8-21-2023 2:00 PM

The Effects of Audiovisual Input on Second Language Learning: A Meta-Analysis

Dru M. Sutton, *Western University*

Supervisor: Webb, Stuart, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Arts degree in Education

Abstract

This meta-analysis investigates the contributions of viewing audiovisual input on second language (L2) learning. We calculated 75 effect sizes from 56 experiments (n = 1954). We assessed the effects of audiovisual input on language learning using a within-group (pre-post) meta-analytic approach. The extent to which fifteen moderator variables influenced results was assessed. Several methodologically and pedagogically relevant results were found. Results showed that a) there was a medium effect of audiovisual input on L2 learning (g = 1.01); b) no differences were found between the effects of viewing audiovisual input on different areas of L2 learning (vocabulary, grammar, pronunciation, speaking, listening proficiency); and c) video category had a significant impact on L2 learning with entertainment-focused videos (e.g., TV series, movies, and mixed videos) yielding lower effects than educational videos (e.g., TED Talks, documentaries, and language-focused). These findings along with future research directions for L2 learning through audiovisual input are discussed.

Keywords: second language learning, meta-analysis, audiovisual input, viewing, video

Summary for Lay Audience

In recent years there has been a large amount of research published that focuses on second language (L2) learning through viewing audiovisual input. These studies typically focus on various types of videos such as movies, TV series, and documentaries among others. Since there have been many studies focusing on learning through viewing audiovisual materials, there is a large amount of data that indicates the benefits of audiovisual input for L2 learning. However, these studies may not be directly compared due to varying methodological choices such as the type of L2 test used, the type of video used, participant demographics, and the area of L2 learning (e.g., vocabulary, grammar, pronunciation). Moreover, it is difficult to understand the extent to which viewing audiovisual materials contributes to L2 learning because there is a lot of variation in the results across studies.

In such cases, a meta-analysis can be conducted to directly compare the various studies. A meta-analysis standardizes the effect of each study so that they can be directly compared, and then pools these effects. This allows researchers to assess the overall impact of a treatment and allows them to assess the influence that certain variables may have on the outcome variable.

This study meta-analyzed research involving learning a L2 through viewing audiovisual input. In total, 56 studies that investigated L2 learning through viewing audiovisual input were analyzed. Methodological variables such as participants' first language, age, target L2 (e.g., grammar, pronunciation, vocabulary), and variables related to the types of tests taken were analyzed as moderator variables (i.e., variables which may have an effect on the learning outcome) .

Overall findings of the meta-analysis showed that viewing audiovisual materials has a medium effect on L2 learning. Results also showed that audiovisual input does not benefit one

area of L2 learning significantly more than others. These findings are important for real-world use as they suggest that not only can watching videos help second language learners, but that several areas of second language learning can benefit from watching video.

Acknowledgements

First, I would like to express gratitude for my supervisor, Dr. Stuart Webb, who has helped me grow as a researcher. Without his expertise, encouragement, and constructive feedback this project would not have been possible. I would also like to extend my gratitude to Dr. Frank Boers who offered his support and knowledge as if I were his student.

I am further grateful to Takumi Uchihara, Zhouhan Jin, Akifumi Yanagisawa, and Su Kyung Kim for answering my endless questions as I learned how to approach my project.

I also wanted to thank my friends and fellow students at Western. Thank you, Alyssa, Ashley, Injung, and Michelle.

Finally, I wanted to thank those closest to me who made life just a little bit easier. From the bottom of my heart, I thank my dad, Heather, Heejin, Derek, and Nick. Without your support I would not have made it this far.

Table of Contents

List of Tables

Table of Appendices

Chapter 1: Introduction

1.0 Introduction

This chapter will provide a brief introduction of the background, goal, and rationale for the present thesis.

1.1 Background

Audiovisual input has been thoroughly studied in recent years (Montero Perez, 2022; Vanderplank, 2010), with research covering various areas of second language (L2) learning (e.g., vocabulary, grammar, speaking). The increased popularity of viewing audiovisual input as a form of L2 learning can largely be attributed to the ease of access to audiovisual materials via the internet.

Audiovisual input is a somewhat unique language learning tool, as audiovisual input encompasses a wide range of materials. Studies may utilize educational videos (e.g., Cintrón-Valentín & García-Amaya, 2021) or videos used for entertainment purposes such as TV series (e.g., Birulés-Muntané & Soto-Faraco, 2016), movies (Sadowska, 2015), and documentaries (Peters & Webb, 2018), among others. While language-focused videos explicitly teach language to learners, there is also a large number of studies that focus on "native" audiovisual input since it is representative of real-life conversation (Quaglio, 2010) and may therefore provide valuable input to learners. The large library of available content also means that there is practically a limitless number of options for learners to choose from, meaning that they can increase motivation by selecting videos that appeal to them.

1.2 Purpose and Rationale

The present study seeks to take advantage of the large number of existing studies on audiovisual input by conducting a meta-analysis. There is a large body of research that investigates audiovisual input on L2 language learning, but to the best of my knowledge this data has yet to be meta-analyzed. This study therefore aggregates data from studies that look at L2 learning across various areas of L2 learning to identify the contributions of viewing audiovisual material on L2 learning.

Although data generally suggests that viewing audiovisual input has a positive effect on L2 learning (Montero Perez, 2022), there are a few reasons why this research is warranted. First, the present study investigates moderator variables that have not been, or that cannot be, investigated in the existing literature. Among these moderator variables are treatment-related, learner-related, and test-related variables. These may not be compared directly in individual studies for various reasons, such as practicality or time. In other cases, they may be investigated, but with small sample sizes.

The second reason is that meta-analyses allow researchers to interpret data from a large number of studies in ways that literature reviews do not. For example, meta-analyses show whether effect sizes are consistent across studies, or whether there is variation among studies (Borenstein et al., 2009). Where variation does exist among studies, a moderator (or subgroup) analysis can be conducted to investigate the different moderating effects certain variables have on a treatment (Borenstein et al., 2009; Harrer et al., 2021). Where there is low heterogeneity between studies (i.e., effect sizes across studies are consistent), a meta-analysis allows researchers to establish the extent to which a treatment affects outcomes. In the context of this

study, a meta-analysis allows us to assess the impact of audiovisual input on L2 language learning, and to identify which moderators may help to contribute to ideal learning conditions.

The third reason is that there has yet to be a meta-analysis conducted that focuses on audiovisual input with no on-screen text. Previous meta-analyses may investigate audiovisual input to some extent (de Vos et al., 2018; Montero Perez et al., 2013; Reynolds et al., 2022), but they do not allow us to determine the effect it has on L2 learning by itself. For example, Montero Perez et al. (2013) and Reynolds et al. (2022) meta-analyze studies using captioning for vocabulary learning, rather than unenhanced audiovisual input. Neither of these studies provide insight into the effectiveness of audiovisual input without captions. De Vos et al. (2018) meta-analyze the contributions of various forms of spoken input on incidental vocabulary learning. However, while this study looks at unenhanced audiovisual input, it provides no effect size for audiovisual input. Moreover, all three of these studies look at vocabulary learning only. The present study seeks to investigate the effects of audiovisual input across all areas of L2 learning possible.

The final rationale for this study is that audiovisual input is widely consumed by L2 learners (Montero Perez, 2022; Peters, 2018). Given its wide consumption and the importance of input on L2 proficiency (Lindgren & Muñoz, 2013), determining the true effect of audiovisual input and moderator variables on L2 learning allows both researchers and teachers to make suggestions on how to properly utilize audiovisual input as a learning tool.

1.3 Theoretical Framework

1.3.1 Multimedia Learning Theory

The Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2014) is the primary theory referenced when explaining the benefits of language learning through viewing audiovisual input. CTML is built on the hypothesis that audio in conjunction with visuals encourages more meaningful learning than either single component presented independently. CTML was initially hypothesized to explain general learning through audiovisual input, but has since been explicitly linked to L2 acquisition (Plass & Jones, 2005). Nonetheless, there are few studies that explicitly investigate the effectiveness of audio versus audiovisual input. There has, however, been increased investigation into the differential effects in recent years, with favorable findings (Dang et al., 2022; Feng & Webb, 2020; Montero Perez, 2022). However, de Vos et al. (2018) conducted a meta-regression on the effects of spoken input on vocabulary acquisition and found that there was no significant learning difference between audio and audiovisual input. Therefore, it is somewhat unclear if the effects of viewing audiovisual input are significantly greater than audio alone in L2 learning. This means definitive evidence of the benefits of viewing audiovisual input over audio alone has not been empirically proven over a large number of studies.

