policy and politics in global primary English*. Oxford University Press.

Fernández Corbacho, A. (2016). *Desarrollo de habilidades tempranas de la destreza lectora en inglés como lengua extranjera* [Unpublished doctoral dissertation]. Universidad de Huelva.

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2020). Vocabulary Learning Through Viewing Captioned or Subtitled Videos and the Role of Learner- and Word-Related Factors. *CALICO Journal*, *37*(3), 233–253. https://doi.org/10.1558/cj.39370

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2021). Promoting incidental vocabulary learning through watching a French Netflix series with glossed captions. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2021.1899244

Figueredo, L. (2006). Using the known to chart the unknown: A review of first-language influence on the development of English-as-a-second-language spelling skill. *Reading & Writing, 19*, 873-905. DOI 10.1007/s11145-006-9014-1

Flanagan, D., & Alfonso, V. (2017). *Essentials of WISC-V assessment*. Wiley.

Flanagan, D., & Kaufman, A. (2009). *Essentials of WISC-IV assessment* (2nd ed.). John Wiley & Sons.

Flege, J., & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26(1)*, 1-34. doi:10.1017/S0272263104026117

French, L.M. and O'Brien, I. (2008) Phonological memory and children's second language grammar learning. *Applied Psycholinguistics 29*, 463–487.

Fried, M. (2015). Construction grammar. In T. Kiss & A. Alexiadou (Eds.), *Syntax-Theory and analysis. An international handbook* (vol.2) (pp. 974-1003). Gruyter Mouton.

Frumuselu, A. C. (2015). *Subtitled television series inside the EFL classroom: Long-term effects upon colloquial language learning and oral production* [Unpublished doctoral dissertation]. Universitat Rovira i Virgili, Tarragona, Spain.

Galimberti, V., & Miralpeix, I. (2018). Multimodal input for Italian beginner learners of English: A study on comprehension and vocabulary learning from undubbed TV series. In C. Coonan, A. Bier, & E. Ballarin (Eds.), *La didattica delle lingue nel nuovo millennio: Le sfide dell'internazionalizzazione* (pp. 615-626). Venice University Press.

Gambier, Y. (2015). Subtitles and language learning (SLL): Theoretical background. In Y. Gambier, A. Caimi & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 63-82). Bern: Peter Lang.

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. https://doi.org/10125/44684

Gathercole, S., Pickering, S., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177-190.

Gesa, F. (2019). *L1 / L2 subtitled TV series and EFL learning: a study on vocabulary acquisition and content comprehension at different proficiency levels.* [Unpublished doctoral dissertation]. Barcelona, University of Barcelona.

Gesa, F., & Miralpeix, I. (2022). Effects of watching subtitled TV series on foreign language learning. Does proficiency level matter? In C. Lütge, *Foreign language learning in the digital age. Theory and pedagogy for developing literacies* (pp. 159-173). Routledge.

Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10.

González-Fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, *41* (4), 481–505, https://doi.org/10.1093/applin/amy057

Grabe, W. (2009). *Reading in a second language. Moving from theory to practice.* Cambridge University press.

Grabe, W., & Jiang, X. (2018). First and second language reading. In J.I. Liontas, T. International Association and M. DelliCarpini (Eds.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1-7). Wiley. https://doi.org/10.1002/9781118784235.e elt0476

Grabe, W., & Stoller, F. (2013). *Teaching and researching reading* (2nd ed.). Routledge.

Grabe, W., & Stoller, F. (2020). *Teaching and researching reading* (3rd ed.). Routledge.

Graham, S. (2022). Self-efficacy and language learning. What it is and what it isn't. *Language Learning Journal, 50*(2), 186–207.

Graham, S., Courtney, L., Marinis, T., & Tonkyn, A. (2017). Early Language Learning: The Impact of Teaching and Teacher Factors. Language Learning, 67(4), 922–958. https://doi.org/10.1111/lang.12251

Granena, G., Muñoz, C., & Tragant, E. (2015). L1 reading factors in extensive L2 reading-while-listening instruction. *System, 55,* 86-89.

Gregersen, T., & MacIntyre, P. D. (2014). *Capitalizing on language learners' individuality: From premise to practice.* Multilingual Matters.

Greving, C. E., & Richter, T. (2021). Beyond the distributed practice effect: Is distributed learning also effective for learning with non-repeated text materials? *Frontiers in Psychology*, *12*, 1-17. https://doi.org/10.3389/fpsyg.2021.685245

Hamada, M. (2021). *Learning words from reading: A cognitive model of word-meaning inference.* Bloomsbury Publishing.

Hamada, M., & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning, 58*(1), 1-31.

Harris, M., & Westermann, G. (2015). *A student's guide to developmental psychology*. Psychology Press.

Hefe, E. (2013). Television subtitles and literacy: Where do we go from here? *Journal of Multilingual and Multicultural Development, 34*(7), 636-652.

Holmes, B., & Myles, F. (2019). *White Paper: Primary languages policy in England. The way forward.* RiPL: www.ripl.uk/policy/

Hoover, W., & Gough, P. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127–160.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*, 207–223.

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition, 38*, 163–175.

Hulstijn, J.H. (2001). Intentional and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge University Press.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty & M. H. Long (Eds.), *The handbook of second language research* (pp. 349–81). Blackwell.

Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (Vol. 5, pp. 2632-2640). Wiley-Blackwell. Retrieved from https://pure.uva.nl/ws/files/1664476/154062_Hulstijn_Inci dental_SLA_.pdf

Hwang, K., & Nation, P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. *Reading in a Foreign Language, 6*(1), 323-335.

Ijalba, E., & Obler, L. (2015). First language grapheme-phoneme transparency effects in adult second language learning. *Reading in a Foreign Language, 27*(1), 47-70.

Irie, K. (2022). Self-efficacy. In T. Gregersen & S. Mercer (Eds.), *The Routledge handbook of the psychology of language learning and teaching*, (pp. 100-111). Routledge.

Jeon, E., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*, 160–212.

Jeon, E. H., & Yamashita, J. (2021). Measuring L2 reading. In P. Winke, & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 265-274). Routledge.

Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality* (1st ed.). Routledge.

Jiang, J., Ouyang, J., & Liu, H. (2019). Interlanguage: A perspective of quantitative linguistic typology. *Language Sciences, 74,* 85-97.

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.) (pp.247-262). Cambridge University Press.

Kasprowicz, R., Marsden, E., & Sephton, N. (2019). Investigating distribution of practice effects for the learning of foreign language verb morphology in the young learner classroom. *The Modern Language Journal, 103*(3), 580-606.

Kim, S.K., & Webb, S. (2022a). Individual difference factors for L2 vocabulary. In S. Li, Hiver, & M. Papi (Eds.), *The Routledge Handbook of Second Language Acquisition and Individual Differences.* (pp. 282-293). Routledge.

Kim, S.K., & Webb, S. (2022b). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning, 72*, 269-319.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning, 57*, 1–44.

Koolstra, C. M., & Beentjes, J. W. J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development, 47*(1), 51–60.

Koolstra, C., van der Voort, T., & van der Kamp, L. (1997). Television's impact on children's reading comprehension and decoding skills: A 3-year panel study. *Reading Research Quarterly, 32*(2), 128-152.

Kormos, J. (2017). *The second language learning processes of students with specific learning difficulties*. Routledge.

Kormos, J., Kosak, M., & Pizorn, K. (2019). The role of low-level first language skills in second language reading, reading-while-listening and listening performance: A study of young dyslexic and non-dyslexic language learners. *Applied Linguistics, 40*(5), 834-858.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*(2), 261-271.

Koskinen, P., Wilson, R., & Jensema, C. (1985). Closed- captioned television: A new tool for reading instruction. *Reading World, 24*(4), 1-7.

Kothari, B., Takeda, J., Joshi, A., & Pandey, A. (2002). Same language subtitling: a butterfly for literacy? *International Journal of Lifelong Education, 21*(1), 55-66. DOI: 10.1080/02601370110099515

Krashen, S. (1996). The case for narrow listening. *System, 24*(1), 97-100.

Krashen, S. (2004). The case for narrow reading. *Language Magazine, 3*(5), 17-19.

Krepel, A., de Bree, E.H. &, de Jong, P.F. (2020). Does the availability of orthography support L2 word learning? *Reading and Writing, 34,* 467–496. https://doi.org/10.1007/s11145-020-10078-6

Krepel, A., de Bree, E. H., Mulder, E., van de Ven, M., Segers, E., Verhoeven, L., & de Jong, P. F. (2021). The unique contribution of vocabulary in the reading development of English as a foreign language. *Journal of Research in Reading, 44*, 453– 474. https://doi.org/10.1111/1467-9817.12350

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology, 35*(1), 65-85.

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science, 42*(3), 373–388. https://doi.org/10.1007/s11251-013-9285-2

Lanes, A., Love, R., Kalman, B., Brenchley, M., & Pickles, M. (2019). Updating the A2 Key and B1 Preliminary vocabulary lists. *Cambridge Research English Research Notes*, 1-9.

Larson-Hall, J. (2010). *A Guide to Doing Statistics in Second Language Research Using SPSS.* Routledge.

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.

Lee, M. & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: an eye-tracking study. *The Modern Language Journal, 102*(3), 557–77.

Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition , 42* (3), 625–651.

Lekkai, I. (2014). Incidental foreign-language acquisition by children watching subtitled television programs. *TOJET: The Turkish Online Journal of Educational Technology, 13*(4), 81-87.

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics, 40*(3), 693–722.

Li, S., Hiver, P., & Papi, M. (2022). Individual differences in second language acquisition: Theory, research and practice. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 3-33). Routledge.

Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z. Han (Ed.), *Understanding second language process* (pp. 27-44). Multilingual Matters.

Lightbown, P., & Spada, N. (2020). Teaching and learning L2 in the classroom: It's about time. *Language Teaching, 53*, 422-432. https://doi.org/10.1017/S0261444819000454

Lin, P.M.S., & Siyanova-Chanturia, A. (2015). Internet television for L2 vocabulary learning. In D. Nunan & J. Richards (Eds.), *Language learning beyond the classroom* (pp. 149-158). New York: Routledge.

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

Linebarger, D. (2001). Learning to read from television: The effects of using captions and narration. *Journal of Educational Psychology, 93*(2), 288–298. https://doi.org/10.1037/0 022-0663.93.2.288

Linebarger, D., Taylor Piotrowski, J., & Greenwood, C. (2010). On-screen print: the role of captions as a supplemental literacy tool. *Journal of Research in Reading, 33*(2), 148-167.

Llanes, A. (2018). Reading in English as a foreign language: Examining differences in reading speed, comprehension, efficacy and L1 cross-linguistic influence across grades. *Investigaciones Sobre Lectura, 9,* 1-49.

Llanes, A., & Tragant, E. (2021). Measuring the effects of repeated exposure to children's graded readers. *Language Teaching Research*, 1-27.

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp.227-246). Cambridge University Press.

Madlener, K. (2015). *Frequency effects in instructed second language acquisition*. G. Kristiansen, & F. Ruiz de Mendoza (Eds.). The Gruyter Mouton.

Marian, V., Bartolotti, J., Van den Berg, A., & Hayakawa, S. (2021). Costs and benefits of native language similarity for non-native word learning. *Frontiers in Psychology, 12*, 1-14.

Mariotti, C. (2015). A survey on stakeholders' perceptions of subtitles as means to promote foreign language learning. In Y. Gambier, A. Caimi & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 83-103). Bern: Peter Lang.

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials. *Studies in Second Language Acquisition, 43*(5), 985–1008. https://doi.org/10.1017/S0272263121000036

Markham, P. (1989). The effects of captioned television videotapes on the listening comprehension of beginning, intermediate and advanced ESL students. *Educational Technology, 29*, 38-41.

Marzá, A., & Torralba, G. (2015). Incidental language learning through subtitled cartoons: Is it possible in a dubbing country? In Y. Gambier, A. Caimi & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 199-219). Peter Lang.

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers and Education, 60*, 95–109.

Masoura, E., & Gathercole, S. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*, 422–429.

Mayer, R. (2014). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.) (pp. 43-71). Cambridge University Press.

Mayer, R. (2022). Cognitive theory of multimedia learning. In R. Mayer, & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed.) (pp. 57-72). Cambridge University Press.

Mayer, R., & Fiorella, L. (2022). Introduction to multimedia learning. In R. Mayer, & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed.) (pp. 3-16). Cambridge University Press.

Mayer, R., Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional videos. *Education Tech Research Dev, 68*, 837–852. https://doi.org/10.1007/s11423-020-09749-6

Mineduc. (2015). *Plan nacional de la lectura 2015-2020*. Retrieved from http://plandelectura.gob.cl/wp-content/uploads/2016/12/Plan-Nacional-Lectura-web-6-12-2016.pdf

Minucci, M., & Cárnio, M. (2010). Movie subtitles reading skills of elementary school children. *Pró-Fono Revista de Actualizaçao Científica, 22*(3), 227-232.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *IRAL – International Review of Applied Linguistics in Language Teaching, 56*(1), 1–24.

Mitchell, A., Jarvis, S., O'Malley, M., & Konstantinova, I. (2015). Working memory measures and L2 proficiency. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 270-284). Multilingual Matters.

Mitchell, R., & Rule, S. (2022). Learning vocabulary in the primary languages classroom: What corpus analysis can tell us. In K. McManus, & M. S. Schmid (Eds.), *How special are early birds? Foreign language teaching and learning* (pp. 37-61). Language Science Press.

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *Plos One, 4*(11), 1-5. http://dx.doi.org/10.1371/journal.pone.0007785.

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

Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. Language Teaching, 1-30. https://doi.org/10.1017/S0261444821000501

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

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal, 99*, 308–328.

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: the effect of two enhancement techniques, *Computer Assisted Language Learning, 31*(1), 1-26. DOI: 10.1080/09588221.2017.1375960

Montero Perez, M., & Rodgers, M. (2019). Video and language learning. *The Language Learning Journal*, 47(4), 403-406. https://doi.org/10.1080/09571736.2019.1629099

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*, 720–739.

Muñoz, C. (2006). *Age and the rate of foreign language learning.* Multilingual Matters.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, *24*(4), 578-596. Doi: 10.1093/applin/amm056

Muñoz, C. (2017a). The role of age and proficiency in subtitle reading: An eye tracking study. *System, 67*, 77-86. https://doi.org/10.1016/j.system.2017.04.015

Muñoz, C. (2017b). The development of language awareness at the transition from Primary to Secondary School. In García-Mayo, M. P. (Ed.) *Learning Foreign Languages in Primary School. Research Insights.* Multilingual Matters.

Muñoz, C. (2017c) Tracing Trajectories of Young Learners: Ten Years of School English Learning. *ARAL*, 37.