1.3.2 Dual Processing Channels

Cognitive load theory argues that memory has a limited capacity and duration (Baddeley, 1986; see also, Sweller et al., 2011). If working memory reaches its capacity limit, then working memory can no longer be allocated to learning (Chen et al., 2009), and therefore, cognitive load needs to be properly managed to create more ideal learning outcomes. Low and Sweller's (2014) modality principle states that limited capacity can be addressed by distributing incoming information into different processing channels, thereby reducing the overall burden on a single channel. According to this principle, each channel can process a different type of information,

which reduces the overall burden on the cognitive systems. When the overall burden on cognitive systems is reduced, more working memory can be allocated toward learning, which explains why multimodal input may be considered more effective than unimodal input in L2 learning.

There are two theories which explain the assumption of dual channels in the cognitive process. The first is Paivio's (1990) dual-coding theory. This theoretical approach argues that there are two coding systems that activate when processing input: a verbal system, and a non-verbal system. The verbal system processes any kind of verbal input, be it spoken words or written words. Therefore, in the context of audiovisual input the verbal system would process speech and subtitles. The non-verbal system processes any non-verbal input, which includes animation, gestures, sound effects, or other imagery. Although Baddeley's (1990) model of working memory also assumes there are dual processing channels, he disagrees with what each system processes. Baddeley suggests that rather than the systems making a distinction between verbal and non-verbal content, they make a distinction between visual input and auditory input. Despite the disagreement in what each system processes, however, both argue in favor of the idea that the cognitive process benefits from having two types of inputs. Consequently, both argue that audiovisual input creates ideal conditions for learning.

1.4 Thesis Organization

This thesis contains three sections. The first chapter provides background information and the justification for the study. Chapter two is the integrated article which presents a literature review, theoretical framework, methodology, results, and a discussion. The third chapter summarizes the findings and limitations and suggests future directions.

Chapter 2: Article

Abstract

This meta-analysis investigates the contributions of viewing audiovisual input on second language (L2) learning. We calculated 75 effect sizes from 56 experiments (n = 1954). We assessed the effects of audiovisual input on language learning using a within-group (pre-post) meta-analytic approach. The extent to which fifteen moderator variables influenced results was assessed. Several methodologically and pedagogically relevant results were found. Results showed that a) there was a medium effect of audiovisual input on L2 learning (g = 1.01); b) no differences were found between the effects of viewing audiovisual input on different areas of L2 learning (vocabulary, grammar, pronunciation, speaking, listening proficiency); and c) video category had a significant impact on L2 learning with entertainment-focused videos (e.g., TV series, movies, and mixed videos) yielding lower effects than educational videos (e.g., TED Talks, documentaries, and language-focused). These findings along with future research directions for L2 learning through audiovisual input are discussed.

2.1 Introduction

Audiovisual input is a form of multimodal input that has been increasingly studied in applied linguistics (Montero Perez, 2022). One reason for the increase in research related to audiovisual input is the wide availability of foreign language videos. In the 1980s, when research on audiovisual input began gaining attention (e.g., Baetens Beardsmore & van Beeck, 1984), accessing video materials required physical media like VHS tapes. In more recent times, however, there has been a huge increase in video streaming platforms such as Netflix, YouTube, Apple TV, and Disney+ that allow for the instant viewing of videos around the world (Montero

Perez, 2022). Wide availability of audiovisual input gives learners easy access to engaging content that they can use as a language learning tool.

The support provided for learning by both aural content and imagery may be more effective for learning than either individual component (Mayer, 2014). Research suggests that this is the case among various areas of second language (L2) learning (e.g., grammar, Cintrón-Valentín & García-Amaya, 2021; vocabulary, Peters & Webb, 2018; pronunciation, Wisniewska & Mora, 2020; and listening, Weyers, 1999). Across studies, the findings consistently show favorable outcomes from learning with audiovisual input. However, many of these studies do not focus on audiovisual input by itself but rather enhanced audiovisual input, which most often contains on-screen text (captions).

On-screen text may act as an additional scaffold to help learners comprehend audiovisual input by displaying the textual form of the audio track (Montero Perez et al., 2018). On-screen text has been extensively studied in the past few years (e.g., Aldukhayel, 2021; Peters, 2019). In fact, there is a larger body of research on captioning than studies that focus solely on audiovisual input (Montero Perez, 2022), with audiovisual input often acting as the control condition. In a general sense, audiovisual input has been disregarded in favor of captioned audiovisual input. Nonetheless, studies have found that not all learners may use captions (Kam et al., 2020). Furthermore, captions may be detrimental to some areas of L2 learning, whereas audiovisual input without on-screen text may be more beneficial (e.g., Wisniewska & Mora, 2020).

Given the large amount of research on audiovisual materials, this study sought to synthesize the existing research to determine the effects of audiovisual input without on-screen text on L2 learning. A synthesis of the existing literature may help to clarify the degree to which

viewing audiovisual input contributes to different areas of L2 learning and further encourage its use as a language learning tool. While a previously conducted meta-analysis by Montero Perez et al. (2013) examined the effects of captioning on vocabulary learning and content comprehension, to the best of our knowledge this is the first meta-analysis conducted to assess the effects of audiovisual input alone. Consequently, this study meta-analyzed data from the existing audiovisual and caption-focused studies (extracting audiovisual-only group data) to provide information about the effects of audiovisual input across different areas of L2 learning (i.e., vocabulary, grammar, pronunciation, speaking, listening proficiency).

2.2 Background

Audiovisual input is generally described as being audio presented alongside either dynamic or static imagery. In applied linguistics research, "audiovisual input" is often synonymously used to mean "video" (i.e., dynamic or moving imagery presented with audio) and most often investigates the effects of TV shows or movies on L2 learning (e.g., Peters et al., 2016; Puimège & Peters, 2019). Since most studies on audiovisual input focus on video, and because audiovisual input may have more than one interpretation, moving forward this study defines audiovisual input as video. This means that films or movies, TV shows, online videos, recorded lectures, or animations are all considered within the scope of this study. Consequently, any type of input that consists of still imagery (e.g., drawings) or where the type of imagery is unclear (e.g., PowerPoint presentations where the type of imagery could be either still or dynamic) were not considered within the scope of this study unless it was explicitly stated that dynamic imagery was used.

Vocabulary learning is the most researched area of L2 learning through viewing audiovisual input (e.g., Puimège & Peters, 2019; Sydorenko, 2010). Although there are generally

favorable findings, there is some variation in the extent to which new items were learned. For example, Peters and Webb (2018) found that participants who watched an L2 documentary learned 19.12% of unknown words on a meaning recall test. A meaning recognition test was also administered, yielding similar results but with a smaller relative gain of just 11.29%. Meanwhile, Peters (2019) investigated the effects of viewing an English documentary on L1 Dutch speaker's form recognition and meaning recall of vocabulary and found less favorable results. While the meaning recall test findings were in line with Peters and Webb's (2018) findings (12.36%), a form recognition test only revealed gains of 3.36%.

The contributions of audiovisual input to L2 pronunciation have also been investigated (e.g., Mohsen & Mahdi, 2021; Wisniewska & Mora, 2020). Wisniewska and Mora (2020) studied the effects of L2 audiovisual input on English phoneme discrimination and accentedness of native L1 Spanish/Catalan speakers. Two groups with no captions were treated, each of which had a different pre-treatment lesson. Wisniewska and Mora found that in a phoneme discrimination task the first group of participants had relative gains of 10.84% from the pre-test to the post-test. These gains indicate that the participants' receptive knowledge of English pronunciation improved. However, the second group of participants had negative gains (-5.22%) indicating that viewing the audiovisual input was not beneficial to their ability to discern English phonemes. An accentedness task determined the effects of audiovisual input on productive knowledge of L2 pronunciation. Gains in the accent rating task were negative for both groups (-6.21%, -1.33%), meaning that neither group's productive knowledge of L2 pronunciation benefited from viewing the audiovisual input. In another study, Mohsen and Mahdi (2021) investigated the effects of viewing L2 video on Arabic speakers' pronunciation of 40 polysyllabic English words. They found that participants who viewed L2 videos without captions

improved in their production of English phonemes, allophones, and consonant clusters. The relative gains on the immediate post-test were 38.74%. The score on a three-week delayed post-test, however, was similar to that of the pre-test, with gains of only 2.20%.

Research has also examined the degree to which listening proficiency may improve through viewing audiovisual input. Weyers (1999) found that American university students exposed to Spanish audiovisual input over an 8-week period had a relative gain of 46.13% on a listening proficiency test. Students who were not exposed to audiovisual input had gains of 24.14%. Terrell (1993) investigated the effects of watching L2 video on L1 English speakers' Spanish listening proficiency. In this study three groups were investigated: students who were not Spanish majors but enrolled in a Spanish course, Spanish majors, and native Spanish speakers. Terrell found an increase in listening proficiency outcomes for both the Spanish majors (28.52%) and for students who were not Spanish majors (66.94%). Overall, in comparison to other areas of language learning, listening skills appear to have very high learning gains.

Studies that investigated the effects of viewing audiovisual input on L2 speaking found more conservative results than those that looked at listening. According to the transfer-appropriate processing (TAP) theory (Morris et al., 1977), if a test is similar to the treatment, it will reveal more learning than a test that is less similar to the treatment. Therefore, since consuming audiovisual input is a receptive activity, one might expect that a receptive form of learning (i.e., listening) would be more easily acquired than a productive one (i.e., speaking). Nonetheless, research reveals that viewing audiovisual input also has positive effects on L2 speaking (Charles & Trenkic, 2015; Weyers, 1999). However, this is not consistent across every study (e.g., negative learning gains of -1.46% were found by Tsai, 2010).

A final area of research has been to examine the effects of audiovisual input on L2 grammar acquisition (e.g., Cintrón-Valentín & García-Amaya, 2021; García–Amaya & Cintrón–Valentín, 2021; Muñoz et al., 2021). Muñoz et al. (2021) found an increase of 18.86% in L2 Spanish grammar knowledge after five weeks of watching an American television show. Utilizing the same audiovisual material, Pattemore and Muñoz (2022) found gains from the pre-test to the immediate post-test of 36.82% for L2 grammar acquisition. Relative gains from pre-test to a one-week delayed post-test were slightly lower at 33.76%. Cintrón-Valentín and García-Amaya (2021) exposed participants to a grammar-focused video. They found positive relative gains from the pre-test to post-test across four areas of L2 Spanish grammar: preterite and imperfect contrast (25.38%), gustar-type verbs (83.10%), the subjunctive in noun clauses (73.91%), and the conditional mood (86.45%).

Together, studies examining the contributions of viewing video on L2 learning reveal varying degrees of gains in knowledge within and across areas. The present meta-analysis aims to clarify the extent to which different aspects of L2 knowledge may be learned through viewing audiovisual input.

2.2.1 Review of Moderator Variables

Age

The majority of studies examining audiovisual input and language learning involve university-aged adults (e.g., Mohsen & Mahdi, 2021; Peters & Webb, 2018). To the best of our knowledge, there are no studies which directly compare the differential effects of audiovisual input on learners of different age groups.

Language Proficiency

Studies typically focus on the intermediate level (e.g., Peters & Webb, 2018); however, there are also studies that involve beginners (e.g., Muñoz et al., 2021) and advanced learners (Wisniewska & Mora, 2020). This meta-analysis may allow for a better understanding of how proficiency might be related to the contributions of audiovisual input on L2 learning.

Learning Target

Vocabulary (both single and multiword items) is perhaps the most investigated area of L2 learning (e.g., Peters & Webb, 2018; Puimège & Peters, 2019; Sydorenko, 2010). However, there has been an increase in the number of studies examining how audiovisual input contributes to L2 learning in other areas. Grammar (e.g., Cintrón-Valentín & García-Amaya, 2021; Muñoz et al., 2021), L2 pronunciation (Mohsen & Mahdi, 2021; Wisniewska & Mora, 2020), speaking (Charles & Trenkic, 2015; Khosh Ayand & Shafiee, 2016), and listening proficiency (Shen, 1991; Weyers, 1999) are all areas of learning through audiovisual input that have been investigated.

Multiple Versus Single Viewing Sessions

When investigating the effects of audiovisual exposure, studies may either investigate the effects of viewing within a single session (e.g., Dang et al., 2022; Puimège & Peters, 2019), or the effects of viewing across multiple sessions (e.g., Charles & Trenkic, 2015; Weyers, 1999). Studies that treat participants over multiple sessions vary considerably in the number of sessions and period of time that treatments occur. Pattemore and Muñoz (2020) examined the effects of viewing audiovisual input over 8 weeks, whereas Muñoz et al. (2021) investigated learning over 8 months. Investigating single and multiple sessions may provide some insight into whether the

amount of exposure to L2 input (Webb & Chang, 2015) and spacing of input (Kim & Webb, 2022) affect L2 learning through viewing audiovisual input.

Language- and Non-Language-Focused Audiovisual Input

There are, broadly, two types of audiovisual input investigated across studies: language-focused (e.g., Cintrón-Valentín & García-Amaya, 2021; Weyers, 1999) and non-language-focused audiovisual input (e.g., Aldera & Mohsen, 2013; Puimège & Peters, 2019; Wisniewska & Mora, 2020). Videos such as movies and television programs are examples of non-language-focused audiovisual input that is typically created for the purpose of information or entertainment. In contrast, educators may use audiovisual input that has a pedagogical focus aimed at explicitly helping viewers acquire some aspect of a second language. This distinction is useful since language-focused videos are more likely to be utilized in the classroom than at home, whereas non-language-focused videos are often consumed by learners outside of the classroom (Peters, 2018). Language-focused videos often control the language present in the video (e.g., Cintrón-Valentín & García-Amaya, 2021), while non-language-focused videos may be more representative of speech used by L1 speakers of the target language (Montero Perez, 2022).

Viewing Time

Low and Sweller (2014) suggest that studies utilizing longer audiovisual materials tend to be ineffective. Since working memory utilizes the cognitive system's dual processing channels (Mayer & Moreno, 1998) and has limited capacity, a longer video may overload working memory impairing learning (Low & Sweller, 2014). However, Webb and Chang (2015) suggest

that as exposure to L2 input increases, so do encounters with target L2 forms and this in turn may lead to increased learning gains.

Test Type

The type of test used to measure learning through audiovisual input may affect the measured effect sizes. Tests are often categorized as receptive which involve understanding the target language (e.g., Dang et al., 2022; Shen, 1991; Weyers, 1999) or productive which involve using the target language (e.g., Cintrón-Valentín & García-Amaya, 2021; Majuddin, 2020; Nguyen & Boers, 2019). Transfer Appropriate Processing Theory (TAP: Morris et al., 1977) proposes that congruency between learning conditions and measurement formats positively affects learning. TAP suggests that because viewing audiovisual input is a receptive learning activity, receptive tests might reveal greater learning gains than productive tests. Tests may also be categorized as aural/oral or written tests. Input-modality – test-modality congruency, which builds upon TAP, states that learning outcomes are better measured when the modality of the input matches that of the test (Jelani & Boers, 2020). Therefore, it is expected that aural/oral tests (e.g., Wisniewska & Mora, 2020) may reveal greater learning gains than written tests (e.g., Majuddin, 2020).

2.3 The Present Study

Research Questions

1. To what extent does viewing unenhanced audiovisual input (i.e., video without on-screen text) contribute to L2 learning?

2. Which empirical variables (learning target, single vs. extended viewing, age, language proficiency, video type, video length, and test type) moderate the effects of viewing audiovisual input?

2.4 Methodology

2.4.1 Literature Search

First, an intensive search of several databases was conducted to identify all relevant literature. The databases searched were: PsychINFO, Education Research Information Center (ERIC), Linguistics and Language Behavior Abstracts, and Google Scholar. In addition, Google Scholar and ProQuest Global Dissertations were also searched for unpublished research to minimize the effects of publication bias (Rosenthal, 1979). Keywords used in searches included the following terms: *audiovisual input, television, movie, video, second language learning, second language, language learning, L2, listening, pronunciation, vocabulary, grammar, speaking, collocations, subtitle, caption.* Results of each search were exported as .ris files and imported into Covidence (*Covidence Systematic Review Software*, n.d.), which allowed for the screening of literature. Covidence also allowed for the automatic removal of duplicate results. The first stage of screening included studies based on relevant titles and abstracts. After this, studies were included or excluded based on full-text screenings. Further studies were identified through a reference search of all studies that were identified as fulfilling the inclusion criteria.

2.4.2 Inclusion Criteria

For inclusion into the meta-analysis, the following criteria must have been met:

1. The study involved exposure to unenhanced audiovisual input for the purpose of measuring at least one area of L2 learning (e.g., pronunciation, grammar acquisition,

vocabulary, speaking). Audiovisual input did not need to be the experimental condition. Studies such as Montero Perez et al. (2018), which focused on the effects of captioning were included if they included an unenhanced audiovisual group as a control condition.