Muñoz, C. (2020a). Cognate recognition by young multilingual language learners: The role of age and exposure. In C. Bardel & L. Sánchez (Eds.), *Third language acquisition: Age, proficiency and multilingualism* (pp. 145–165). Language Science Press. DOI: 10.5281/zenodo.4138743

Muñoz, C. (2020b). Boys like games and girls like movies. Age and gender differences in out-of-school contact with English. *Revista Española de Lingüística Aplicada, 33*(1). https://doi.org/10.1075/resla.18042.mun

Muñoz, C. (2022). Audiovisual input in L2 learning. *Language, Interaction and Acquisition*, *13*(1), 125-143.

Muñoz, C., Cadierno, T., & Casas, I. (2018). Different Starting Points for English Language Learning: A Comparative Study of Danish and Spanish Young Learners. *Language Learning, 68*(4), 1–34. https://doi.org/10.1111/lang.12309

Muñoz, C., Pattemore, A. & Avello, D. (2022). Exploring repeated captioning viewing as a way to promote vocabulary learning: Time lag between repetitions and learner factors. *Computer Assisted Language Learning (CALL).*

Muñoz, C., Pujadas, G., & Pattemore, A. (2021). Audio-visual input for learning L2 vocabulary and grammatical constructions. *Second Language Research,* 1-25. https://doi.org/10.1177/02676583211015797

Muñoz, C., & Spada, N. (2018). Foreign language learning from early childhood to young adulthood. In A. De Houwer, & L. Ortega, (Eds.), *The Cambridge handbook of bilingualism* (pp. 233-249). Cambridge University Press.

Myles, F. (2022). Attitudes and motivation towards learning French in primary school. The role of developmental changes between the ages of 5 and 7. *Language Teaching for Young Learners, 1-24.* https://doi.org/10.1075/ltyl.22005.myl

Myles, F., Tellier, A., & Holmes, B. (2019). Embedding languages policy in primary schools in England: Summary of the RiPL White Paper proposing solutions. *Languages, Society & Policy,* 1-9.

Nakata, T., & Webb, S. (2016). Vocabulary learning exercises: Evaluating a selection of exercises commonly featured in language learning materials. In B. Tomlinson (Ed.), *SLA research and materials development for language learning* (pp. 123-138). Oxon, UK: Routledge.

Nassaji, H. (2014). The role and importance of lower-level processes in second language reading. *Language Teaching, 47*(1), 1-37. DOI: 10.1017/S0261444813000396

Nation, I.S.P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 15- 29). Routledge. https://doi.org/10.4 324/9780429291586

Nation, I.S.P., & Webb, S. (2011). *Researching and analysing vocabulary*. Boston: Heinle Cengage learning.

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.

Newton, J., & Nation, I.S.P. (2021). Extensive listening. In J. Newton, & I.S.P. Nation (Eds.), *Teaching ESL/EFL listening and speaking*. Routledge.

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

Paivio, A. (2014). Bilingual dual coding theory and memory. In R. Heredia & J. Altarriba (Eds.), *Foundations of bilingual memory* (pp. 41-62). Springer.

Pallant, J. (2016). *SPSS survival manual. A step by step guide to data analysis using IBM SPSS* (6th Ed.). McGraw Hill Education.

Papp, S. (2020). Phonics and literacy instruction for young learners in EFL. *Part of the Cambridge papers in ELT series*. [pdf] Cambridge: Cambridge University Press.

Paribakht, T., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy*, (pp. 174–200). Cambridge University Press.

Pattemore, A. & Muñoz, C. (2020). Learning L2 constructions from captioned audio-visual exposure: The effect of learner-related factors. *System, 93.* https://doi.org/10.1016/j.s ystem.2020.102303

Pattemore, A. & Muñoz, C. (2022a). Captions and learnability factors in learning grammar from audio-visual input. *The JALT CALL journal, 18*(1)*,* 83-109.

Pattermore, A. & Muñoz, C. (2022b). *The effects of binge-watching and spacing on learning L2 multi-word units from captioned TV series* [Manuscript submitted for publication]. Department of Modern Languages and Literatures and English Studies, University of Barcelona.

Pattemore, A., Suárez, M. D. M., & Muñoz, C. (2020). Exploring L2 TV mode preferences and perceptions of learning. In K.-M. Frederiksen, S. Larsen, L. Bradley & S. Thouësny (Eds), *CALL for widening participation: short papers from EUROCALL 2020* (pp. 272-278). Research-publishing.net. https://doi.org/10.14705/rpnet.2020.48.1200

Pattemore, M., & Serra, J. (2021). *Traditional and innovative factors that explain L2 reading skills* [Poster presentation]. EuroSLA conference 30th, University of Barcelona, Spain.

Pellicer-Sánchez, A. (2022). Multimodal reading and second language learning. *International Journal of Applied Linguistics, 172*(1). https://doi.org/10.1075/itl.21039. pel

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Llanes, A., & Serrano, R. (2018). L2 reading and reading-while-listening in multimodal learning conditions: An eye-tracking study. *ELT Research Papers*. British Council. Retrieved from https://www.teachingenglish.org.uk/sites/teacheng/files/pub_H191%20ELT_L2%20reading%20and%20reading-while-listening%20in%20multimodal_FINAL.pdf

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, A. (2020). Young learners' processing of multimodal input and its impact on reading comprehension. An eye tracking study. *Studies in Second Language Acquisition,* 1-22.

Pellicer-Sánchez, A., & Webb, S. (2022). *Examining the role of lexical coverage in reading* [Conference presentation]. BAAL Vocabulary Special Interest Group's conference, University of Exeter, UK.

Pérez Cañado, M. L. (2006). The effects of explicit spelling instruction in the Spanish EFL classroom: diagnosis, development and durability. *Language Awareness, 15*(1), 20-37.

Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L. (2007). Reading in two writing systems: Accommodation and assimilation of the brain's reading network. *Bilingualism: Language and Cognition, 10*(2), 131–146.

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

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly, 53*(4). 1008-1032.

Peters, E. (2020). Factors affecting the learning of single word items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 125-142). Routledge.

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System, 63*, 134-148. https://doi.org/10.1016/j.system.2016.10.002

Peters, E., Noreillie, A. S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning, 69*(3), 747-782.

Peters, E., & Muñoz, C. (2020). Introduction to the special issue. Language learning from multimodal input. *Studies in Second Language Acquisition*, 42, 489-497. https://doi.org/10.1017/S0272263120000212

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 1–27. https://doi.org/10.1017/S0272263117000407

Pinter, A. (2017). *Teaching young learners* (2nd ed.). Oxford University Press.

Pinter, A. (2022). Using interviews with children in L2 research. In Y. Goto Butler, & B. Huang (Eds.) *Research methods for understanding child second language development.* Routledge.

Pixabay. (n.d.). *Stunning free images & royalty free stock*. https://pixabay.com/

Porter, A. (2017). Verbal working memory and foreign language learning in English primary schools: Implications for teaching and learning. In J. Enever, & E. Lindgren (Eds.), *Early language learning. Complexity and mixed methods* (pp. 65-84). Bristol: Multilingual Matters.

Porter, A. (2020). Learning French sound/spelling links in English primary school classrooms. *EuroAmerican Journal of Applied Linguistics and Languages*, *7*(1), 78-107.

Price, K. (1983). Closed-captioned TV: an untapped resource. *MATSOL Newsletter*, *12*, 1-8.

Proctor, C. P., Carlo, M., August, D., & Snow, C. E. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology*, 97, 246–256.

Prophète, E., Sundqvist, P., & Peters, E. (2022). *Exposure and grammar: Young L2 learners' extramural English and grammaticality judgement prior to formal instruction* [Poster presentation]. EuroSLA conference 31st, University of Fribourg, Switzerland.

Puimège, E., & Peters, E. (2019a). Learners' English vocabulary knowledge prior to formal instruction: The effect of learner-related and word-related factors. *Language Learning*, *69*, 943–977.

Puimège, E., & Peters, E. (2019b). Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal, 47*(4), 424-438, DOI: 10.1080/09571736.2019.163
8630

Pujadas, G. (2019). Language learning through extensive TV viewing. A study with adolescent EFL learners [Unpublished doctoral dissertation]. Barcelona, Universitat de Barcelona. https://www.tdx.cat/handle/10803/668091#page=1

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

Pujadas, G., & Muñoz, C. (2020). Examining adolescent EFL learners' TV viewing comprehension through captions and subtitles. *Studies in Second Language Acquisition, 42*(3), 551-575. doi:10.1017/S0272263120000042

Pujadas, G., & Muñoz, C. (2022). *When to switch off captions?* [Conference presentation]. EuroSLA 31st conference, University of Fribourg, Switzerland.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *The Quarterly Journal of Experimental Psychology*, *55*(4), 1339-1362. http://www.cogsci.mq.edu.au/research/resources/nwdb/nwdb.html

Raudszus, H., Segers, E., & Verhoeven, L. (2021). Patterns and predictors of reading comprehension growth in first and second language readers. *Journal of Research in Reading*, 44, 400– 417. https://doi.org/10.1111/1467-9817.12347.

Renaissance. (n.d.). ATOS for text [online text analyzer]. Retrieved from http://www1.renaissance.com/Products/Accelerated-Reader/ATOS/ATOSAnalyzer-for-Text?_ga=2.224324059.1746568400.1644356884-1641808666.1644356884

Rieder, A. (2003). Implicit and explicit learning in incidental vocabulary acquisition. *Views*, 12, 24–39.

Rivas, F. (2019, February 21). ¿Por qué los chilenos leen poco? Jóvenes autores creen que el problema puede empezar en los colegios. *Bio Bio Chile*. Retrieved from https://www.biobiochile.cl/noticias/artes-y-cultura/libros/2019/02/21/por-que-los-chilenos-leen-poco-jovenes-autores-creen-que-el-problema-puede-empezar-en-los-colegios.shtml

Riveros, V. (forthcoming). An exploration of the impact of literature-to-film adaptations on Chilean TEFL undergraduates' reading motivation and interest: A classroom-based study (doctoral dissertation, University of York, UK).

Rodgers, M.P.H. (2013). English language learning through viewing television: an investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions [Unpublished doctoral dissertation]. New Zealand, Victoria University of Wellington.

Rodgers, M. P. H. (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.P.H. (2020). The images in television programs and the potential for learning unknown words. The relationship between on-screen imagery and vocabulary. In S. Webb (Ed.), *Approaches to learning, testing, and researching vocabulary* (pp. 191-211). Amsterdam: John Benjamins Publishing Company.

Rodgers, M.P.H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly, 45* (4), 689–717. DOI: 10.5054/tq.2011.268062

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

Rogers, J. (2015). Learning second language syntax under massed and distributed conditions. *TESOL Quarterly, 49*, 857–866.

Rogers, J. (2017). The Spacing Effect and its Relevance to Second Language Acquisition. *Applied Linguistics*, *38*, 906–911.

Rogers, J. (2021). Input spacing in second language classroom settings: Replications of Bird (2010) and Serrano (2011). *Language Teaching, 54*(3), 424-433.

Rogers, J., & Cheung, A. (2020a). Input spacing and the learning of L2 vocabulary in a classroom context. *Language Teaching Research, 24*(5), 616–641.

Rogers, J., & Cheung, A. (2020b). Does it matter when you review? Input spacing, ecological validity, and the learning of L2 vocabulary. *Studies in Second Language Acquisition*, 1–19. https://doi.org/10.1017/S0272263120000236

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16,* 183–186.

Sadoski, M. (2008). Dual Coding theory. Reading comprehension and beyond. In C. Collins, & S. Parris (Eds.), *Comprehension instruction. Research-based best practices* (pp. 38-49). The Guilford Press.

Sadoski, M., Willson, V., Holcomb, A., & Boulware-Gooden, R. (2004). Verbal and nonverbal predictors of spelling performance. *Journal of Literacy Research, 36*, 461-478.

Sadoski, M., & Paivio, A. (2007). Toward a unified theory of reading. *Scientific Studies of Reading, 11*(4), 337-356.

Sadoski, M., & Paivio, A. (2013). A dual theoretical model of reading. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (pp. 886-922). Newark: International Reading Association.

Santillana. (2014). *Lenguaje y comunicación para cuarto básico serie Bicentenario*. Chile: Santillana.

Scheffler, P., Jones, C., & Dominska, A. (2020). The Peppa Pig television series as input in pre-primary EFL instruction: A corpus-based study. International *Journal of Applied Linguistics*, 1-15.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 1–32). Cambridge University Press.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.

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

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal, 95*, 26–43.

Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition, 38*, 607–646.

Serfaty, J., & Serrano, R. (2022a). Lag effects in grammar learning: A desirable difficulties perspective. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716421000631

Serfaty, J., & Serrano, R. (2022b). *Differential lag effects for vocabulary and grammar learning* [Conference presentation]. EuroSLA 31st conference, University of Fribourg, Switzerland.

Serrano, R. (2011). The time factor in EFL classroom practice. *Language Learning, 61*, 117-145.

Serrano, R. (2012). Is intensive learning effective? Reflecting on the results from cognitive psychology and the second language acquisition literature. In C. (Ed.), *Intensive exposure experiences in second language learning* (pp. 3–22). Bristol, England: Multilingual Matters.

Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System*, *35*, 305– 321.

Serrano, R. & Pellicer-Sánchez, A. (2019). Young L2 learners' online processing of information in a graded reader during reading-only and reading-while-listening conditions: A study of eye-movements. *Applied Linguistics Review*, *13*(1), 49-70. https://doi.org/10.1515/applirev-2018-0102

Serrano, R., & Huang, H. Y. (2018). Learning vocabulary through assisted repeated reading: how much time should there be between repetitions of the same text? *TESOL Quarterly, 52*(4), 971–994. https://doi.org/10.1002/TESQ.445

Serrano, R., & Huang, H. (2021). Time distribution and intentional vocabulary learning through repeated reading: a partial replication and extension. *Language Awareness*, 1–19. https://doi.org/10.1080/09658416.2021.1894162

Service, E., & Simard, D. (2022). How measures of working memory relate to L2 vocabulary. In J. Schwieter, & Z. Wen (Eds.), *The Cambridge handbook of working memory and language* (pp. 529-549). Cambridge University Press.

Singleton, D., & Pfenninger, S. (2019). The age debate: A critical overview. In S. Garton & F. Copland (Eds.), *The Routledge handbook of teaching English to young learners* (pp. 30-43). Routledge.

Socket, G., & Kusyk, M. (2015). Online informal learning of English: frequency effects in the uptake of chunks of language from participation in web-based activities. In T. Cadierno & S.W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 153-177). Berlin, Germany: de Gruyter.

Sparks, R.L. (2021). Identification and characteristics of strong, average, and weak foreign language readers: The simple view of reading model. *The Modern Language Journal, 105*, 507-525. https://doi.org/10.1111/modl.12711

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360–407.

Stæhr, L.S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal, 36*(2), 139–152.