2. Studies with within-group and between-group designs were included. If the study used a between-group design, only data from the audiovisual input group was used.
3. Studies must measure gains through a pre-test and a post-test. If a study did not include a pre-test or post-test, it was excluded. Following other meta-analyses (e.g., Uchihara et al., 2019), if more than one post-treatment test was conducted then the scores of the tests were averaged if they were conducted at the same time. In cases where post-tests were conducted at different intervals, only data from the first was included in the analysis unless there were additional treatment sessions between each post-test (e.g., Charles & Trenkic, 2015), in which case the post-test scores were averaged.
4. Neither the target language nor first language needed to be English. For example, Cintrón-Valentín & García-Amaya (2021), where participants' L1 was English and L2 was Spanish, was included along with Pattemore and Montero Perez (2022), where participant's L1 was Dutch and the target language was Spanish.
5. Studies that investigated content comprehension were not included (e.g., Rodgers & Webb, 2017), because content comprehension is the application of L2 skills rather than the learning of a specific skill.
6. Two-way interactions (e.g., Zoom calls or interactive videos) were not included, because the foci of these media are interactive components rather than audiovisual input.

7. If a study included multiple areas of L2 learning (e.g., Wisniewska & Mora, 2020, which measured listening and pronunciation) only one area was used in the analysis to avoid dependency between the two effect sizes.
8. The study provided enough statistical data to calculate an effect size (e.g., mean, standard deviation, and number of participants for pre- and post-tests).
9. The study was written in English.

Apart from studies excluded during the literature search according to the abovementioned criteria, several studies were excluded from the meta-analysis for practical reasons. The first to be excluded were studies which included the same participants but different areas of language learning. For example, Pattemore & Montero Perez (2022) and Montero Perez et al. (2022) used the same participants but focused on different areas of language learning. Although the general approach for this study is to average ESs that use the same participants, each of these studies looked at two under-investigated areas of language learning (listening: Montero Perez et al., 2022; and grammar: Pattemore & Montero Perez, 2022), which means that valuable data would be lost by averaging these two ESs. Therefore, the decision was made to exclude the study that examined the more well-researched area of language learning among the two. Montero Perez et al. (2022) was excluded from analysis. Similarly, Wisniewska & Mora (2020) looked at both speaking (shadowing) and pronunciation. Since pronunciation was less investigated, a decision to exclude data from the shadowing test was made. Dang et al. (2022b) was excluded for the same reason. Data from this study included data from Dang et al. (2022a), but in addition to investigating collocation learning, learning of individual vocabulary items was also investigated. Since the studies used the same participants and the sample of studies looking at collocation

learning was smaller, the first study (Dang et al., 2022a) was included. One additional study was excluded due to its outlying ES. More information on this study can be found in Appendix B.

2.4.3 Coding

Studies included in the meta-analysis were coded for dependent and moderator variables. Many studies included a battery of tests to measure the dependent variable (e.g., Aldera & Mohsen, 2013; Feng & Webb, 2020). In these cases, an ES was calculated for each test, then all ESs from a group were averaged so that each group of participants contributed only one ES to the overall model. Averaging the tests eliminates the dependency among effect sizes which occurs when testing the same participants with different measures (Hedges, 2019).

Moderator variables were coded as follows: learner-related variables (L1, L2, age, language proficiency), methodological variables (learning target, video type, language-focused video, video length, viewing plus, number of sessions, receptive or productive test, and written or aural test), and study-related variables (publication type and region). Procedures for coding study-related moderators and the results for these variables can be found in Appendix C. Following Li (2016) all subgroups with a sample size smaller than $k < 3$ were excluded so that contrasts could be more meaningfully made. Furthermore, all studies that did not report data required for inclusion in a given moderator analysis were excluded from that analysis.

Age

Age of participants was coded into two broad categories: adult and young learners. This meant that while there was a large sample size for adult learners ($k = 69$), the sample sizes for secondary ($k = 4$) and younger learners ($k = 2$) produced results that may not be meaningful. Aggregating secondary and younger learners helped to create a more meaningful comparison.

Therefore, adult learners were those that were in university or older, while young learners were those that were in secondary school or younger.

Language Proficiency

Language proficiency was coded into four categories based on the language level reported by authors: beginner, intermediate, intermediate-advanced, and advanced. These levels corresponded with CEFR proficiency levels. Intermediate-advanced participants were reported to be B2 to C1. Not all studies reported the proficiency of their participants (e.g., Weyers, 1999), and others (e.g., Feng & Webb, 2020) used metrics such as prior vocabulary knowledge that does not directly measure proficiency. These studies were therefore excluded from this moderator analysis.

First Language

L1 was coded based on the study's reporting of participants. The L1 was only coded if all participants were L1 speakers of the same language. If any participant of the group spoke another L1, then the study was coded as "Multi" to represent a mixed group of L1s. Several studies did not report the first language of participants, and rather than making assumptions based on the other features of the study (e.g., region, institution participants attended, etc.) these studies were simply excluded from this moderator analysis.

Target Language

Among seventy-five effect sizes, seventy-one of these targeted English learning while four targeted other L2s. Since the data was greatly skewed toward English, two groups were coded: English and non-English.

Learning Target

Several L2 learning targets were coded. Single-word items (vocabulary in which the target construction consisted of a single word), multi-word items (vocabulary in which the target construction consisted of more than one word), grammar, pronunciation (when the learning target required explicit focus on pronunciation-related features, such as morpheme discrimination), listening (which was characterized more broadly than pronunciation, where the overarching meaning is the focus as opposed to phonological details), and speaking (where participants needed to produce a cohesive thought verbally). Some studies mixed single- and multi-word items and were coded as mixed.

Number of Sessions

Studies were coded based on whether the audiovisual input was viewed in a single session or multiple sessions. Coding of this variable was straightforward: if a study involved treatment at one point in time, it was coded as a single session study; if a study involved treatment over several sessions, it was coded as having multiple sessions.

Language-Focused Versus Non-Language Focused

The videos used as materials were sorted into one of two categories. Language-focused videos were those that were focused specifically on teaching an aspect of language. Studies were coded as either “y” for being language-focused videos and “n” for being non-language-focused videos. Where a video was educational in nature, but did not focus on language learning (e.g., a lecture) it was considered to be non-language focused.

Audiovisual Type

Several broad categories of audiovisual input were also identified and coded. Animation included any type of animated video. Any video that focused on informing about a specific topic with a non-language focus was coded as documentary. Longer, standalone videos that served primarily entertainment purposes, regardless of the topic, were categorized as movies. Television serials or other videos that only represented part of an overarching story or theme were coded as TV Series. Videos that were recordings of non-language focused academic lectures were coded as lectures. Due to the number of studies investigating language learning from viewing TED Talks, the decision was made to code them as their own category. Language-focused videos were also coded as their own category. Two further categories (news and various clips) were coded but were excluded from the moderator analysis due to small sample sizes.

Viewing Time

Total viewing time was coded as: short (0 to 10 minutes), medium (11 to 30 minutes), and long (31 minutes or longer). The categories are representative of the total time spent viewing audiovisual input, rather than the average length of each viewing session.

Viewing Plus

Studies were categorized according to viewing behavior. Viewing studies were those in which participants viewed audiovisual input from start to finish without additional activities or manipulation of the video. Viewing plus studies were characterized as studies which included activities related to the audiovisual input, an audiovisual-based curriculum, control over the audiovisual input (e.g., allowed for pausing, rewinding, replaying, etc.).

Written vs. Aural/Oral Tests

Tests included in this analysis were divided into two different categories: written or aural/oral. These categories represented the modality through which tests were administered. Written tests were tests which solely relied on written cues and responses, whereas aural/oral tests required either the participant to respond to auditory input or that the participant responded orally to either written or auditory input. Studies that used mixed tests (i.e., they used each type of modality at different points) and did not report scores for sections of each test were excluded from this analysis. Since this moderator variable is concerned with the differences between individual test types, this subgroup analysis was conducted using the effect sizes of every individual test, rather than averaging the results of multiple tests.

Productive vs. Receptive Tests

Tests were coded as either productive or receptive. Tests coded as productive were those in which participants had to produce L2 forms (e.g., fill-in-the-blank tests). Receptive tests were those where participants were required to recognize an L2 form or meaning (e.g., multiple-choice tests). Tests were coded as mixed if they used both receptive and productive formats in a single test and were excluded from this analysis. Like the previous moderator variable, this moderator analysis was concerned with each individual test type and therefore did not average effect sizes across tests.

2.4.4 Reliability of Coding

In order to test the reliability of the coding procedure, 10 studies (18% of the primary studies and 23% of the total effect sizes) were selected at random and coded by a second rater who had previously conducted meta-analyses on L2 learning. The percentage of agreement between the two raters was initially 94%. After discussion, all disagreements in coding were resolved and the remaining studies were checked to ensure consistency.