Suárez, M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: do proficiency and aptitude make a difference? *The Language Learning Journal, 47*(4), 1-21. https://doi.org/10.1080/09571736.2019.1616806

Suárez, M., Gilabert, R., & Moskvina, N. (2021). The mediating role of vocabulary size, working memory and attention in early vocabulary learning under different TV genres. *Tesol Journal, 12*(4), 1-29. https://doi.org/10.1002/tesj.637

Subsaga. (n.d.). Charlie and Lola [online scripts]. Retrieved from https://subsaga.com/bbc/childrens/charlie-and-lola/

Sun-Alperin, M. K., & Wang, M. (2008). Spanish-speaking children's spelling errors with English vowel sounds that are represented by different graphemes in English and Spanish words. *Contemporary Educational Psychology, 33*, 932-948.

Suzuki, Y. (2021). The cognitive approach. In T. Gregersen & S. Mercer (Eds.), *The Routledge handbook of the psychology of language learning and teaching* (pp. 7-21). Routledge.

Suzuki, Y., Nakata, T., & DeKeyser, R. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *Modern Language Journal*, *103*, 713-720.

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals, 38*(3), 422-427.

Teng, F. (2019a). Incidental vocabulary learning for primary school students: the effects of L2 caption type and word exposure frequency. *The Australian Educational Researcher*, *46*, 113- 136.

Teng, F. (2019b). Maximizing the potential of captions for primary school ESL students' comprehension of English language video. *Computer Assisted Language Learning*, 1-27. https://doi.org/10.1080/09588221.2018.1532912

Teng, F. (2021). *Language learning through captioned videos. Incidental vocabulary acquisition.* Routledge.

Teng, F. (2022). Vocabulary learning through videos: captions, advance-organizer strategy, and their combination, *Computer Assisted Language Learning, 35*(3), 518-550. DOI: 10.1080/09588221.2020.1720253

Teng, F., & Zhang, D. (2021). The associations between working memory and the effects of multimedia input on L2 vocabulary learning. *International Review of Applied Linguistics in Language Teaching,* 1-29. https://doi.org/10.1515/iral-2021-0130

Torres-Díaz, R., Mosquera, R., Ontivero, M., Romero, Y., González, E., Alvarez-Rivero, A., Ojeda, J., Peón, B., Recio, & B., Valdés-Sosa, M. (2020). Text segmentation ability predicts future reading efficiency in Spanish-speaking children. *System, 204,* 1-6. https://doi.org/10.1016/j.actpsy.2020.103015

Toscano-Fuentes, C.M., & Julián de Vega, C. (2018). Videos musicales en el aula de inglés de primaria para la mejora de la fluidez lectora. *Tejuelo, 28*, 43-66.

Tragant, E., & Vallbona, A. (2018). Reading while listening to learn: Young EFL learners' perceptions. *ELT Journal*. https ://doi.org/10.1093/elt/ccy00 9

Tragant, E., Llanes, À., & Pinyana, À. (2019). Linguistic and non-linguistic outcomes of a reading-while-listening program for young learners of English. *Reading and Writing, 32***,** 819–838. https://doi.org/10.1007/s11145-018-9886-x

Tragant, E., & Pellicer-Sánchez, A. (2019). Young EFL learners' processing of multimodal input: Examining learners' eye movements. *System*, *80*, 212-223. https://doi.org/10.1016/j.system.2018.12.002

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux. *Journal of Learning Disabilities, 45*, 453–66.

Uchihara, T., Webb, S. & Yanagisawa, A. (2019). The Effects of Repetition on Incidental Vocabulary Learning: A Meta-Analysis of Correlational Studies. *Language Learning*, *69* (3), 559 – 599.

Ullman, M., & Lovelett. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research, 34,* 39- 65.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505. https://doi.org/10.3758/BF031 92720

Van Heuven, W., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190.

Vanderplank, R. (1988), The value of teletext sub-titles in language learning. *ELT Journal, 42*(4), 272-281. https://doi.org/10.1093/elt/42.4.272

Vanderplank, R. (2015). Thirty years of research into captions/same-language subtitles and second/foreign language learning: Distinguishing between 'effects of' subtitles and 'effects with' subtitles for future research. In Y. Gambier, A. Caimi & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 19-40). Bern: Peter Lang.

Vanderplank, R. (2016). '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. DOI: 10.1017/s0261444813000207.

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, 47*(4), 407-423. https://doi.org/10.1080/09571736.2019.1610033

Van Lommel, S., Laenen, A., & d'Ydewalle, G. (2006). Foreign grammar acquisition while watching subtitled television programs. *British Journal of Educational Psychology, 76*, 243–258. https://doi.org/10.1348/000709905X38946

van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 33*, 1–24.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609-624.

Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and writing*, *25*(8), 1805–1818. https://doi.org/10.1007/s11145-011-9346-3

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*, 219–258.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*, 130–163.

Web FX. (n.d.). Readability test [online text analyzer]. Retrieved from https://www.webfx.com/tools/read-able/

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. Richards (Eds.), *Language learning beyond the classroom* (pp. 159-168). New York: Routledge.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225-239). Routledge.

Webb, S. (2021). Research investigating lexical coverage and lexical profiling: What we know, what we don't know, and what needs to be examined. *Reading in a Foreign Language, 33*(2), 287-302.

Webb, S., & Chang, A. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Journal/La Revue canadienne des langues vivantes, 68*(3), 267–290.

Webb, S., & Chang, A. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research, 19*(6), 667–686.

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

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

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

Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary learning activities? A meta-analysis. *The Modern Language Journal, 20,* 1-24. https://doi.org/10.1111/modl.12671

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children (WISC-IV) administration and scoring manual* (4th ed.). The Psychological Corporation.

Wechsler, D. (2014). *Technical manual for the Wechsler intelligence scale for children* (5th ed.). Pearson.

Weiss, L. G., Saklofske, D. H., Holdnack, J. A., & Prifitera, A. (2019). WISC-V: Advances in the Assessment of Intelligence. In *WISC-V Clinical Use and Interpretation* (2nd ed.) (pp. 1-21). Academic Press.

Wen, Z. (2015). Working memory in second language acquisition and processing: The phonological/executive model. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 41-62). Bristol: Multilingual Matters.

Wen, Z., & Jackson, D. (2022). Working memory. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp.54-66). Routledge.

Willis, M., & Ohashi, Y. (2012). A model of L2 vocabulary learning and retention. *Language Learning Journal, 40*(1), 125–137.

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

Wright, A. (2010). *Pictures for language learning*. Cambridge University Press.

Wright, C. (2015). Working memory and L2 development across the lifespan: A commentary. In Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 285-300). Bristol: Multilingual Matters.

Yamashita, J. (2002). Mutual compensation between L1 reading ability and L2 language proficiency in L2 reading comprehension. *Journal of Research in Reading, 25*, 81–95.

Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System, 70*, 38–49.

Zabalbeascoa, P., González-Casillas, S., & Pascual-Herce, R. (2015). Bringing the SLL project to life: Engaging Spanish teenagers in learning while watching Foreign language audiovisuals. In Y. Gambier, A. Caimi & C. Mariotti (Eds.), *Subtitles and language learning. Principles, strategies and practical experiences* (pp. 105-126). Bern: Peter Lang.

Appendices

Appendix 1.

Onsite pilot-testing group

Procedures: a sample episode of Charlie and Lola (But that is MY book) was pilot tested with two groups of fourth graders (20 male and 20 female) from school 1 in November 2019. Having watched the episode, they completed a comprehension and vocabulary-focused activity in groups of four students. It is important to mention that the tasks that were pilot tested with these groups had to be completely adapted with the actual experimental groups due to the pandemic (e.g. physical distance). After completing the activity, the participants were asked to answer the questionnaire below. The Cronbach alpha coefficient obtained for items 1-6 was .702, which is considered acceptable (Pallant, 2016).

Encuesta

Nombre: ________________________________ **Apellido:** ______________________________ **Curso:** ______________

Contesta esta encuesta con **sinceridad**. **No** hay respuestas buenas ni malas, por lo que te pedimos **no dejar** preguntas sin responder.
Lee los enunciados y selecciona con una X bajo el cuadro que represente tu opinión (desde 0=No estoy de acuerdo a 6=Estoy **muy** de acuerdo). ***Responde los enunciados de práctica junto a tu profesor/a.***

Enunciados	0 **No** estoy de acuerdo	1	2	3	4	5	6 Estoy **muy** de acuerdo
Me encanta el helado de chocolate.							
No me gustan los animales.							
1. Comprendí la mayor parte del video.							
2. Pude leer y seguir los subtítulos en inglés.							
3. Me gustó ver el video.							
4. Este tipo de videos es apropiado para los niños de mi edad.							
5. Me encantaría aprender inglés viendo episodios de Charlie and Lola.							
6. El video era muy aburrido.							
7. Las instrucciones de la actividad en grupo eran claras.							
8. La actividad en grupo fue difícil de resolver.							

¿Te gustaría continuar con este tipo de clases de inglés? **Encierra: a)** si **b)** no **c)** me da igual.

¿Por qué? ____

¿Te gustaría continuar viendo "Charlie and Lola" en inglés? **Encierra: a)** si **b)** no **c)** me da igual.

¿Por qué? ____

In addition, the EFL teachers that observed the pilot session filled in the questionnaire below.

Encuesta profesores

Nombre profesor/a: ____ **Curso observado:** ______________________________

Conteste esta encuesta con **sinceridad** de acuerdo a lo que observe. **No** hay respuestas buenas ni malas, por lo que le pedimos **no dejar** preguntas sin responder.
Lea los enunciados y seleccione con una X bajo el cuadro que represente su opinión (desde 0=**No** estoy de acuerdo a 6=Estoy **muy** de acuerdo).

Enunciados	0 **No** estoy de acuerdo	1	2	3	4	5	6 Estoy **muy** de acuerdo
1. En general, los alumnos comprendieron la mayor parte del video.							
2. La velocidad de los subtítulos me pareció apropiada.							
3. Los alumnos parecieron disfrutar viendo el video.							
4. El video me pareció apropiado para la edad de los alumnos.							
5. Creo que los subtítulos ayudan a los alumnos a comprender el video.							
6. Volvería a utilizar un capítulo de Charlie y Lola en mis clases.							
7. Las instrucciones de la actividad en grupo eran claras.							
8. Los alumnos se frustraron intentando hacer la actividad en grupos.							
9. Es posible que haya dos tipos de actividades en una misma sala. Los alumnos son capaces de trabajar de forma autónoma.							
10. El comportamiento de los alumnos fue inapropiado.							

Registre el tiempo promedio que necesitaron los alumnos para hacer la actividad escrita en grupos: ____________

Comentarios adicionales respecto a la sesión (recuerde que su feedback será muy bien recibido).

____
____
____
____
____

Appendix 2.

Online pilot-testing group

Due to the pandemic, we were unable to pilot test the pen-and-paper instruments at school 1 onsite. Therefore, all the measures were computerized (Google Forms). Six parents from Chile enrolled their children to participate in this online pilot investigation (April-May, 2020), where each child had to complete a set of pre and posttests, and attend 15 viewing sessions. These activities were done individually (one-to-one sessions) in order to assess the materials qualitatively throughout the process. The sessions were implemented on a daily basis from Monday to Friday for a maximum of 40 minutes. The participants were asked to keep their camera and microphone on all the time. At the end of the process, they completed a questionnaire about their perceptions of the viewing experience. The main characteristics of the participants were as follows:

Student	Gender	Age	School	Level of English	Out-of-school activities
1	Female	12	Private	A2	Limited, watching YouTube videos
2	Male	11	Private	A2	Limited, online videogames
3	Female	10	Public	Pre-A1	Limited or non-existent
4	Female	11	Private	A1	Limited, online videogames
5	Male	9	Private	A1	Limited or non-existent
6	Male	10	Private	A1	Limited or non-existent

Appendix 3.

Technique feature analysis on the construction-focused after-viewing activity

Motivation	
Is there a clear vocabulary learning goal?	1
Does the activity motivate learning?	1
Do the learners select the words?	0
Noticing	
Does the activity focus attention on the target words?	1
Does the activity raise awareness of new vocabulary learning?	1
Does the activity involve negotiation?	0
Retrieval	
Does the activity involve retrieval of the word?	1
Is it productive retrieval?	0
Is it recall?	1
Are there multiple retrievals of each word?	0
Is there spacing between retrievals?	0
Generation	
Does the activity involve generative use?	0
Is it productive?	0
Is there a marked change that involves the use of other words?	0
Retention	
Does the activity ensure successful linking of form and meaning?	1
Does the activity involve instantiation?	0
Does the activity involve imaging?	0
Does the activity avoid interference?	1

Appendix 4.

Episode 1: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 1

A

Name: ______________________________ Class: _______________

1. En este episodio, ¿cuál fue la enseñanza?:
a) Lola debe parar de comer golosinas.
b) Es importante probar las comidas primero.
c) Es entretenido viajar por el mundo para conocer nuevos alimentos.
d) Debemos compartir nuestra comida.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es verdadera?
a) Charlie odia los tomates.
b) Charlie logró convencer a Lola de comer algunos alimentos.
c) Lola sólo come comida chatarra.
d) Los padres de Charlie y Lola se enojaron por el desorden de la cocina.
e) No lo sé.

3. Charlie dijo que este alimento proviene de Júpiter:
a) Huevos
b) Papas
c) Plátanos
d) Zanahorias
e) No lo sé.

4. Charlie dijo que este alimento está hecho de nubes:
a) Coliflor
b) Puré
c) Manzana
d) Tallarines
e) No lo sé.

5. De estas alternativas, selecciona los **DOS** alimentos que se mencionan porque a Lola **NO** le gustan:
☐ Pera
☐ Huevos
☐ Caramelos
☐ Limones
☐ Champiñones
☐ Hierbas

6. De estas alternativas, selecciona los **DOS** alimentos que **NO** se mencionan en el capítulo:
☐ Arroz
☐ Tomates
☐ Champiñones
☐ Atún
☐ Naranjas
☐ Frutillas

Episode 1

B

Name: ______________________________ Class: _______________

1. En este episodio, ¿cuál fue la enseñanza?:
a) Lola debe parar de comer golosinas.
b) Es importante probar las comidas primero.
c) Es entretenido viajar por el mundo para conocer nuevos alimentos.
d) Debemos compartir nuestra comida.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es verdadera?
a) Charlie odia los tomates.
b) Charlie logró convencer a Lola.
c) Lola sólo come comida chatarra.
d) Los padres de Charlie y Lola se enojaron por el desorden en la cocina.
e) No lo sé.

3. Charlie dijo que este alimento era "green ***drops*** from Greenland":
a) Baked beans
b) Tomatoes
c) Peas
d) Mushrooms
e) No lo sé.

4. Según Charlie, quienes comen "fish fingers"?
a) Whales
b) Aliens
c) Dancers
d) Mermaids
e) No lo sé.

5. De estas alternativas, selecciona el alimentos que se menciona porque a Lola **NO** le gusta:
a) Pear
b) Sweets
c) Lemons
d) Sausages
e) No lo sé.

6. ¿Qué ves en la imagen?
a) Cabbage
b) Eggs
c) Corn
d) Bread
e) No lo sé.

Appendix 5.