2.4.5 Data Analysis

Pooling of effect sizes and moderator analyses were conducted using the R packages meta (Schwarzer, 2023), metafor (Viechtbauer, 2022), and dmetar (Harrer et al., 2019). The meta package requires pre-calculated ESs when running analyses with within-group data. For this reason, ESs and standard errors for each study were first calculated in Excel (*Microsoft Excel*, 2018).

A within-group (or pre-post contrast) approach was taken because many studies that have an audiovisual condition primarily investigated L2 learning through captions (e.g., Pattemore & Muñoz, 2020) or glossing (e.g., Montero Perez et al., 2018) and therefore audiovisual input without on-screen text often served as the control condition, rather than the treatment. Since there are few studies that look at audiovisual input with no on-screen text as the experimental condition, between-group ESs could not be calculated for the majority of studies. Nonetheless, there was still a large number of studies that used an audiovisual treatment group, and the within-group approach was able to garner a large sample of primary studies.

2.4.6 Calculating Effect Size

This study used Hedges' *g* as the measure for effect size. Hedges' *g* is a correction factor for the widely used effect size Cohen's *d*, which helps correct for an overestimation of Cohen's d for studies with small sample sizes (Borenstein & Hedges, 2019). The formulae for calculating effect sizes, and standard errors (which are needed for calculating the weight of studies) can be found in Appendix A. Since most studies do not report pre-post correlation, the average of studies that did report it was used as recommended by Borenstein et al. (2009). A sensitivity analysis was conducted to test the robustness of the borrowed correlation. The sensitivity

analysis showed that the borrowed correlation did not affect the overall pooled effect to a large degree (see Appendix B for more information).

Since within-group meta-analyses tend to overestimate the size of the effect (Cumming, 2012), Plonsky and Oswald (2014) suggest that effect sizes are interpreted as $d = 0.60$ being small, $d = 1.00$ being medium, and $d = 1.40$ being a large effect. Therefore, these thresholds were used for interpreting the results.

2.4.7 Random- and Mixed-Effects Approach

In order to address the first research question, a random-effects approach was chosen. A random-effects model assumes that there is not a single effect size of the overall population, but rather, there is a subset of effect sizes (Harrer et al., 2021; Plonsky & Oswald, 2014). Borenstein et al. (2009) argue this approach is good for meta-analyses in educational fields because populations vary considerably (e.g., in age, education, L1, L2). Furthermore, since between-study heterogeneity was expected to be high due to audiovisual input being the treatment in some studies and the control in others, as well as the aforementioned variations in population, a random-effects model was deemed most appropriate. The DerSimonian-Laird estimator (DerSimonian & Laird, 1986) was used to calculate heterogeneity variance, tau-squared (τ^2). Furthermore, the Knapp-Hartung adjustment (Knapp & Hartung, 2003) was applied because Borenstein et al. (2009) argue a random-effects model should not be compared against a *Z*-distribution (i.e., a normal distribution), but rather a *t*-distribution.

To address the second research question, a moderator (or subgroup) analysis was conducted using a mixed-effects approach. In a mixed-effects model it is assumed that each subgroup has a true effect size because they should have been taken from the same population (i.e., the population shares a unifying characteristic). Meanwhile, each subgroup contributes to

an array of effect sizes which defines the overall effectiveness of a treatment (Harrer et al., 2021).

2.5 Results

The analyses revealed a significant effect of audiovisual input on language learning, g = 1.01, 95% CI [0.77, 1.26], which is a medium effect following the benchmarks proposed by Plonsky and Oswald (2014). This indicates that audiovisual input has an overall moderate effect on language learning. High heterogeneity among included studies was expected given the differing populations and research designs. Since there was a large number of effect sizes included in the meta-analysis (k = 75), heterogeneity of the pooled effect was measured with I^2, as Q-tests are susceptible to bias with a large sample size and should not be relied upon as the sole test of heterogeneity in such cases (Harrer et al., 2021). The I^2 value was 0.946 (CI 95% [.938; .953]). This value suggests that 94.6% of heterogeneity was due to variation in the true effect of studies, rather than sampling error, which means that studies vary considerably in their true effects. Since I^2 is calculated using Q, it is still susceptible to some of the same issues. In particular, study precision influences I^2 to a high degree (Harrer et al., 2021). Nonetheless, a high I^2 value was expected since there are various moderating variables amongst the included studies. In such instances of high heterogeneity, moderator analyses are appropriate (Borenstein et al., 2009; Harrer et al., 2021).

The second research question was concerned with addressing the moderating effects of empirical variables on L2 learning. Q-tests were used to interpret whether the results of a moderator analysis were significant, while effect sizes gave insight into how effective each condition is compared to others in the same subgroup analysis. Moreover, confidence intervals

help in determining if an effect size calculation is robust. If the confidence interval crosses zero, it suggests that results may not be reliable.

Table 1, Moderator Analysis

			95% CI				*Q* tests	
Variables	*k*	*g*	LL	UL	*p*	I^2	*Q*	*p*
Age							0.09	0.76
Adult	69	1.00	0.74	1.26	0.00	.95		
Young	6	1.13	0.09	2.17	0.04	.89		
Language proficiency							8.93	**0.03**
Beginner	3	1.55	-1.20	4.29	0.14	.92		
Intermediate	29	1.19	0.63	1.75	0.00	.97		
Intermediate-Advanced	8	0.54	0.17	0.90	0.01	.91		
Advanced	13	1.11	0.74	1.39	0.00	.84		
L1							19.42	**0.01**
Arabic	5	1.69	-0.78	4.16	0.13	.98		
Multi	10	0.54	0.28	0.79	0.00	.81		
Chinese	11	1.11	0.30	1.91	0.01	.96		
Korean	4	0.54	0.13	0.95	0.03	.70		
Japanese	3	0.42	-1.67	2.52	0.48	.88		
Dutch	6	0.73	0.38	1.08	0.00	.84		

Vietnamese	11	1.51	0.99	2.03	0.00	.91		
Spanish/Catalan	4	0.40	-0.50	1.29	0.25	.96		
L2							0.69	0.41
English	71	0.99	0.73	1.24	0.00	.94		
Non-English	4	1.35	0.05	2.63	0.05	.96		
Learning target							11.19	0.08
Single word	38	1.30	0.89	1.70	0.00	.94		
Grammar	6	0.65	-0.58	1.89	0.23	.98		
Speaking	5	0.47	0.01	0.93	0.05	.79		
Listening	10	0.97	0.36	1.57	0.01	.95		
Multi Word	8	0.60	-0.01	1.45	0.07	.94		
Pronunciation	3	0.77	-2.78	4.33	0.45	.96		
Single + Multi	5	0.86	0.19	1.53	0.02	.80		
Multiple vs. single session							0.02	0.88
Single	46	1.00	0.73	1.27	0.00	.94		
Multiple	29	1.04	0.55	1.54	0.00	.96		
Language-focused							2.41	0.12
Language-focused	5	1.73	0.38	3.08	0.00	.98		
Non-language-focused	70	0.95	0.70	1.20	0.02	.94		
Video Type							30.04	**0.00**

Animation	5	1.54	-1.08	4.16	0.18	.97		
Documentary	14	1.49	0.52	2.47	0.01	.96		
Movies	15	0.53	0.20	0.74	0.00	.88		
TV Series	16	0.52	0.20	0.84	0.00	.93		
Lecture	3	1.45	-0.58	3.47	0.09	.94		
TED Talk	12	1.53	1.05	2.01	0.00	.90		
Various	3	0.63	0.08	1.19	0.04	.62		
Language-focused	5	1.73	0.38	3.08	0.02	.98		
Viewing Time							3.39	0.18
Short	10	1.16	0.23	2.09	0.02	.92		
Medium	20	1.23	0.82	1.64	0.00	.94		
Long	24	0.68	0.21	1.18	0.01	.95		
Viewing Plus							3.64	0.06
Yes	30	1.33	0.87	1.78	0.00	.96		
No	45	0.82	0.54	1.11	0.00	.93		
Test Type							0.67	0.41
Written	63	1.00	0.65	1.35	0.00	.96		
Aural/Oral	45	1.20	0.87	1.54	0.00	.96		
Test Type							2.07	0.15
Receptive	68	1.18	0.84	1.52	0.00	.96		
Productive	42	0.87	0.60	1.14	0.00	.96		

Age

Age did not significantly affect the extent to which audiovisual input promoted learning ($Q = 0.09$, $p = .76$). A significant medium effect was found for adult learners ($g = 1.00$, 95% CI [0.74, 1.26], $p < .00$) and young learners ($g = 1.13$, 95% CI [0.09, 2.17]). However, the sample size for young learners was relatively low ($k = 6$).

Language Proficiency

There was a significant difference ($Q = 8.93$, $p = .03$) between participants' reported language proficiency and the effects of viewing audiovisual input on L2 learning. For intermediate learners ($g = 1.19$, 95% CI [0.63, 1.75], $p < .00$) and advanced learners ($g = 1.11$, 95% CI [0.74, 1.39], $p < .00$) there was a medium effect. However, there was no effect for intermediate-advanced learners ($g = 0.54$, 95% CI [0.17, 0.90], $p = .01$). Results for beginner learners were insignificant ($p = .14$). The lower effect of intermediate-advanced learners could possibly be a result of the sample ($k = 8$) being derived from just 4 primary studies.

First Language

The *Q*-test indicated significant differences between first languages ($Q = 19.42$, $p = .01$). There were significant medium effects found for Chinese ($g = 1.11$, 95% CI [0.30, 1.91], $p = .01$) and Vietnamese ($g = 1.51$, 95% CI [0.99, 2.03], $p < .00$). A small, but significant, effect was found for Dutch ($g = 0.73$, CI 95% [0.38, 1.08], $p = .00$). Results from mixed L1s ($g = 0.54$, 95% CI [0.28, 0.79], $p < .00$) and Korean ($g = 0.54$, 95% CI [0.13, 0.95], $p = .03$) were significant but had no effect. Arabic ($p = .13$), Japanese ($p = .48$), and Spanish/Catalan ($p = .25$) produced insignificant results.

Target Language

The *Q*-test result examining differences between the target language subgroups was not significant ($Q = 0.69$, $p = .41$). English as a target language had a significant small effect ($g = 0.99$, 95% CI [0.73, 1.24], $p < .00$) that bordered on medium. There was a large and significant effect for non-English target languages ($g = 1.35$, 95% CI [0.05, 2.63], $p < .05$), however it is worth noting that there was a small sample in this group ($k = 4$) which is likely the cause of such a large confidence interval.

Learning Target

Results from the *Q*-test suggest that there was no significant difference between learning targets ($Q = 11.19$, $p = .08$). Audiovisual viewing had a medium and significant effect on the learning of single words ($g = 1.30$, 95% CI [0.89, 1.70], $p < .00$). In the learning of listening skills ($g = 0.97$, 95% CI [0.36, 1.57]. $p = .02$) and mixed single- and multi-word units ($g = 0.86$, 95% CI [0.19, 1.53], $p = .02$) a small effect was found. No effect was found for speaking ($g = 0.47$, 95% CI [0.01, 0.93], $p = .05$). Learning targets of multi-word units, grammar, and pronunciation were all insignificant. This is likely due to the relatively small sample sizes of these categories.

Multiple vs. Single Sessions

There was no significant difference between single or multiple treatment sessions ($Q = 0.02$, $p = .88$). There was a medium and significant effect of multiple sessions ($g = 1.04$, 95% CI [0.55, 1.54], $p < .00$), and single sessions ($g = 1.00$, 95% CI [0.73, 1.27], $p < .00$). These results suggest that while multiple treatment sessions may have a greater effect, this difference may be negligible.

Language-focused video

There was no significant difference between whether the audiovisual input was language-focused or not (Q = 2.41, p = .12). The effects of viewing language-focused audiovisual input were significant and produced a large effect (g = 1.73, 95% CI [0.38, 3.08], p < .00), but had a large confidence interval due to a relatively small number of studies (k = 5). The effect of non-language-focused audiovisual input was small but had a much smaller confidence interval (g = 0.95, 95% CI [0.70, 1.20], p = .02). The large confidence interval of language-focused videos makes the results of this analysis somewhat unclear.

Audiovisual Input Type

There was a significant difference between the effects of different types of audiovisual input (Q = 30.04, p = .00). There were significant large effects found for documentaries (g = 1.49, 95% CI [0.52, 2.47], p = .00) and TED Talks (g = 1.53, 95% CI [1.05, 2.01], p = .00). Results for language-focused videos are the same as the previous subgroup analysis. There was no effect found for movies (g = 0.53, 95% CI [0.20, 0.74], p = .00) or TV series (g = 0.52, 95% CI [0.20, 0.84], p = .00). Mixed video types had a small and significant effect (g = 0.63, 95% CI [0.08, 1.19], p = .04), however all ESs came from the same primary study and these findings should therefore be interpreted cautiously. Two types of audiovisual input had insignificant p-values and confidence intervals that spanned zero: animation (p = .18) and lectures (p = .09).

Viewing Time

Viewing time had no significant effect on language learning (Q = 3.39, p = .18). A medium effect was found for short (g = 1.16, 95% CI [0.23, 2.09], p = .02) and medium viewing times (g = 1.23, 95% CI [0.82, 1.64], p = .00), while long viewing times had a small effect (g = 0.68, 95% CI [0.21, 1.18], p = .00). These results suggest that while shorter periods of time spent

viewing audiovisual input might have a slightly higher effect, it is not significantly more impactful.

Viewing Plus

There was not a significant difference between whether or not viewing was supplemented with other activities ($Q = 3.64$, $p = .06$). Viewing plus yielded a medium effect ($g = 1.33$, 95% CI [0.87, 1.78], $p = .00$), while viewing alone had a small effect on language learning ($g = 0.82$, 95% CI [0.54, 1.11], $p = .00$). These results suggest that while additional activities may have a positive impact on language learning through audiovisual input, these benefits are not significant.

Written vs. Oral/Aural Tests

There was no significant difference found between written and aural/oral tests ($Q = 0.67$, $p = .41$). There was a medium effect found for both written ($g = 1.00$, 95% CI [0.65, 1.35], $p = .00$) and aural/oral tests ($g = 1.20$, 95% CI [0.87, 1.54], $p = .00$). However, aural/oral tests had a higher average effect. It is worth noting that this subgroup analysis is likely to have a substantial amount of covariance between effect sizes since tests within groups are not averaged to yield independent effect sizes. Results should therefore be interpreted cautiously.

Productive vs. Receptive Tests

There was no significant difference between whether a test was receptive or productive ($Q = 2.07$, $p = .15$). There was a medium effect for receptive ($g = 1.18$, 95% CI [0.84, 1.52], $p = .00$) and a small effect for productive tests ($g = 0.87$, 95% CI [0.60, 1.14], $p = .00$). Like the previous moderator variable, this analysis utilizes individual test scores meaning it is susceptible to a fair degree of covariance that is not statistically accounted for.

2.6 Discussion

2.6.1 Pooled Effect

This meta-analysis aimed to quantify the extent to which viewing audiovisual input contributes to L2 learning. Results of the meta-analysis yielded a Hedges' g of 1.01, which represents the average gain in terms of standard deviation from pre-test to post-test. Using Plonsky and Oswald's (2014) benchmarks, this is a medium effect. These results suggest that there was a clear and significant improvement in L2 learning after viewing audiovisual input. While previous literature generally shows that L2 learning occurs through viewing audiovisual input (Montero Perez, 2022), the influence of unenhanced video had yet to be quantified. Furthermore, while individual studies show a positive effect on language learning, this synthesis of available research helps to solidify the extent to which this effect contributes to L2 learning. This may help increase acceptance of audiovisual input as a tool within the language learning classroom. This is particularly important given that many still fail to acknowledge the validity of audiovisual input as a L2 learning tool (Webb, 2015).

Although a comparison to other meta-analyses that look at previous areas is important, methodological choices make a direct comparison somewhat difficult. In Montero Perez et al. (2013) unenhanced audiovisual input served as the control group which means that their meta-analysis was only able to tell us that captions benefited learners more than audiovisual input alone. This study builds on these results by giving a baseline of the extent to which audiovisual input contributes to language learning. While the results from the current study are favorable, they also help to promote the use of captions since Montero Perez et al. (2013) found captions to contribute to greater learning than unenhanced audiovisual input.

2.6.2 Moderators

Age did not have a significant influence on the impact of audiovisual input on language learning. This suggests that learners of all ages may benefit from viewing audiovisual input. However, it is important to note that the sample size was largely skewed toward adult language learners (young learners, $k = 6$, adult learners, $k = 69$). Thus, it is clear that adult L2 learners benefit from audiovisual input. However, the small sample size for young learners indicates a need for further investigation. This has been a common finding across meta-analyses in applied linguistics (e.g., de Vos et al., 2018; Lee et al., 2015). Although studies often recruit adults for practical reasons, the scope of existing research would benefit from further research involving younger learners.