Episode 2: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 2

A

Name: ______________________ Class: ____________

1. En este episodio, ¿qué aprendió Lola?:
a) Que no es bueno engañar a la hada de los dientes.
b) Que es normal que a esa edad se nos caigan los dientes.
c) Que la hada de los dientes es millonaria.
d) Que debe cepillar sus dientes cada día.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola no quiere volver a tener un diente suelto.
b) Lola cree que las jirafas viven en granjas.
c) Lola compró un pollo con su moneda.
d) Charlie escondió el diente de Lola.
e) No lo sé.

3. Cuando Lola perdió su diente, ¿Cómo solucionó su problema?
a) Le pidió ayuda a Lotta.
b) Sonrió mientras dormía para demostrar que le faltaba un diente.
c) Le envió un mensaje a la hada de los dientes.
d) Escondió un diente falso bajo la almohada.
e) No lo sé.

4. ¿Cómo logró Lola que se cayera su diente?
a) Lo empujó con su lengua.
b) Comió un toffee.
c) Comió una manzana.
d) Giró su diente hasta que se saliera.
e) No lo sé.

5. ¿Cuál de estas afirmaciones es **falsa**?
a) Lola no cuida sus dientes.
b) Marv es el mejor amigo de Charlie.
c) Lotta cree en la hada de los dientes.
d) Lola no sabía que existía la hada de los dientes.
e) No lo sé.

B

Name: ______________________ Class: ____________

1. En este episodio, ¿qué aprendió Lola?:
a) Que no es bueno engañar a la hada de los dientes.
b) Que es normal que a esa edad se nos caigan los dientes.
c) Que la hada de los dientes es millonaria.
d) Que debe cepillar sus dientes cada día.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola no quiere volver a tener un diente suelto.
b) Lola cree que las jirafas viven en granjas.
c) Lola compró un pollo con su moneda.
d) Charlie escondió el diente de Lola.
e) No lo sé.

3. Cuando Lola perdió su diente, ¿Cómo solucionó su problema?
a) Le pidió ayuda a Lotta.
b) Sonrió mientras dormía para demostrar que le faltaba un diente.
c) Le envió un mensaje a la hada de los dientes.
d) Escondió un diente falso bajo la almohada.
e) No lo sé.

4. En la frase, "wobbly tooth", qué significa "wobbly"?
a) Que tiene caries.
b) Chueco.
c) Suelto, flojo.
d) Perdido.
e) No lo sé.

5. Une las traducciones con su palabras/frases claves. Te sobrarán dos palabras.

Pillow _____	a) Pista de auto
Whale _____	b) Dientes de leche
Tooth fairy _____	c) Hada de los dientes
Track _____	d) Almohada
Baby teeth _____	
Race _____	

Appendix 6.

Episode 3: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 3: Say Cheese

A

Name: ______________________ Class: ____________

1. En este episodio, ¿qué aprendió Lola?:
a) Que debe ser más cuidadosa con su ropa.
b) Que jamás debió recortar las fotos.
c) Que siempre hay una solución a los problemas.
d) Que no debe volver a comer en el colegio.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola quería que a su mamá le agradara la foto.
b) Lola logró mantenerse limpia.
c) Lola había salido muy bien en la foto del colegio del año pasado.
d) Lotta manchó a Lola por accidente.
e) No lo sé.

3. ¿Cuál de estas estrategias usó Charlie para que Lola no se manchara?
a) Le pidió ayuda a Lotta.
b) Distrajo a Lola para que no jugara en los charcos de agua.
c) Le dio un delantal a Lola.
d) Se quedó con la colación/snack de Lola.
e) No lo sé.

4. Al final de la historia, ¿cómo solucionó Lola su problema?
a) Se sacó una nueva foto de colegio.
b) Arregló su foto de colegio con trozos de fotos antiguas.
c) Usó un computador para arreglar su foto.
d) Le pidió ayuda a su mamá.
e) No lo sé.

5. ¿Cuál de estas afirmaciones es **falsa**?
a) Lola se manchó con pintura.
b) A Lola le encanta jugar con Charlie en los charcos de agua.
c) A Lola le encanta pintar.
d) Charlie tiene mucha paciencia.
e) No lo sé.

B

Name: ______________________ Class: ____________

1. En este episodio, ¿qué aprendió Lola?:
a) Que debe ser más cuidadosa con su ropa.
b) Que jamás debió recortar las fotos.
c) Que siempre hay una solución a los problemas.
d) Que no debe volver a comer en el colegio.
e) No lo sé.

2. En la frase "Be careful, Lola". ¿Qué significa "careful"?
a) Cuidadoso/cuidadosa.
b) Alegre.
c) Limpio/limpia.
d) Ordenado/ordenada.
e) No lo sé.

3. En la frase: "Lola has mud on her shoes", qué significa mud?
a) Agua.
b) Barro.
c) Chicle.
d) Basura.
e) No lo sé.

4. ¿Qué significa puddle?
a) Barro.
b) Charco de agua.
c) Mancha.
d) Patio del colegio.
e) No lo sé.

5. ¿Qué objeto ves en la imagen?
a) A water tray.
b) An apron.
c) A school bell.
d) A table.
e) No lo sé.

6. En la frase "Mum is pleased", qué significa "pleased"?
a) Contento/contenta.
b) Triste.
c) Enojado/enojada.
d) Preocupado/preocupada.
e) No lo sé.

Appendix 7.

Episode 4: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 4

A

Name: ______________________ Class: ____________

1. En este episodio, ¿qué aprendió Lola?
a) A ahorrar dinero para comprar juguetes.
b) A ser generosa con su mejor amiga.
c) Que los regalos no son importantes.
d) Que su amigo Soren Lorensen no es bueno.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola creía que a Lotta no le gustaría el regalo tanto como a ella.
b) Charlie tuvo un accidente y Lola lo cuidó.
c) A Lola no le alcanzaba el dinero para comprar ningún regalo.
d) Charlie no quería ir al cumpleaños de Lotta.
e) No lo sé.

3. ¿Cuál de estas afirmaciones es **falsa**?
a) Lola quería regalarle un pony a Lotta pero su patio es muy pequeño.
b) Lola quería que el regalo de Lotta fuese especial.
c) Lola quería quedarse con el regalo de Lotta.
d) Charlie quiere ser doctor cuando sea grande.
e) No lo sé.

4. ¿Qué hizo Charlie para ayudar a Lola?
a) Le regaló algunas monedas para comprar el regalo.
b) Convenció a Lola de hacer lo correcto.
c) Acusó a Lola con su mamá por abrir el regalo.
d) Le entregó el regalo a Lotta.
e) No lo sé.

5. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola le compró una pelota pequeña a Lotta.
b) Lotta le devolvió el regalo a Lola.
c) Lola tiene un amigo imaginario.
d) A Lotta no le gustó tanto el regalo de Lola.
e) No lo sé.

B

Name: ______________________ Class: ____________

1. En este episodio, ¿qué aprendió Lola?:
a) A ahorrar dinero para comprar juguetes.
b) A ser generosa con su mejor amiga.
c) Que los regalos no son importantes.
d) Que su amigo imaginario Soren Lorensen no es bueno.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola creía que a Lotta no le gustaría el regalo tanto como a ella.
b) Charlie tuvo un accidente y Lola lo cuidó.
c) A Lola no le alcanzaba el dinero para comprar ningún regalo.
d) Charlie no quería ir al cumpleaños de Lotta.
e) No lo sé.

3. Lola quería que el regalo de Lotta fuese "useful". ¿Qué significa "useful"?
a) Bonito.
b) Especial.
c) Útil.
d) Caro.
e) No lo sé.

4. Lola le quería regalar "wings" a Lotta. ¿Qué significa "wings"?
a) Mariposa.
b) Alas.
c) Avión.
d) Una cuerda de saltar.
e) No lo sé.

5. ¿Qué le regaló Lola a Lotta?
a) A doctor kit.
b) A small ball.
c) A pony.
d) A box of chocolates.
e) No lo sé.

6. ¿Qué ves en la imagen?
a) A coat.
b) A bandage.
c) A birthday present.
d) A toy shop.
e) No lo sé.

Appendix 8.

Episode 5: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 5

A

Name: ______________________ Class: ____________

1. En este episodio, ¿qué aprendió Lola?
a) Que las arañas son muy peligrosas.
b) Que las arañas son inteligentes y no debe temerles.
c) Que los insectos no son agradables.
d) Que en su casa deben exterminar las arañas.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola y Charlie encontraron dos arañas.
b) Charlie le enseña a Lola a sacar las telas de araña.
c) A Charlie no le gustan las arañas.
d) Charlie encontró una araña en su habitación.
e) No lo sé.

3. ¿Cuál de estas afirmaciones es **falsa**?
a) Lola le hizo una casa a su araña.
b) A Lola no le gusta ningún tipo de insecto.
c) Lola encontró a Sidney en el lavamanos.
d) Charlie atrapa a las arañas con un vaso.
e) No lo sé.

4. ¿Qué problema tiene Lola en este episodio?
a) No quiere que Sidney tenga frío en el jardín y necesita encontrarlo.
b) Charlie pisó a Sidney por accidente.
c) Lotta quiere quedarse con la araña de Lola.
d) Lola es atacada por una araña.
e) No lo sé.

5. ¿Cuál de estas afirmaciones es **verdadera**?
a) Charlie no quiere que Lola juegue con arañas.
b) Charlie le enseña a Lola sobre las arañas.
c) Lola lee un libro sobre las arañas.
d) Sidney es gigante y tiene 20 patas.
e) No lo sé.

B

Name: ______________________ Class: ____________

1. En este episodio, ¿qué aprendió Lola?:
a) Que las arañas son muy peligrosas.
b) Que las arañas son inteligentes y no debe temerles.
c) Que los insectos no son agradables.
d) Que en su casa deben exterminar las arañas.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola y Charlie encontraron dos arañas.
b) Charlie le enseña a Lola a sacar las telas de araña.
c) A Charlie no le gustan las arañas.
d) Lola encontró una araña en el comedor.
e) No lo sé.

3. Lola dice que Sidney es "hairy". ¿Qué significa "hairy"?
a) Bonito.
b) Peludo.
c) Divertido.
d) Oscuro.
e) No lo sé.

4. Charlie dice que las arañas son "clever". ¿Qué significa "clever"?
a) Tímidas.
b) Peligrosas.
c) Inteligentes.
d) Rápidas.
e) No lo sé.

5. ¿Qué organiza Lola para sus arañas?
a) A tea party.
b) A class.
c) A shopping day.
d) A sports day.
e) No lo sé.

6. ¿Qué ves en la imagen?
a) A spider coat.
b) A spider web.
c) A spider crab.
d) A spider box.
e) No lo sé.

Appendix 9.

Episode 6: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 6

A

Name: ______________________ Class: ____________

1. En este episodio, ¿qué aprendió Lola?
a) Que su profesora no la quiere.
b) Que su personaje en la obra era aburrido.
c) Que todos los personajes son importantes en la obra.
d) Que el personaje del sol es más importante que las hojas de los árboles.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola finalmente no participa en la obra.
b) Lola quería ser el sol en la obra.
c) Lola convenció a su profesora de ser el sol en la obra.
d) Lola se cae en el escenario al final de la obra.
e) No lo sé.

3. ¿Cuál de estas afirmaciones es **falsa**?
a) Charlie intentó subirle el ánimo a su hermana.
b) La obra se trata de las vacaciones de verano.
c) Charlie ayudó a su hermana con la vestimenta.
d) Charlie le enseñó algo nuevo a su hermana.
e) No lo sé.

4. ¿Qué problema tiene Lola en este episodio?
a) Perdió su vestimenta.
b) Mrs. Hanson eligió a otra niña para ser el sol en la obra.
c) Se enfermó y no podía actuar en la obra.
d) Manchó su disfraz antes de subir al escenario.
e) No lo sé.

5. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola siempre ha amado el color café.
b) Mrs. Hanson eligió a dos niñas para ser el sol en la obra.
c) Lola finalmente queda contenta con su personaje.
d) Lola no encontró nada amarillo para hacer un disfraz.
e) No lo sé.

B

Name: ______________________ Class: ____________

1. En este episodio, ¿qué aprendió Lola?:
a) Que su profesora no la quiere.
b) Que su rol en la obra era aburrido.
c) Que todos los roles son importantes en la obra.
d) Que el sol es más importante que las hojas de los árboles.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola finalmente no participa en la obra.
b) Lola quería ser el sol en la obra.
c) Lola convenció a su profesora de ser el sol en la obra.
d) Lola se cae en el escenario al final de la obra.
e) No lo sé.

3. Lola crea su propio "sun costume." ¿Qué significa "costume"?
a) Abrigo.
b) Disfraz.
c) Guantes.
d) Cartulina.
e) No lo sé.

4. Lola es elegida como "autumn leaf." ¿Qué significa "leaf"?
a) Viento.
b) Hoja.
c) Rama.
d) Chubascos.
e) No lo sé.

5. ¿Qué significa "school play"?
a) Obra de teatro.
b) Competencia escolar.
c) Juego de mesa.
d) Una escuela divertida.
e) No lo sé.

6. ¿Qué ves en la imagen?
a) A green carpet.
b) A wide forest.
c) White snow.
d) Creepy-crawlies.
e) No lo sé.

Appendix 10

Episode 7: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 7

A

Name: ______________________________ Class: ______________

1. En este episodio, ¿qué aprendió Lola?
a) Que ella es mejor que Lotta cuidando mascotas.
b) Que es importante compartir los cuidados de una mascota.
c) Que no debiesen haber perros callejeros.
d) Que es peligroso llevar a las mascotas al parque.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola tiene un perro llamado Sizzles.
b) Marv le regala su perro a Lola.
c) Lola quiere tener un perro como mascota.
d) A Charlie le gusta correr con Sizzles en el parque.
e) No lo sé.

3. ¿Cuál de estas afirmaciones es **falsa**?
a) Lola cree que Sizzles es inteligente.
b) Marv tiene dos perros de color café.
c) Sizzles no debe comer chocolates ni caramelos.
d) Lola se encargó de cuidar a Sizzles en el parque.
e) No lo sé.

4. ¿Qué problema tiene Lola en este episodio?
a) Lotta no le permitió jugar con Sizzles.
b) Le dio chocolate a Sizzles y se intoxicó.
c) Perdió a Sizzles.
d) Necesita convencer a su papá de quedarse con Sizzles.
e) No lo sé.

5. ¿Cuál de estas afirmaciones es **verdadera**?
a) Sizzles perdió su collar.
b) Sizzles tiene una chapa con la dirección de su casa.
c) Charlie y Lola fueron al parque con su mamá.
d) Sizzles siempre se pierde.
e) No lo sé.