It has been suggested that more advanced learners might benefit more from unenhanced audiovisual input than lower-level learners (Leveridge & Yang, 2013), however, findings from this subgroup analysis found similar outcomes among intermediate- and advanced-level learners. Although there was a significant difference between groups, this is likely due to the intermediate-advanced group having a small primary study size ($k = 4$). Looking at only the effect sizes between the intermediate ($g = 1.19$, 95% CI [0.63, 1.75]) and advanced ($g = 1.11$, 95% CI [0.74, 1.39]) groups, effect sizes are similar with confidence intervals that closely align. There is no obvious reason that a group with a proficiency situated between intermediate and advanced learners would benefit significantly less from audiovisual viewing unless: 1) the videos presented to these learners were beyond their level since their proficiency was at an intermediary position; 2) the smaller sample size skewed results.

The first language of learners had a significant influence on the effect size. One might expect that learners with L1s similar to English (e.g., Dutch) would benefit more from

audiovisual input than those with L1s that are grammatically, phonologically, and lexically distant (e.g., Chinese or Vietnamese). Results from this subgroup analysis, however, showed the opposite to be true. The reason for this outcome is unclear and warrants further investigation.

Target L2 was not a significant moderator of L2 learning. While both English and non-English L2s reached significance, just four effect sizes involved target languages other than English. The results reveal a need to research the learning of languages other than English.

A key area of interest in this study was the L2 learning target. The aim of this moderator variable was to determine whether audiovisual contributes to a range of aspects of L2 learning rather than just individual areas. The results showed that there was no significant difference between L2 targets. This suggests that a range of skills can be improved through viewing unenhanced video. However, it should be noted that three L2 targets did not reach significance. The lack of significance is likely due to the relatively small sample sizes of these groups (grammar, $k = 6$; multi-word, $k = 8$; pronunciation, $k = 3$). The effects for all three insignificant areas were small. However, because the results were not significant, these results should be interpreted with caution. Although the majority of studies focused on vocabulary learning (single items, $k = 38$; multi-word items, $k = 8$; mixed, $k = 5$), there were still a large number of studies focusing on other areas of L2 learning ($k = 24$). For this reason, although individual subgroups may not be significant, there is a strong enough basis to suggest that learning may occur to a similar extent across various areas of L2 learning. Nonetheless, clarifying the effects on different areas of language learning through further research is crucial. In particular, the relatively small number of studies investigating pronunciation and speaking indicate that further research in these areas is warranted.

Treatment related variables also provided some insight into the current state of audiovisual input research. There were no significant differences between studies which used multiple and single treatment sessions. At first glance this could be interpreted as there being no difference between the two types of treatments and therefore single sessions should be used for the most efficient learning approach. However, the results may also indicate that multiple viewing sessions allowed participants to retain a similar level of knowledge for a longer period of time because knowledge gained in the first session could still be recalled much later. Although single sessions allow for initial acquisition of target items, spacing learning across multiple sessions benefits retention of knowledge (Kim & Webb, 2022)..

The results also showed that language-focused videos had a large effect on learning, while non-language-focused videos had a small effect (which bordered on medium) with no significant difference between the two types of videos. The larger effect size for language-focused videos may be due the focus on learning language features, slower speech rates, and controlled use of language to make the input more comprehensible for learners.

Audiovisual input was further categorized by type. While there was no effect of movies and TV series on learning, there were large effects from viewing both TED Talks and documentaries. A possible explanation is that TED Talks and documentaries are educational in nature, whereas TV series and movies are primarily concerned with providing entertainment. Therefore, unfamiliar language is more likely to be contextualized and spoken clearly in TED Talks and documentaries so that viewers can better comprehend the content. On the other hand, types of videos that serve to entertain the audience may focus less on helping viewers to understand unfamiliar language. A possible solution to this issue is watching television programs and movies with captions, which support learning and understanding (Montero Perez, 2022).

Among the other video types, language-focused videos had the largest effect among all the groups, and animations and lectures did not reach significance, perhaps because their sample sizes (k = 5, k = 3, respectively) were smaller than other audiovisual types. Following these findings, learners should be encouraged to consume educational media. While there was still a positive effect of entertainment-focused media, there may be a need for the use of learning support through captions, repeated viewing, pre and post viewing activities to enhance learning and understanding (Webb, 2015).

Viewing time was not a significant factor in language learning. There was a slightly smaller effect of longer viewing over shorter viewing. One explanation for this is that when more content is viewed, there is a lower density of target L2 features, which may draw attention away from target items. This is especially true in studies that focus on incidental learning where participants are not aware of the L2 targets. Another possibility is that there may be more target items for participants to attend to over longer viewing times. This would mean that participants had to learn more items which may put a burden on their cognitive processing facilities. A final observation of the data for viewing time is that few studies used short-form videos. This is an area that might be investigated in future studies given the increasing popularity of short-form content (e.g., TikTok, YouTube Shorts, Instagram Reels).

The moderator analysis also showed that viewing plus did not have a significantly greater effect on language learning than viewing alone. Nonetheless, the effect of viewing plus was medium (g = 1.33), while viewing alone had a small effect (g = 0.82). Therefore, while this analysis does not reach statistical significance, the effects of viewing plus are demonstrably larger than viewing without additional activities. While it should be expected that supplementing audiovisual input with activities that reinforce L2 learning increases gains, this finding suggests

that there is merit to integrating audiovisual input into the curriculum, rather than simply guiding students to consume audiovisual input outside of the classroom. This in turn may have the positive effect of reducing hesitancy to use audiovisual input in the classroom (Webb, 2015). When interpreting these results, it should be cautioned that some studies (e.g., Weyers, 1999) treated participants with audiovisual input over a period of time and had activities structured around the input. Therefore, all learning may not necessarily be attributed to audiovisual input.

Test related variables provided some valuable insight about how measurement may affect the interpretation of L2 learning gains. First, aural/oral and written tests were examined to determine if there was a test congruency effect. Results show that while there was not a significant difference between these two formats, there was a slightly larger effect found for aural/oral tests than written tests. This suggests that while both tests can detect learning that has occurred, congruent tests may reveal slightly larger learning gains. An important consideration in this analysis is that effect sizes from individual tests were pooled. Because individual tests were pooled, there is a fair degree of covariance between effect sizes and results should therefore be interpreted with caution.

The moderator analysis also showed that there was no significant difference between receptive and productive tests; however receptive tests did have a slightly higher overall effect (receptive, $g = 1.18$; productive, $g = 0.87$). The slightly larger gains on receptive tests may be attributed to transfer appropriate processing (Morris et al., 1977); the congruency between the receptive learning condition (viewing audiovisual input) and receptive test formats leads to larger gains than incongruent learning and testing conditions. Similar to the oral/aural and written test subgroup analysis, results need to be interpreted with caution due to unaccounted for covariance.

2.7 Conclusion

Overall, there seems to be an overall benefit of viewing audiovisual input across all areas of L2 learning. Not only is there a positive effect, but positive effects were found across different L2 learning targets, for learners at different proficiency levels, and across different regions. Findings also suggest that learners may benefit greatest from activities used to supplement learning through viewing audiovisual input. Therefore, teachers should be encouraged to integrate L2 audiovisual input into the curriculum. Should learners view audiovisual input without captions in their own time, they may benefit the most from viewing education-focused content such as TED Talks or documentaries, rather than more entertainment-focused materials like movies and TV series. When viewing movies and TV series, learners should be encouraged to use captions to further support their L2 learning.

2.8 Limitations and Future Directions

There are several limitations of this study that should be considered. First, this study uses a pre-post contrast approach to calculating effect sizes. While this allows for inclusion of a greater number of studies, pre-post contrasts tend to have ESs that are positively biased because of the pre-post correlation (Plonsky & Oswald, 2014). To help counteract this, ESs were interpreted with Plonsky and Oswald's (2014) benchmarks which account for the larger ESs. Second, multiple ESs were taken from primary studies. This means that there is some dependency between some ESs that is not accounted for (Hedges, 2019). Nonetheless, Hedges (2019) suggests that where relatively few studies report multiple ESs, this approach is unlikely to threaten validity.

The next limitation is that meta-analyses are reliant on the studies available. The impact of audiovisual input on different areas of language learning was a focus of this study. Apart from

vocabulary-focused studies there is still a lack of research in all areas of L2 learning. Although there has been increased research on grammar learning through audiovisual input (e.g., Cintrón-Valentín & García-Amaya, 2021; García–Amaya & Cintrón–Valentín, 2021; Pattemore & Muñoz, 2022), there is a need for greater research in other areas of L2 learning. By increasing the scope of research, the benefits of viewing audiovisual input for L2 learning may be more clearly determined.

The sample is also limited by a lack of research on audiovisual input with younger learners. Since people of all ages engage in both viewing audiovisual input and L2 learning, it is important that research investigates the effects of treatments across age groups. For this reason, it is also suggested that research also look at older learners (i.e., above university age).