B

Name: ______________________________ Class: ______________

1. En este episodio, ¿qué aprendió Lola?
a) Que ella es mejor que Lotta cuidando mascotas.
b) Que es importante compartir los cuidados de una mascota.
c) Que no debiesen haber perros callejeros.
d) Que es peligroso llevar a las mascotas al parque.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es verdadera?
a) Lola tiene un perro llamado Sizzles.
b) Marv le regala su perro a Lola.
c) Lola quiere tener un perro como mascota.
d) A Charlie le gusta correr con Sizzles en el parque.
e) No lo sé.

3. ¿Cuál de estas afirmaciones es falsa?
a) Lola cree que Sizzles es inteligente.
b) Marv tiene dos perros de color café.
c) Sizzles no debe comer chocolates ni caramelos.
d) Lola se encargó de cuidar a Sizzles en el parque.
e) No lo sé.

4. ¿Cómo traducirías al español la frase "He ***will*** be OK with Lola and Lotta"?
a) Estará bien con Lola y Lotta.
b) Estuvo bien con Lola y Lotta.
c) Está bien con Lola y Lotta.
d) Ha estado bien con Lola y Lotta.
e) No lo sé.

5. Este episodio se llama "We DO promise honestly we can look after your dog."
En este caso, ¿qué crees que significa "***do*** promise"?
a) Que no cumplieron la promesa de cuidar al perro.
b) Que realmente prometen cuidar al perro.
c) Que necesitan prometer que cuidarán bien al perro.
d) Que pueden hacer una promesa para cuidar al perro.
e) No lo sé.

6. ¿Qué ves en la imagen?
a) A dog tag.
b) A dog lead.
c) A dog bed.
d) A dog bowl.
e) No lo sé.

Appendix 11.

Episode 8: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 8

A

Name: ______________________________ Class: _______________

1. En este episodio, ¿qué aprendió Lola?
a) Que Charlie nunca se asusta con nada.
b) La manera de sorprender a Charlie para asustarlo.
c) Que no es bueno contar historias de terror.
d) Que ella y sus amigos son muy valientes.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola y Lotta asustaron a Sizzles.
b) A Marv no le gustan las historias de terror.
c) Lotta es muy buena asustando a sus amigos.
d) Lola and Lotta lograron asustar a Charlie mientras veía una película.
e) No lo sé.

3. ¿Cuál de estas afirmaciones es **falsa**?
a) Lotta arruinó el plan de Lola para asustar a Charlie cuando le dió hipo.
b) Lola y Charlie invitaron a Lotta y a Marv a una pijamada.
c) A Lotta no le gustan las historias de terror.
d) Lola le cuenta a sus amigos la historia de una granja embrujada.
e) No lo sé.

4. ¿Qué problema tiene Lola en este episodio?
a) Lola no sabe como asustar a Charlie.
b) Lola le tiene miedo a la oscuridad.
c) No quiere escuchar historias de terror.
d) Charlie la abandona en un bosque.
e) No lo sé.

5. ¿Cuál de estas afirmaciones es **verdadera**?
a) Sizzles asustó a Charlie.
b) El gatito de la vecina se llamaba "Rocky."
c) Charlie se asustó al ver el gatito de la vecina.
d) Lola nunca logró asustar a Charlie.
e) No lo sé.

B

Name: ______________________________ Class: _______________

1. En este episodio, ¿qué aprendió Lola?
a) Que Charlie nunca se asusta con nada.
b) La manera de sorprender a Charlie para asustarlo.
c) Que no es bueno contar historias de terror.
d) Que ella y sus amigos son muy valientes.
e) No lo sé.

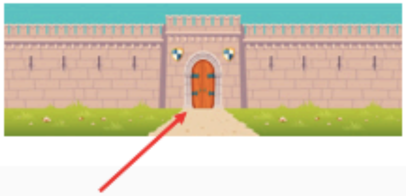
2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola y Lotta asustaron a Sizzles.
b) A Marv no le gustan las historias de terror.
c) Lotta es muy buena asustando a sus amigos.
d) Lola and Lotta lograron asustar a Charlie mientras veía una película.
e) No lo sé.

3. Lola dice que en el castillo hay "sticky spiders." ¿Qué significa "sticky"?
a) Asquerosas.
b) Pegajosas.
c) Peludas.
d) Venenosas.
e) No lo sé.

4. ¿Cómo traducirías al español la palabra "kitten"?
a) Escalera.
b) Mascota.
c) Gatito.
d) Perro cachorro.
e) No lo sé.

5. Lola dice que la escalera es "creaky." ¿Qué significa "creaky"?
a) Que tiene muchos peldaños.
b) Que cruje.
c) Que tiene la forma de un caracol.
d) Que es peligrosa.
e) No lo sé.

6. ¿Qué ves en la imagen?
a) A castle window.
b) A window castle.
c) A castle door.
d) A door castle.
e) No lo sé.



Appendix 12.

Episode 9: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 9

A

Name: ______________________________ Class: _______________

1. En este episodio, ¿qué aprendió Lola?
a) Que debe tener paciencia.
b) Que quiere ser dentista cuando sea grande.
c) Que es importante tener tiempo para jugar y descansar.
d) Que Charlie sólo quiere jugar con Marv.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola jugó con Lotta a ser dentistas.
b) Lola tiene mucha imaginación.
c) Charlie tiene un amigo llamado "Flip Flop."
d) Soren Lorenson ayudó a Lola a servir en la cafetería.
e) No lo sé.

3. ¿Cuál de estas afirmaciones es **falsa**?
a) A Sizzles le gusta comer queso.
b) Charlie se aburrió de esperar a Lola para jugar con ella.
c) Charlie desordenó todos los libros de la casa.
d) Lola no quiso jugar con Lotta porque tenía muchas cosas que hacer.
e) No lo sé.

4. ¿Qué problema tiene Lola en este episodio?
a) Imagina que está muy ocupada y no tiene tiempo de jugar con Charlie.
b) Marv no quiere que ella juegue a las cartas.
c) Lotta no quiere jugar con Lola.
d) Su mamá le pidió limpiar toda la casa y piensa que es injusto.
e) No lo sé.

5. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola imagina que debe servirle comida a Sizzles en la cafetería.
b) Lotta llamó a Marv para jugar con él.
c) Marv cree que el juego de cartas que inventó Charlie es aburrido.
d) Soren Lorenson se aburrió de jugar con Lola al dentista.
e) No lo sé.

B

Name: ______________________________ Class: _______________

1. En este episodio, ¿qué aprendió Lola?
a) Que debe tener paciencia.
b) Que quiere ser dentista cuando sea grande.
c) Que es importante tener tiempo para jugar y descansar.
d) Que Charlie sólo quiere jugar con Marv.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola jugó con Lotta a ser dentistas.
b) Lola tiene mucha imaginación.
c) Charlie tiene un amigo llamado "Flip Flop."
d) Soren Lorenson ayudó a Lola a servir en la cafetería.
e) No lo sé.

3. Lola repite muchas veces la frase "I am busy." ¿Qué significa "busy"?
a) Aburrida.
b) Ansiosa.
c) Ocupada.
d) Castigada.
e) No lo sé.

4. Lola le repitió muchas veces a Soren Lorenson "open wide". ¿Qué significa "wide"?
a) Grande y ancha.
b) Pequeña y cerrada.
c) Limpia y desinfectada.
d) Tener caries.
e) No lo sé.

5. ¿Qué ordenó Charlie en la cafetería?
a) Sausage soup and pea yoghurt.
b) Cake and coffee.
c) Tomato soup and eggs.
d) Strawberries and juice.
e) No lo sé.

6. ¿Qué ves en la imagen?
a) A door office.
b) An office wall.
c) A wall office.
d) An office door.
e) No lo sé.

Appendix 13.

Episode 10: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 10

A

Name: ________________________________ Class: ________________

1. En este episodio, ¿qué aprendió Lola?
a) Que debe tener paciencia.
b) Que es peligroso cargar maletas tan pesadas.
c) Que todo tiene solución con un poco de imaginación.
d) Que debemos compartir nuestros juguetes.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola llevó una mochila a la casa de Lotta.
b) Lola olvidó su maleta en casa.
c) Charlie jugó "flip flop" con Marv.
d) Lola y Lotta jugaron al cuento de la Caperucita Roja.
e) No lo sé.

3. ¿Cuál de estas afirmaciones es **falsa**?
a) Lola quería llevar muchas cosas a la casa de Lotta.
b) Charlie perdió una pieza de su pista para autos.
c) Lola jugó al cuento de la Cenicienta con Lotta.
d) Marv organizó una pijamada para los cuatro amigos.
e) No lo sé.

4. ¿Qué problema tiene Lola en este episodio?
a) No quiere compartir sus juguetes con Lotta.
b) Charlie le escondió los juguetes a Lola.
c) Lola olvidó su maleta con juguetes en casa.
d) Lola se aburre en la casa de Lotta.
e) No lo sé.

5. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola escondió la pieza de la pista de autos a propósito.
b) Charlie y Marv no jugaron con la pista de autos porque le faltaba una pieza.
c) Lola se disfrazó de hada madrina.
d) Charlie invitó a jugar a Soren Lorenson.
e) No lo sé.

B

Name: ________________________________ Class: ________________

1. En este episodio, ¿qué aprendió Lola?
a) Que debe tener paciencia.
b) Que es peligroso cargar maletas tan pesadas.
c) Que todo tiene solución con un poco de imaginación.
d) Que debemos compartir nuestros juguetes.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola llevó una mochila a la casa de Lotta.
b) Lola olvidó su maleta en casa.
c) Charlie jugó "flip flop" con Marv.
d) Lola y Lotta jugaron al cuento de la Caperucita Roja.
e) No lo sé.

3. ¿Qué olvidó Lola en casa?
a) A train.
b) A toothbrush.
c) A suitcase.
d) A book.
e) No lo sé.

4. ¿Qué ves en la imagen?
a) A glove.
b) A track.
c) A slipper.
d) A boot.
e) No lo sé.

5. ¿Cuál de estos objetos puso Lola en su maleta?
a) A wand.
b) An apple.
c) A broom.
d) A doll.
e) No lo sé.

6. Estas son las "stripy tights" de Lola. ¿Qué significa "stripy"?
a) Ajustadas.
b) Rayadas.
c) Cómodas.
d) Entretenidas.
e) No lo sé.

Appendix 14.

Episode 11: Meaning-focused (A) and construction-focused (B) after viewing activities

Episode 11

A

Name: ______________________________ Class: _____________

1. En este episodio, ¿qué aprendió Lola?
a) Que hay que ser cuidadoso con las pertenencias de otras personas.
b) Que debe ser generosa con su amiga Lotta.
c) Que tiene una obsesión por las cosas suaves y blandas.
d) Que debe compartir sus juguetes.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola le prestó a Lotta su libro favorito.
b) Charlie olvidó su chaqueta en la oficina de correos.
c) A Charlie no le gusta beber leche.
d) Lola perdió la chaqueta de Lotta.
e) No lo sé.

3. ¿Cuál de estas afirmaciones es **falsa**?
a) Lola y Charlie fueron a la biblioteca a leer libros.
b) Charlie es muy amable con su hermana.
c) A Lotta le encanta su chaqueta blanca.
d) Lotta manchó su chaqueta con leche rosada.
e) No lo sé.

4. ¿Qué problema tiene Lola en este episodio?
a) Rompió la chaqueta de Lotta.
b) Debe mantener la chaqueta de Lotta como nueva.
c) Olvida la chaqueta de Lotta en el colegio.
d) Lotta no quiere compartir sus cosas con ella.
e) No lo sé.

5. ¿Cuál de estas afirmaciones es **verdadera**?
a) Marv le prestó uno de sus autos a Charlie.
b) Una niña robó la chaqueta de Lotta.
c) A Lola le costó decirle la verdad a Lotta.
d) Lotta le prestó su estuche a Lola.
e) No lo sé.

B

Name: ______________________________ Class: _____________

1. En este episodio, ¿qué aprendió Lola?
a) Que hay que ser cuidadoso con las pertenencias de otras personas.
b) Que debe ser generosa con su amiga Lotta.
c) Que tiene una obsesión por las cosas suaves y blandas.
d) Que debe compartir sus juguetes.
e) No lo sé.

2. ¿Cuál de estas afirmaciones es **verdadera**?
a) Lola le prestó a Lotta su libro favorito.
b) Charlie olvidó su chaqueta en la oficina de correos.
c) A Charlie no le gusta beber leche.
d) Lola perdió la chaqueta de Lotta.
e) No lo sé.

3. ¿Qué le prestó **Lola a Lotta**?
a) A pencil case.
b) A school bag.
c) A handbag.
d) A coat.
e) No lo sé.

4. ¿De que se trataba el libro que leyó Lola en la biblioteca?
a) Pearls.
b) Mammals.
c) Shells.
d) Pets.
e) No lo sé.

5. ¿Qué ves en la imagen?
a) An umbrella.
b) A trolley.
c) A feather.
d) An apron.
e) No lo sé.

6. ¿Cómo era la chaqueta de Lotta?
a) Fluffy.
b) Comfy.
c) Clippy.
d) Wet.
e) No lo sé.

Appendix 15.

Target words of the EFL picture vocabulary test

K1 nouns (14)	K2 nouns (14)
address	bean
aunt	brush
bread	ceiling
coat	chain
daughter	cheese
dinner	dish
farmer	jam
finger	mirror
glass	onion
grass	pocket
pain	silver
ring	sink
speaker	tent
wood	wool
K1 adjectives (6)	K2 adjectives (6)
afraid	brave
cheap	crowded
dry	foreign
hurt	polite
thirsty	spare
wet	upset
K1 verbs (5)	K2 verbs (5)
agree	bake
climb	belong
kick	boil
park	dive
rest	improve

Appendix 16.

Pilot testing of the Picture EFL vocabulary test

The Picture EFL vocabulary test was administered online at two private schools in Chile. Students were asked to keep their camera on during the whole session. School A, located in Santiago, instructed an intensive English program (8 hours a week) where History and Science were taught through the target language. School B, located in the 10th region of Chile, is the same institution addressed as school 2 in the experimental groups. However, the participants from the experimental group were in year 3 at the moment this instrument was pilot tested, therefore they did not participate in this experience.

Group	Year level	School
Group 1 (*N*=32)	Third grade	A
Group 2 (*N*=34)	Third grade	A
Group 3 (*N*=31)	Third grade	A
Group 4 (*N*=33)	Third grade	A
Group 5 (*N=20*)	Fourth grade	B
Group 6 (*N=14*)	Fifth grade	B
Group 7 (*N=13*)	Sixth grade	B

Appendix 17.

Questions asked to the teachers that attended the piloting of the Picture EFL vocabulary test (online form).

1. Name and last name:
2. Class and school where the test was administered:
3. What's your opinion about the vocabulary test?
4. Do you think that the time used to administer the test was appropriate?
5. Do you think that the time the students were given to answer each question was appropriate?
6. Do you think that there was any external factor that affected their performance? Which one(s)?
7. Do you think that the level of difficulty is appropriate for the age group?
8. What would you modify to improve the instrument? Explain.

Appendix 18.

Pen-and-paper format of the EFL picture vocabulary test (sample page of the answer sheet)

Vocabulary activity

Name: .... **Class:** ....

Apple: I like apples.	A	B	?
	C	D	

1. Address: This is my address.	A	B	?
	C	D	
2. Beans: I love beans.	A	B	?
	C	D	
3. Aunt: My aunt is adorable.	A	B	?
	C	D	
4. Brush: You should use a brush.	A	B	?
	C	D	
5. Bread: I like bread.	A	B	?
	C	D	
6. Ceiling This is the ceiling.	A	B	?
	C	D	
7. Afraid: I am afraid.	A	B	?
	C	D	
8. Brave: I am brave.	A	B	?
	C	D	
9. Agree: I agree with you.	A	B	?
	C	D	
10. Bake: It's time to bake.	A	B	?
	C	D	
11. Cheap: This is cheap.	A	B	?
	C	D	
12. Climb: Can you climb?	A	B	?
	C	D	
13. Crowded: It is crowded.	A	B	?
	C	D	

Appendix 19.

Text segmentation test in Spanish

Page 1

DESAFÍO DE LECTURA

NOMBRE: ______________________________ CURSO: ________

EJEMPLO:

MINOMBREESJUANITA

Page 2

PABLOCAMINABARUMBOALCOLEGIO
ESTABACONTENTOAUNQUEUNPOCOPREOCUPADO
ENELCAMINOIBAPENSANDOCOMOESTARANSUSCOMPAÑEROS
QUIENSERASUPROFESORAYCOMOLEIRAESTEAÑO
MIENTRASSEHACÍAESTASPREGUNTAS
CASISINDARSECUENTALLEGOALCOLEGIOYUNAVEZENELPATIO
SUSCOMPAÑEROSLOVIERONYCORRIERONASALUDARLO
ALENCONTRARSEENTREAMIGOSOLVIDOSUSPREOCUPACIONES
MIENTRASCONVERSABANALEGREMENTE
SONOLACAMPANALLAMANDOLOSACLASES

Appendix 20.

Text segmentation test in English

Page 1

READING CHALLENGE

NAME: ____ CLASS: __________

EXAMPLE:

THISISACHAIR

Page 2

GOODMORNINGMYNAMEISPETER
TODAYISVERYHOT
ITISSUMMERANDTHESUNISVERYSTRONG
IKNOWHOWICANGETCOOL
ICANGOTOTHESWIMMINGPOOLINTHEAFTERNOON
ISTHEPOOLOPENORCLOSED
INEEDMYPHONETOCALLTHEPOOL
MYGREENSWIMSUITISINMYBLUEBAG
IAMREADYTOHAVEFUNWITHMYFRIENDS
ANDSHAREABASKETWITHDELICIOUSFRUITS

Appendix 21.

Questionnaire on L1-reading habits and attitude towards reading

Cuestionario

Nombre:____Curso:____________

Lee las preguntas cuidadosamente y responde *con sinceridad, no hay respuestas buenas ni malas*. No dejes ninguna en blanco.

1. Imagina que estás de cumpleaños y un amigo te regala un libro. ¿Te gustaría este regalo?

0 1 2 3 4

2. Selecciona la alternativa que más te representa. ***Aparte*** de lo que te piden leer en el colegio…

a) No leo nada más porque no me gusta.
b) Muy de vez en cuando me intereso por leer otras cosas.
c) Cada mes o cada dos meses leo un libro que me guste.
d) Me gusta leer y leo una vez por semana.
e) Me encanta leer y leo todos los días o casi todos los días.

3. Cuando eras más pequeño/pequeña, ¿Alguien te leía cuentos?

a) Nunca o rara vez.
b) Si, a veces.
c) Si, a menudo.
d) Si, todos los días.

4. En el verano pasado, ¿leíste algún libro?

a) Ninguno.

b) Si, uno.

c) Si, más de uno.

5. ¿Vas a la biblioteca a pedir libros **extras** para leer? **NO** los que te piden leer en el colegio.

a) Nunca.

b) Muy rara vez.

c) Una o dos veces al mes.

d) Todas las semanas.

6. ¿Alguna vez has pedido que te regalen un libro? Por ejemplo, para la navidad o tu cumpleaños.

a) Nunca.

b) Una vez.

c) Pocas veces.

d) Varias veces.

7. ¿Te gusta leer?

0 1 2 3 4

8. Imagina que tienes una alcancía y la rompes. ¿Comprarías un libro con el dinero?

a) Definitivamente no.

b) Tal vez, pero lo dudo.

c) Si, es algo que me gustaría comprar.

The scoring criterion for each question was as follows:

Question 1: Likert scale 0-4

Question 2:
a) 0 points
b) 1 point
c) 2 points
d) 3 points
e) 4 points

Question 3:
a) 0 points
b) 1 point
c) 2 points
d) 3 points

Question 4
a) 0 points
b) 1 point
c) 2 points

Question 5:
a) 0 points
b) 1 point
c) 2 points
d) 3 points

Question 6:
a) 0 points
b) 1 point
c) 2 points
d) 3 points

Question 7: Likert scale 0-4

Question 8:
a) 0 points
b) 1 point
c) 2 points

The maximum score was 25 and the Experimental groups' total scores ranged from 0 to 24 points (N=94, M=12.54, SD= 5.2). An independent-samples T-test revealed that the differences in total scores obtained by fourth and fifth graders were not statistically significant (M=12.64, SD=5.43 vs M=12.44, SD=5.04, respectively; t (92)= .181, p=.857, r=.01). Overall, 16% of the students reported reading once a week, whereas 24% indicated that they read every day or almost every day. Conversely, of the 56,7% remaining students, 7,3% never reads for pleasure, 24% rarely does it, and 26% reads once a month or every two months.

Appendix 22.

Questionnaire on students’ perceptions of the treatment

Final questionnaire

Name:____Class:___________

Lee las preguntas cuidadosamente y responde con sinceridad. No dejes ninguna en blanco.

1. ¿Entendiste los episodios?

0 1 2 3 4

○ ○ ○ ○ ○

2. ¿Pudiste leer y seguir los subtítulos en inglés?

0 1 2 3 4

○ ○ ○ ○ ○

3. De estas alternativas, ¿Qué te ayudó más a entender los videos? (marca sólo 1)

a) Las imágenes.

b) Los subtítulos en inglés.

c) El audio.

4. ¿A qué le pusiste más atención al ver los videos? (marca sólo 1)

a) A las imágenes.

b) A los subtítulos.

c) Al audio.

d) A nada. Me cansaba al no entender mucho y dejaba de poner atención rápidamente.

5. ¿Crees que se puede aprender inglés viendo Charlie and Lola?

0 1 2 3 4

○ ○ ○ ○ ○

6. ¿Te gustó ver los videos?

0 1 2 3 4

○ ○ ○ ○ ○

7. ¿Crees que aprendiste inglés viendo Charlie and Lola?

a) Si, bastante*.

b) Un poco*.

c) No mucho*.

d) Nada.

¿Qué cosas nuevas aprendiste en inglés al ver Charlie y Lola? Puedes seleccionar más de una casilla *

- ☐ Palabras de vocabulario
- ☐ Pronunciación
- ☐ Gramática
- ☐ Frases
- ☐ A comprender mejor los videos en inglés
- ☐ Como se escriben algunas palabras
- ☐ A leer los subtítulos en inglés
- ☐ Otro ______________________

8. ¿Te gustaría ver más videos con subtítulos en inglés en la clase de inglés?

0	1	2	3	4
○	○	○	○	○

Appendix 23.

Digits test (examiner's notes)

Forward digit span

	Intento	Respuesta	Puntos		Total
1	2-9		0	1	0 1 2
	5-4		0	1	
2	3-9-6		0	1	0 1 2
	6-5-2		0	1	
3	5-4-1-7		0	1	0 1 2
	9-1-6-8		0	1	
4	8-2-1-9-6		0	1	0 1 2
	7-2-3-4-9		0	1	
5	5-7-3-6-4-8		0	1	0 1 2
	3-8-4-1-7-5		0	1	
6	2-1-8-9-4-3-7		0	1	0 1 2
	7-8-5-2-1-6-3		0	1	
7	1-8-4-2-7-5-3-6		0	1	0 1 2
	2-7-9-6-3-1-4-8		0	1	
8	7-2-6-1-9-4-8-3-5		0	1	0 1 2
	4-3-8-9-1-7-5-6-2		0	1	
9	6-2-5-3-1-9-8-5-4-7		0	1	0 1 2
	9-4-3-8-7-5-2-9-6-1		0	1	

Backward digit span

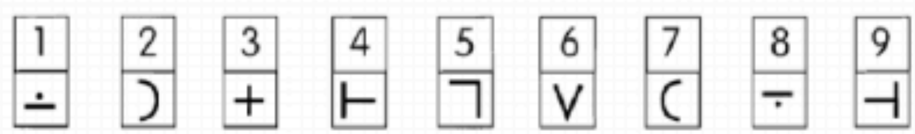
	Intento	**Respuesta esperada**	**Respuesta**	**Puntos**		**Total**
Ej	9-4	4-9		**0**	**1**	**0 1 2**
	5-6	6-5		**0**	**1**	
1	2-1	1-2		**0**	**1**	**0 1 2**
	1-3	3-1		**0**	**1**	
2	3-9	9-3		**0**	**1**	**0 1 2**
	8-5	5-8		**0**	**1**	
3	2-3-6	6-3-2		**0**	**1**	**0 1 2**
	5-4-1	1-4-5		**0**	**1**	
4	4-5-8	8-5-4		**0**	**1**	**0 1 2**
	2-7-5	5-7-2		**0**	**1**	
5	7-4-5-2	2-5-4-7		**0**	**1**	**0 1 2**
	9-3-8-6	6-8-3-9		**0**	**1**	
6	2-1-7-9-4	4-9-7-1-2		**0**	**1**	**0 1 2**
	5-6-3-8-7	7-8-3-6-5		**0**	**1**	
7	1-6-4-7-5-8	8-5-7-4-6-1		**0**	**1**	**0 1 2**
	6-3-7-2-9-1	1-9-2-7-3-6		**0**	**1**	
8	8-1-5-2-4-3-6	6-3-4-2-5-1-8		**0**	**1**	**0 1 2**
	4-3-7-9-2-8-1	1-8-2-9-7-3-4		**0**	**1**	
9	3-1-7-9-4-6-8-2	2-8-6-4-9-7-1-3		**0**	**1**	**0 1 2**
	9-8-1-6-3-2-4-7	7-4-2-3-6-1-8-9		**0**	**1**	

Appendix 24.

Coding test

Claves

Nombre: ______________________________



Reactivos muestra

2	1	4	6	3	5	2	1	3	4	2	1	3	1	2	3	1	4	2	6	3

1	2	5	1	3	1	5	4	2	7	4	6	9	2	5	8	4	7	6	1	8

7	5	4	8	6	9	4	3	1	8	2	9	7	6	2	5	8	7	3	6	4

5	9	4	1	6	8	9	3	7	5	1	4	9	1	5	8	7	6	9	7	8

2	4	8	3	5	6	7	1	9	4	3	6	2	7	9	3	5	6	7	4	5

2	7	8	1	3	9	2	6	8	4	1	3	2	6	4	9	3	8	5	1	8

Appendix 25.

Movers A

Part 1

– 5 questions –

Listen and draw lines. There is one example.

Ben Sally Nick Kim

Jane Paul Mary

Part 2

– 5 questions –

Listen and write. There is one example.

Going to the zoo

	Going to zoo today by:	train
1	Name of zoo:	Jungle ....
2	Number of different kinds of animals:	....
3	Can give food to:	....
4	Animal food in store next to:	....
5	Food on train:	.... and lemonade

Part 3

– 5 questions –

Mrs Castle is telling Sally about the people in her family and about their different hobbies. Which is each person's favourite hobby?

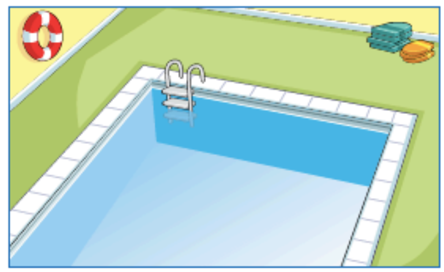
Listen and write a letter in each box. There is one example.

	her parents	H
	her uncle	
	her son	
	her cousin	
	her brother	
	her daughter	

A

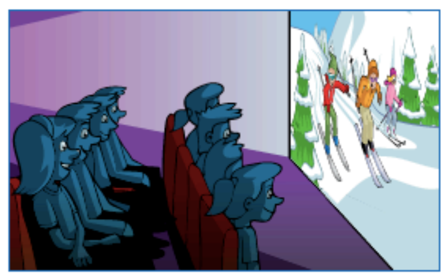
B

C

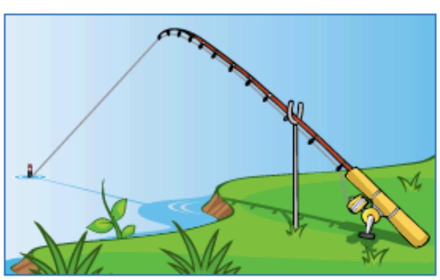


D

E



F



G

H

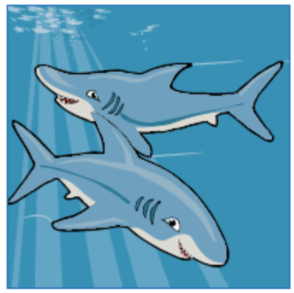
Part 4

– 5 questions –

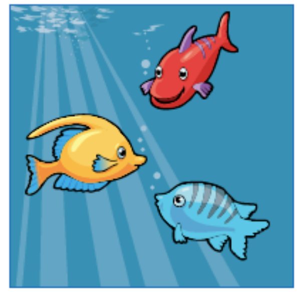
Listen and tick (✓) the box. There is one example.

What is the DVD about?

A ✓



B



C

1 Who is Vicky's piano teacher?

A

B

C

2 What clothes does Nick want to wear at school today?

A

B

C

3 Where did Peter find the shell?



A ☐ B ☐ C ☐

4 What is Daisy doing now?



A ☐ B ☐ C ☐

5 What sport did Anna get a cup for?



A ☐ B ☐ C ☐

Appendix 26.

Movers B

Part 1

– 5 questions –

Listen and draw lines. There is one example.

Mark Clare Peter Jane

Fred Daisy Vicky

Part 2

– 5 questions –

Listen and write. There is one example.

Grandma's work

	Worked at:	............City............ Hospital
1	Had to wash:	the .... in the morning
2	Number of work days every week:	....
3	Wore:	blue ....
4	At work, Grandma had:	lots of ....
5	The hospital was for:	.... only

Part 3

– 5 questions –

Mrs First is telling Paul about the people in her family. What are the people doing now?

Listen and write a letter in each box. There is one example.

her son **B**

her uncle

her daughter

her cousin

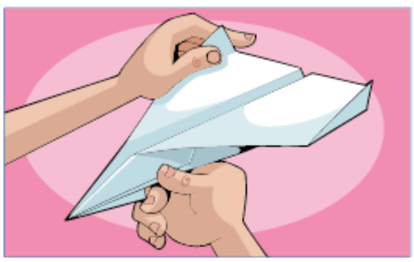
her brother

her sister

Cambridge Assessment English

A

B



C

D

E

F

G

H

Part 4

– 5 questions –

Listen and tick (✓) the box. There is one example.

Which is Charlie's favourite animal?

A ☐ B ☐ C ✓

1 Which man is the girl's teacher?

A ☐ B ☐ C ☐

2 What did Sally lose?

A ☐ B ☐ C ☐

Appendix 27.

SR efficacy test A

Page 1

<u>Lectura en español</u>

Lee este texto en **silencio** y **al ritmo que lees normalmente**, de modo que lo puedas entender bien. Recuerda que debes leerlo solo **una vez**. **Después de leer**, la examinadora te entregará una hoja con preguntas sobre el texto.

No olvides levantar la mano apenas **termines** de leer. Hemos puesto los círculos rojos en el texto para ayudarte a recordarlo.

Page 2

<u>Claudio, el pelotero</u>

Claudio era un niño fanático del fútbol. Esa mañana estaba desayunando unas ricas galletas y solo deseaba terminar su desayuno para ir al gimnasio de su comuna a jugar un partido de fútbol con sus amigos. Tomó su bicicleta, pero cuando atravesaba la calle principal se le cruzó un auto y Claudio se asustó tanto que pisó el freno delantero y cayó hacia adelante, dando tres volteretas y perdiendo totalmente el conocimiento. Cuando se despertó, estaba en el hospital y no se acordaba de nada. En realidad, sí se acordaba de algo, ese día tenía que jugar un partido, pero ahora no iba a ser posible porque se había roto una pierna, un brazo y se le habían caído tres dientes. Claudio estuvo un mes en el hospital y luego en cama en su casa por dos meses más. Cada día solo pensaba en su gran pasión: el fútbol. Sus amigos lo visitaban diariamente, conversaba y jugaba con ellos play station. Estaba mucho tiempo frente al computador y se aficionó a resolver pasatiempos como las "sopas de letras". Claudio comprendió que además del fútbol hay otras cosas importantes, como los amigos, la lectura, y por supuesto, las sopas de letras.

Lectura en español: Claudio, el pelotero

Nombre: ______________________ **Apellido:** ______________________

Curso: ______________

I. Responde las siguientes preguntas sobre el texto leído.

1. Claudio tuvo un accidente…

a) En la mañana.
b) A la hora de almuerzo.
c) Un día jueves.
d) Un fin de semana.
e) No lo sé.

2. Luego del accidente, mientras Claudio estaba en casa…

a) Se sentía muy solo.
b) No dejaba de pensar en el fútbol.
c) Se dio cuenta de que ya no le gustaba el fútbol.
d) Sólo quería volver a subirse a su bicicleta.
e) No lo sé.

3. Después de esta experiencia…

a) Claudio se alejó del fútbol y de sus amigos.
b) Claudio se dio cuenta de que también habían otras actividades entretenidas.
c) Claudio terminó odiando la lectura y las sopas de letras.
d) Claudio hizo aún más amigos.
e) No lo sé.

4. El día del accidente…

a) Claudio iba camino a la escuela.
b) Claudio iba camino al estadio a ver a su equipo favorito.
c) Claudio iba camino al gimnasio a jugar un partido con sus amigos.
d) Claudio iba distraído comiendo galletas.
e) No lo sé

5. En el texto aparece que Claudio perdió el conocimiento. En el texto **perder el conocimiento** significa :

a) Olvidarse de todo y para siempre.
b) Perder la conciencia y sentirse confundido por un rato.
c) Perder la cabeza y volverse completamente loco.
d) Tener un dolor de cabeza tan intenso que no te deja pensar.
e) No lo sé.

6. La enseñanza de la historia de Claudio es que…

a) No debes andar tan fuerte en bicicleta.
b) No es bueno ser tan fanático del fútbol.
c) Hay que disfrutar de todas las cosas que son importantes en nuestra vida.
d) No es bueno jugar tantas horas en el computador.
e) No lo sé.

Appendix 28.

SR efficacy test B

Page 1

Lectura en español

Lee este texto en **silencio** y **al ritmo que lees normalmente**, de modo que lo puedas entender bien. Recuerda que debes leerlo solo **una vez**. **Después de leer**, la examinadora te entregará una hoja con preguntas sobre el texto.

No olvides levantar tu mano cuando **termines** de leer. Hemos puesto los círculos rojos en el texto para ayudarte a recordarlo.

Page 2

El amable don Francisco

A don Francisco nunca le gustó mucho la lectura, pero hace unos años decidió abrir una librería en su barrio, la cual se transformó en su pasión. Por esto, a él le encantaba compartir y conversar largamente con sus vecinos y clientes. Don Francisco era muy amable y generoso. Su librería era un lugar muy especial, ya que todos los que entraban se podían sentar en unos cómodos sillones a leer un buen libro. Además, don Francisco les ofrecía galletas con un té, un mate o un café. Cada día, abría su librería a las 9:00 de la mañana y no la cerraba hasta la noche. Vendía muchos libros, sobre todo en diciembre en la época navideña, donde la campanilla de la puerta no dejaba de sonar en todo el día. A diario, Don Francisco tenía mucho trabajo y también se cansaba, pero al mismo tiempo, era muy feliz atendiendo a sus vecinos y clientes. En su librería, todos encontraban los libros que necesitaban, por eso también era tan famosa y estaba siempre llena de gente. Don Francisco era muy querido por todos sus clientes, quienes agradecían tener siempre las puertas abiertas de la librería para disfrutar de la lectura.

Lectura en español: El amable don Francisco

Nombre: ________________________ **Apellido:** ________________________

Curso: _______________

I. Responde las siguientes preguntas sobre el texto leído.

1. Don Francisco…

a) Amaba leer.
b) Le regalaba libros a sus vecinos.
c) Tenía su librería abierta las 24 horas del día.
d) Conversaba mucho con sus clientes.
e) No lo sé.

2. Don Francisco atendía a sus clientes con…

a) Leche y pan con mantequilla.
b) Café y queque.
c) Té, mate o café.
d) Chocolate caliente.
e) No lo sé.

3. La época en la que don Francisco tenía más trabajo era…

a) El comienzo del año escolar.
b) Los fines de semana.
c) La navidad.
d) Todo el año tenía la misma cantidad de clientes.
e) No lo sé.

4. Es cierto que a don Francisco…

a) No le gustaba admitir niños en la librería.
b) Nunca le había gustado mucho la lectura.
c) No le gustaba el sonido de la campanilla de la puerta.
d) Le gustaba beber café todo el día.
e) No lo sé

5. Del texto podemos concluir que…

a) Don Francisco ya estaba cansado de tanto trabajar y atender a tanta gente.
b) Don Francisco regalaba todos sus productos y no ganaba mucho dinero.
c) Para don Francisco, sus clientes y vecinos eran muy importantes.
d) Don Francisco tenía siempre muy pocos libros disponibles.
e) No lo sé.

6. Los clientes querían mucho a don Francisco porque…

a) Tenía los libros a muy buen precio.
b) Su librería era el único lugar donde podían comer algo gratis.
c) Era muy generoso y atento con ellos.
d) Les permitía llevar libros sin pagar por ellos.
e) No lo sé.

Appendix 29.

SR efficacy test

Page 1

Lectura en español

Lee este texto en **silencio** y **al ritmo que lees normalmente**, de modo que lo puedas entender bien. Recuerda que debes leerlo solo **una vez**. **Después de leer**, la examinadora te entregará una hoja con preguntas sobre el texto.

No olvides levantar tu mano cuando **termines** de leer. Hemos puesto los círculos rojos en el texto para ayudarte a recordarlo.

Page 2

Juan, el niño veloz

Juan era un niño muy alegre, bromista y conversador. Ya estaba en quinto básico, por lo que ya se sentía mayor. En clases, no paraba de hablar y siempre contaba chistes para hacer reír a sus amigos, a pesar de que la profesora estuviera explicando algo importante. En ocasiones, cuando ponía atención, participaba en la clase y aportaba con muy buenas ideas. Se sentía orgulloso cuando todos lo felicitaban. En los días de pruebas, siempre quería terminar primero y se apuraba mucho para lograrlo. Le gustaba gritar que había terminado primero y salía corriendo al pasillo para burlarse de sus compañeros por la ventana. Esto siempre lo hacía sentir como un ganador. Sin embargo, después de cada prueba, cuando sus compañeros comparaban sus respuestas, Juan se daba cuenta de que había dejado algunas partes de la prueba en blanco y de que no había leído bien las preguntas para responder de manera correcta. A veces, también olvidaba escribir su nombre o marcar las respuestas en el lugar correcto. Un día, la profesora conversó con Juan para ayudarlo a cambiar su actitud y desde ese momento, logró reflexionar sobre su comportamiento para mejorar sus notas y la relación con sus compañeros.

Lectura en español: Juan, el niño veloz

Nombre: ______________________ **Apellido:** ______________________

Curso: ______________

I. Responde las siguientes preguntas sobre el texto leído.

1. Es cierto que Juan…

a) Ayudaba a sus compañeros.
b) Distraía a sus compañeros.
c) Era un ejemplo para sus compañeros.
d) Era castigado todas las semanas.
e) No lo sé.

2. En los días de pruebas,

a) Juan era admirado por su rapidez.
b) Juan creía que era bueno terminar rápido.
c) Los compañeros se burlaban de los errores de Juan.
d) La profesora expulsaba a Juan de la sala de clases.
e) No lo sé.

3. Después de cada prueba…

a) Los compañeros de Juan revisaban su prueba con la profesora.
b) Los compañeros le decían a Juan en qué se había equivocado.
c) Juan debía volver a la sala de clases a escribir su nombre en la hoja.
d) Juan se daba cuenta de que no había hecho las cosas correctamente.
e) No lo sé.

4. Es cierto que…

a) Los profesores siempre culpan a Juan de todo.
b) Los profesores felicitan a Juan cuando cuenta chistes.
c) Juan no lograba hacer reír a sus compañeros.
d) Juan es capaz de compartir muy buenas ideas.
e) No lo sé

5. Del texto podemos concluir que…

a) Juan era el alumno preferido de su profesora.
b) Juan es capaz de mejorar su actitud.
c) Juan se convirtió en el mejor estudiante de la clase.
d) Todos rechazaban a Juan.
e) No lo sé.

6. La enseñanza de la historia de Juan es que…

a) Los alumnos que más interrumpen en clases son los que más saben.
b) Es importante pensar antes de actuar.
c) Las notas no son lo más importante.
d) No es una buena idea comparar las respuestas después de una prueba.
e) No lo sé.

Appendix 30.

ER efficacy test A

Page 1

<u>**Lectura en inglés**</u>

Lee este texto en **silencio** y **al ritmo que lees normalmente**, de modo que lo puedas entender bien. Recuerda que debes leerlo solo **una vez**. **Después de leer**, la examinadora te entregará una hoja con preguntas sobre el texto.

No olvides levantar la mano cuando **termines** de leer. Hemos puesto los círculos rojos en el texto para ayudarte a recordarlo.

Page 2

<u>**Lizards**</u>

Hi! My name is Peter and I want to have a lizard for a pet. Some people don't like lizards but I think they are beautiful animals. We can find them in different sizes and colors. Lots of lizards are small but some of them are very big. Lizards can be green, grey or yellow. They have a long tail at the end of their body too. Some lizards like eating spiders and some like eating fruit. A lizard can run on its four legs. Many lizards live in trees, but, at the beach, you can find some lizards on the sand. They love sleeping in the sun.

Lectura en inglés: Lizards

Nombre: ______________________ **Apellido:** ______________________

Curso: ______________

I. Responde las siguientes preguntas sobre el texto leído.

1. Es cierto que Peter…

a) Tiene una lagartija de mascota en su casa.
b) Le tiene miedo a las lagartijas.
c) Quisiera tener una lagartija de mascota.
d) Encuentra que las lagartijas son amigables.
e) No lo sé.

2. Peter explica que…

a) No todas las lagartijas son iguales.
b) Las lagartijas son grandes y tienen una lengua muy larga.
c) Hay lagartijas rojas, verdes y negras.
d) A mucha gente le encantan las lagartijas.
e) No lo sé.

3. Según Peter…

a) Las lagartijas solo se encuentran en la playa.
b) A las lagartijas les encanta dormir al sol.
c) Hay lagartijas que se esconden en el pasto.
d) Las lagartijas corren a gran velocidad.
e) No lo sé.

4. La intención de Peter en el texto es:

a) Describir a un animal que le gusta y quisiera tener en casa.
b) Pedirle al lector que cuide las lagartijas porque son animales hermosos.
c) Lograr que a todos les guste las lagartijas.
d) Describir a su mascota.
e) No lo sé.

5. Las lagartijas se alimentan de…

a) Insectos y frutos del bosque.
b) Arañas y agua.
c) Frutas y verduras.
d) Frutas y arañas.
e) No lo sé.

Appendix 31.

ER efficacy test B

Page 1

<u>**Lectura en inglés**</u>

Lee este texto en **silencio** y **al ritmo que lees normalmente**, de modo que lo puedas entender bien. Recuerda que debes leerlo solo **una vez**. **Después de leer**, la examinadora te entregará una hoja con preguntas sobre el texto.

No olvides levantar la mano cuando **termines** de leer. Hemos puesto los círculos rojos en el texto para ayudarte a recordarlo.

Page 2

<u>**Parrots**</u>

Hi! My name is Rosa and I love parrots because they are great birds. Some families have one for a pet because they are smart and beautiful. My best friend has a parrot named Jack. I don't have one in my house but I have a picture of a colorful parrot in my room. They can be green, blue, yellow, red, or a mix of many colors. They like eating fruits and insects but some like eating flowers, too. They have big, round heads and two legs. They cannot run but they can fly and talk. They can repeat words and sounds. I think they are very funny.

Lectura en inglés: Parrots

Nombre: ______________________ Apellido: ______________________

Curso: ______________

I. Responde las siguientes preguntas sobre el texto leído.

1. Es cierto que Rosa…

a) Tiene un loro de mascota en su habitación.
b) Tiene la foto de un loro.
c) Quiere tener un loro en su casa.
d) Le regaló un loro a su mejor amiga.
e) No lo sé.

2. Rosa explica que...

a) Algunos loros pueden correr.
b) Los loros son inteligentes.
c) Los loros pueden aprender a cantar.
d) Los loros son verdes.
e) No lo sé.

3. Según Rosa….

a) Los loros son grandes mascotas.
b) Los loros no deben estar en jaulas.
c) Los loros pueden repetir hasta 20 palabras.
d) Hay loros que no pueden volar.
e) No lo sé.

4. La intención de Rosa en el texto es:

a) Pedir a la gente que proteja a los loros.
b) Describir a un animal que le gusta mucho.
c) Hablar de los cuidados que deben tener los loros.
d) Explicar las razones de por qué quiere tener un loro como mascota.
e) No lo sé.

5. Según Rosa, los loros...:

a) No deben comer flores.
b) Comen gusanos y hierbas.
c) Comen frutas, flores y miel.
d) Comen flores, insectos y frutas.
e) No lo sé.

Appendix 32.

ER efficacy test C

Page 1

<u>**Lectura en inglés**</u>

Lee este texto en **silencio** y **al ritmo que lees normalmente**, de modo que lo puedas entender bien. Recuerda que debes leerlo solo **una vez**. **Después de leer**, la examinadora te entregará una hoja con preguntas sobre el texto.

No olvides levantar la mano cuando **termines** de leer. Hemos puesto los círculos rojos en el texto para ayudarte a recordarlo.

Page 2

<u>**Dolphins**</u>

Hi! My name is Bob. Dolphins are my favorite animal because they are playful. I cannot have one for a pet at home but I have a picture of a very beautiful dolphin in my bedroom. There are many types of dolphins. They can swim very fast and their colour can be a mix of black, gray and white. They sleep at night and they like to play in the water during the day. They can jump very high. They are smart and they can use sounds to "talk". They live with their families in seas and rivers. They have small teeth and they love to eat fish.

Lectura en inglés: Dolphins

Nombre: ______________________ **Apellido:** ______________________

Curso: ______________

I. Responde las siguientes preguntas sobre el texto leído.

1. Es cierto que Bob…

a) Tiene un delfín en su casa.
b) Tiene la foto de un delfín.
c) Quiere tener un delfín como mascota.
d) Le tiene miedo a los delfines.
e) No lo sé.

2. Bob explica que …

a) Los delfines viven solos.
b) Los delfines son todos iguales.
c) Los delfines tienen dientes grandes.
d) Los delfines necesitan dormir de noche.
e) No lo sé.

3. Según Bob….

a) Los delfines son menos inteligentes que las ballenas.
b) Los delfines no pueden nadar tan rápido.
c) Los delfines usan sonidos para comunicarse.
d) Todos los delfines son grises.
e) No lo sé.

4. La intención de Bob en el texto es:

a) Proteger a los delfines porque están en extinción.
b) Describir a un animal que le gusta mucho.
c) Alertar de los cuidados que deben tener los delfines.
d) Explicar las razones de por qué ha pedido un delfín como mascota.
e) No lo sé.

5. Bob explica que los delfines:

a) Comen algas.
b) Comen durante la noche.
c) Comen peces.
d) Son hervíboros.
e) No lo sé.

Appendix 33.

Sample dictation test

Dictation

Name: [redacted] Class: 5º B

1. My pants are ____________.
2. Sorry, I'm Pencil.
3. Be caerful, please.
4. I like ____________.
5. My brother is cleaver.
6. I love this costom.
7. The floor is creaky.
8. Two drops of milk.
9. The feary is generous.
10. My pet is floofy.
11. Let's go to the forest.
12. My pet is heary.
13. Don't forget your hands back.
14. The kitten is lovely.
15. Don't forget its lect.
16. I found a beautiful life.
17. Oh! I saw a woneyt.
18. You need a pendrate.
19. My pants have month.

20. Pecs are round.

21. I like your blucaR.

22. I am slufs with the results.

23. I didn't see the podle.

24. Add a sosach.

25. I can see a shel.

26. My sleepees are brown.

27. My hands are stiking.

28. My suit case is big.

29. This is a long treak.

30. Push the trolly.

31. That object is beatifull.

32. He has a want.

33. It is climbing the wear.

34. The corridor is white.

35. This animal has winds.

36. Don't sit on the woping chair.

Appendix 34.

Multiple-choice vocabulary test

Multiple choice

Name: ____. Class: ___________________

1. The ________________ are pink (zapatillas de levantarse).

a) Vazes
b) Sandals
c) Slippers
d) Cushions
e) No sé

2. My hands are _____________________ (pegajosas).

a) Fliff
b) Soft
c) Chestnut
d) Sticky
e) No sé

3. Be ___________________ (cuidadoso/cuidadosa).

a) Scary
b) Cleady
c) Careful
d) Evil
e) No sé

4. I love __________________ (arvejas).

a) Aubergines
b) Clurps
c) Celery
d) Peas
e) No sé

5. The prince needs a _____________________ (venda).

a) Cast
b) Bandage
c) Tawl
d) Rope
e) No sé

6. Shhhh! the floor is ____________________ (que cruje).

a) Bumpy
b) Creaky
c) Noisy
d) Feafy
e) No sé

7. Look at the ___________________ (hada).

a) Dween
b) Fairy
c) Witch
d) Godmother
e) No sé

8. Birds have ____________________ (alas).

a) Feathers
b) Wings
c) Beaks
d) Noods
e) No sé

9. Wait for me in the ________________________ (bosque).

a) Bush
b) Beeth
c) Forest
d) Crab
e) No sé

10. Your _____________________ is adorable (gatito/gatita).

a) Spake
b) Puppy
c) Kitten
d) Beetle
e) No sé

11. My sister is jumping on a _____________________ (charco).

a) Puddle
b) Plook
c) Hermit
d) Staith
e) No sé

12. The table is ______________________ (ancha).

a) Empty
b) Narrow
c) Wide
d) Crelp
e) No sé

13. Don't sit on the _______________________ chair (suelta).

a) Blath
b) Flat
c) Tight
d) Wobbly
e) No sé

14. I love my ______________________ cat (suave y peludo).

a) Hard
b) Clurve
c) Comfy
d) Fluffy
e) No sé

15. You have to wear a ______________ (disfraz).

a) Costume
b) Prelt
c) Skirt
d) Apron
e) No sé

16. There are _____________________ of water on the table (gotas).

a) Glunts
b) Clouds
c) Drops
d) Streams
e) No sé

17. My ________________ is full of food (carro).

a) Trolley
b) Croid
c) Basket
d) Race
e) No sé

18. My ______________________ is green (cartera, bolso).

a) Wallet
b) Handbag
c) Frimp
d) Tower
e) No sé

19. Use the ______________________ to walk the dog (correa para perros).

a) Lead
b) Bence
c) Necklace
d) Rope
e) No sé

20. The ______________________ is pretty (sirena).

a) Whale
b) Mermaid
c) Ghost
d) Dwaint
e) No sé

21. Technology is ______________________ (útil).

a) Smart
b) Useful
c) Fruft
d) Charming
e) No sé

22. Some men are ______________________ (peludos).

a) Hairy
b) veilful
c) Scary
d) Creepy
e) No sé

23. I can include this ______________________ in my collection (hoja de árbol).

a) Grass
b) Leaf
c) Rice
d) Flomb
e) No sé

24. I need my ______________________ to sleep tight (almohada).

a) Mattress
b) Skarm
c) Pillow
d) Tray
e) No sé

25. I like collecting ______________________ (conchas).

a) Badges
b) Shrimps
c) Shells
d) Blarves
e) No sé

26. I have to pack my ______________________ (maleta).

a) Suitcase
b) Purse
c) Antler
d) Fliffer
e) No sé

27. Harry Potter has a ______________________ (varita mágica).

a) Wizard
b) Brand
c) Cluss
d) Wand
e) No sé

28. This ______________________ is enormous! (telaraña).

a) Web
b) Dust
c) tralph
d) Creepy-crawly
e) No sé

29. My shoes are full of ______________________ (barro).

a) Whale
b) Mud
c) Clay
d) Brelp
e) No sé

30. We need ______________________ to make hot dogs (salchichas).

a) Beans
b) Eggs
c) Sausages
d) Jums
e) No sé

31. We need the ______________________ for the car race (pista).

a) Track
b) Wheel
c) Dipper
d) Blarb
e) No sé

32. I love ______________________ (Repollo).

a) Celery
b) Cabbage
c) Zarf
d) Coin
e) No sé

33. Call me later, I'm ______________________ (ocupado/ocupada).

a) Sleepy
b) Busy
c) Frain
d) Mean
e) No sé

34. My T-shirt is ________________________ (rayada).

a) Checked
b) Dween
c) Stripy
d) Dreadful
e) No sé

35. She's so ________________________ (inteligente).

a) Brave
b) Funny
c) Crelp
d) Clever
e) No sé

36. My parents are ________________________ (contentos, satisfechos).

a) Pleased
b) Handsome
c) Worried
d) Stoud
e) No sé

Appendix 35.

Informed consent

Barcelona, marzo 2021

Autorización

Investigadora: Daniela Avello García
Directora y supervisora del proyecto: Dra. Carme Muñoz Lahoz
Institución: Universitat de Barcelona

Estimados padres,

Mi nombre es Daniela Avello, investigadora predoctoral del área de inglés de la Universidad de Barcelona. Me dirijo a ustedes con el objetivo de solicitar su autorización para que su hijo/a participe en un set de actividades en inglés durante el primer trimestre de este año 2021. Estas actividades serán implementadas dentro de las clases de inglés y tienen por objetivo favorecer y fortalecer la exposición al idioma, ya que las investigaciones han demostrado consistentemente que este factor tiene un rol fundamental en el aprendizaje de lenguas extranjeras.

Es importante destacar que la información recolectada es estrictamente ***confidencial***, por lo que la revisión de esta será de ***exclusiva responsabilidad de la investigadora***. Igualmente, ***no*** se revelará información identificativa de los participantes bajo ningún concepto. Aunque se pide el nombre y apellidos de los alumnos/as para identificar las actividades, se les asignará un número aleatorio en la base de datos, y cualquier referencia a sus respuestas en futuras publicaciones se realizará siempre con este número ***anónimo***.

La participación del estudiante se puede confirmar, ya sea a través de la colilla incluida en este documento o el link a continuación: https://forms.gle/1CASwy6UwR6Hb1xr7

En caso de cualquier consulta, no duden en contactarme a daniela.avello@ub.edu

Muchas gracias por su colaboración.

Daniela Avello García

Yo, ____, apoderado/a de el/la estudiante ____, he leído la carta enviada por la investigadora predoctoral Daniela Avello y autorizo a mi hijo/a a participar de las actividades que se realizarán en las clases de inglés durante el primer trimestre del año 2021.

Firma: ______________________

Appendix 36.

Written-word form recall: Partial knowledge scale (PKS) scoring criteria

Two points were given to each 100% accurate response while one point was awarded to those responses that approached the target form as a result of previous encounters (pretest) or the treatment (posttest). The list of possible answers was specified by adapting the criteria used by Gesa (2019), which was, in turn, adapted from Muñoz (2006):

- The answers that appeared to have been influenced, in their entirety, by the oral prompts and the transparent sound-symbol correspondence patterns of Spanish were discarded (e.g. 'cleva' vs. clever or 'estraipy' vs. stripy).
- Only **one** spelling mistake per target word was accepted for two-syllable words. This was not the case for one-syllable words since there were fewer L2 orthographic patterns available to infer that the participant had encountered the target word before (sample two-syllable words: 'trolly', 'troley', pudle).
- If the answer resulted in a word that already exists in English and has a different meaning, the answer was awarded zero points (e.g. 'paddle' instead of puddle).
- The written representation had to approach the pronunciation of the target word.

The following spelling mistakes were tolerated:

a) Missing consonant or use of double consonants in a position where it was not required (e.g. as in 'carefull' or 'cabage').

b) Wrong vowel/diphthong or missing vowel (e.g. 'floffy', 'costum', 'sosage'; 'feiry').

d) Wrong cluster or digraph (e.g. 'cabbach' vs. cabbage).

e) The use of a vowel instead of the final graphemes 'y' or 'w' (e.g. fluffi).

The list of possible answers was as follows:

Stripy:	Stripi, stripe, strippy,strypy, stripey.
Busy:	Busi, bussy.
Careful:	Carefol, carefull, carful, carefoul.
Cabbage:	Cabage, cabbige, cabbach, cabbege.
Clever:	-

Costume:	Costum.
Creaky:	Creacky, creaki, creeky, creky.
Drop:	-
Fairy:	Feiry, fairi.
Fluffy:	Flufy, fluffi, fluffy, flaffy.
Forest:	Forrest.
Hairy:	Heiry, heary.
Handbag:	Handbug.
Kitten:	Kiten, kitteen.
Lead:	-
Leaf:	-
Mermaid:	Marmaid.
Bandage:	Bandige, bandach, bendage, bandege, bandige.
Mud:	-
Pea:	-
Pillow:	Pilow, pillou.
Pleased:	-
Puddle:	Pudle, puddl, poddle.
Sausage:	Sosage, sausach, susage, sausege, saussage, sousage.
Shell:	-
Slipper:	Sliper.
Sticky:	Stiky, sticki, stycky.
Suitcase:	Sutcase, siutcase.
Track:	-
Trolley:	Troley, trolly, trollee.
Useful:	Usefol, usefull, usiful, usful.
Wand:	-
Web:	-
Wide:	-
Wing:	-
Wobbly:	Wobly, wubbly, wobbli.

Appendix 37.

WWFMR: Time pairwise contrasts per fourth-grade viewing distribution group with vocabulary knowledge as covariate

Viewing distribution	Time Pairwise Contrasts	Contrast Estimate	Std. Error	t	df	Adj. Sig.	95% CI Lower	95% CI Upper
Once a week	Pretest - Posttest	-.256	.030	-8.503	1006	.000	-.328	-.184
	Pretest - Delayed	-.261	.032	-8.178	1369	.000	-.332	-.189
	Posttest - Delayed	-.005	.013	-.351	7367	.726	-.030	.021
Twice a week	Pretest - Posttest	-.194	.034	-5.646	713	.000	-.271	-.117
	Pretest - Delayed	-.171	.026	-6.702	551	.000	-.232	-.110
	Posttest - Delayed	.023	.014	1.632	6654	.103	-.005	.050
Three times a week	Pretest - Posttest	-.175	.019	-9.446	7367	.000	-.219	-.131
	Pretest - Delayed	-.155	.018	-8.433	3840	.000	-.196	-.113
	Posttest - Delayed	.020	.009	2.184	7367	.029	.002	.039
Four times a week	Pretest - Posttest	-.212	.032	-6.576	1662	.000	-.284	-.140
	Pretest - Delayed	-.188	.022	-8.466	1563	.000	-.242	-.135
	Posttest - Delayed	.023	.011	2.105	2902	.035	.002	.045

The sequential Bonferroni adjusted significance level is .05.
Confidence interval bounds are approximate.

Appendix 38.

WWFMR: The influence of viewing distribution in fourth and fifth graders with vocabulary knowledge as covariate

Source	F	df1	df2	Sig.
Corrected Model	66.840	7	193	.000
Vocabulary picture	159.022	1	56	.000
Viewing distribution	.048	1	66	.827
Level	.065	1	67	.799
Time	155.912	2	3419	.000
Level * Time	6.749	2	5162	.001

Probability distribution: Binomial
Link function: Logit
a. Target: Multiple choice

Appendix 39.

Online sessions