Furthermore, there are few studies that investigate audiovisual input with a true control (e.g., Aldukhayel, 2021; Dang et al., 2022; Feng & Webb, 2020; Wisniewska & Mora, 2020). In order to conduct a between-group meta-analysis, more studies need to be conducted with a test-only group. A between-group meta-analysis would allow for the control of intragroup correlations that is present in a pre-post approach (Plonsky & Oswald, 2014).

Overall findings of this research confirm the utility of viewing audiovisual input for L2 learning. These findings should help encourage both learners and educators to use this tool to improve language proficiency.

Chapter 3: Conclusion

This chapter provides an overview of findings from Chapter 2, as well as future directions for research, and implications for pedagogy.

3.1 Review of Findings

Overall, audiovisual input was shown to have a medium effect on L2 learning (g = 1.02). This means that learners can expect to gain knowledge of a second language by watching unenhanced audiovisual input. This knowledge may be acquired both through viewing with supplementary activities or viewing alone. However, results suggest that viewing alone only has a small effect in comparison to viewing with supplementary activities, which had a large effect. Learners may therefore benefit from using both approaches, where viewing with activities may be a good choice for the classroom, viewing alone for leisure may still provide some benefit. It is also worth noting that the type of audiovisual input learners choose to consume will make a large difference in the extent to which L2 learning gains are made. For example, learning through language-focused videos had the largest effect on L2 learning, while non-language-focused videos had a lesser effect. This analysis was broken down further by categorizing videos by type (e.g., movie, TV series, documentary, etc.) where it was found that entertainment-focused media led to minimal gains, while educational-focused content (e.g., language-focused videos, documentaries, and TED Talks) contributed to greater L2 gains.

Several moderators did not have any effect on the results. There was no significant difference between the areas of L2 learning nor the age of participants. The type of test used (whether aural/oral or written and productive or receptive) also showed no significant findings.

3.2 Future Directions

Overall findings of this research confirm the utility of viewing audiovisual input for L2 learning. These findings should help encourage both learners and educators to use this tool to improve language proficiency. Findings also study suggest future directions for researchers. For one, when comparing other areas of L2 learning to vocabulary learning, there is a dearth of research. There has been increased research on grammar learning through audiovisual input (e.g., Cintrón-Valentín & García-Amaya, 2021; García–Amaya & Cintrón–Valentín, 2021; Pattemore & Muñoz, 2022), however, other areas of L2 learners do not seem to be attracting as much attention. By increasing the scope of research, the benefits on other areas of language learning may be more clearly determined. There is also a lack of research on audiovisual input and younger learners. Although results from this meta-analysis showed that the young learner subgroup reached significance, research heavily favors university-aged participants. Since people of all ages engage in L2 learning, it is important that research investigates the effects of treatments across age groups. For this reason, it is also recommended that research look at older learners (i.e., above university age). Since audiovisual content is consumed across all age groups, it would be useful to more clearly determine the extent to which each age group can learn from audiovisual input.

Furthermore, there are few studies that investigate audiovisual input with a true control group (e.g., Aldukhayel, 2021; Dang et al., 2022; Feng & Webb, 2020; Wisniewska & Mora, 2020). In order to conduct a between-group meta-analysis, more studies across various areas of language learning would need to be conducted with a test-only group. A between-group meta-analysis would allow for the control of intragroup correlations that is present in a pre-post approach (Plonsky & Oswald, 2014).

References

Aldera, A. S., & Mohsen, M. A. (2013). Annotations in captioned animation: Effects on vocabulary learning and listening skills. *Computers & Education, 68,* 60–75. https://doi.org/10.1016/j.compedu.2013.04.018

Aldukhayel, D. (2021). The effects of captions on L2 learners' comprehension of vlogs. *Language Learning,* 14.

Baddeley, A. (1990). *Human memory*. Lawrence Erlbaum Associates.

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

Baetens Beardsmore, H., & van Beeck, H. (1984). Multilingual television supply and language shift in Brussels. *International Journal of the Sociology of Language, 48,* 65–79.

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PLOS ONE, 11*(6), e0158409. https://doi.org/10.1371/journal.pone.0158409

Borenstein, M., & Hedges, L. V. (2019). Effect sizes for meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd edition, pp. 207–243). Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Charles, T. J., & Trenkic, D. (2015). Speech 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. 173–198). Peter Lang.

Chen, I.-J., Chang, C.-C., & Lee, Y.-C. (2009). Applications of cognitive load theory to multimedia-based foreign language learning: An overview. *Educational Technology, 49*(1), 34–39.

Cintrón-Valentín, M. C., & García-Amaya, L. (2021). Investigating textual enhancement and captions in L2 grammar and vocabulary. *Studies in Second Language Acquisition*, *43*(5), 1068–1093. https://doi.org/10.1017/S0272263120000492

Clarivate. (2023). *Web of Science Master Journal List*. Web of Science Group. https://mjl.clarivate.com/

Covidence systematic review software. (n.d.). [Computer software]. Veritas Health Innovation. www.covidence.org

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge, Taylor & Francis Group.

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

de Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, *68*(4), 906–941. https://doi.org/10.1111/lang.12296

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

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

García–Amaya, L., & Cintrón–Valentín, M. C. (2021). The effects of textually enhanced captions on written elicited imitation in L2 grammar. *The Modern Language Journal*, *105*(4), 919–935. https://doi.org/10.1111/modl.12740

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2019). *dmetar: Companion R package for the guide 'doing meta-analysis in R* (0.0.9000) [Computer software]. http://dmetar.protectlab.org/

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman & Hall/CRC Press. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/

Hedges, L. V. (2019). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Third ed.). Russell Sage Foundation.

Jelani, N. A. M., & Boers, F. (2020). Examining incidental vocabulary acquisition from captioned video: Does test modality matter? In S. Webb (Ed.), *Approaches to learning, testing and researching L2 vocabulary* (Vol. 109, pp. 169–189). John Benjamins Publishing Company. https://doi.org/10.1075/bct.109.itl.00011.jel

Kam, E. F., Liu, Y.-T., & Tseng, W.-T. (2020). Effects of modality preference and working memory capacity on captioned videos in enhancing L2 listening outcomes. *ReCALL*, *32*(2), 213–230. https://doi.org/10.1017/S0958344020000014

Khosh Ayand, F., & Shafiee, S. (2016). Effects of English and Persian subtitles on oral fluency and accuracy of intermediate Iranian EFL learners. *International Journal of Applied Linguistics and English Literature*, *5*(3). https://doi.org/10.7575/aiac.ijalel.v.5n.3p.133

Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning*, *72*(1), 269–319. https://doi.org/10.1111/lang.12479

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345–366. https://doi.org/10.1093/applin/amu040

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

Li, S. (2016). A Meta-Analysis The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*(4), 801–842. https://doi.org/10.1017/S027226311500042X

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

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 227–246). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.010

Majuddin, E. (2020). *Incidental and intentional acquisition of multiword expressions from audio-visual input: The effects of typographically enhanced captions and repetition* [Doctoral dissertation]. Victoria University of Wellington.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*(2), 312–320. https://doi.org/10.1037/0022-0663.90.2.312

Microsoft Excel. (2018). [Computer software]. Microsoft Corporation. https://office.microsoft.com/excel

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

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

Montero Perez, M., Simon, E., Goethals, P., & Pattemore, A. (2022). *Effects of captioned video on L2 speech perception in intermediate L2 learners of Spanish*. CALICO, Seattle, WA, USA.

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

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

Muñoz, C., Pujadas, G., & Pattemore, A. (2021). Audio-visual input for learning L2 vocabulary and grammatical constructions. *Second Language Research*, 026765832110157. https://doi.org/10.1177/02676583211015797

Nguyen, C., & Boers, F. (2019). The Effect of Content Retelling on Vocabulary Uptake From a TED Talk. *TESOL Quarterly*, *53*(1), 5–29. https://doi.org/10.1002/tesq.441

Paivio, A. (1990). *Mental representations*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Pattemore, A., & Montero Perez, M. (2022). *Learning of Spanish subjunctive through original version audio-visual input*. International Conference of the Spanish Society for Applied Linguistics (AESLA), Gran Canaria, Spain.

Pattemore, A., & Muñoz, C. (2020). Learning L2 constructions from captioned audio-visual exposure: The effect of learner-related factors. *System*, *93*, 102303. https://doi.org/10.1016/j.system.2020.102303

Pattemore, A., & Muñoz, C. (2022). Captions and learnability factors in learning grammar from audio-visual input. *The JALT CALL Journal*, *18*(1), 83–109. https://doi.org/10.29140/jaltcall.v18n1.564

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

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, *53*(4), 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E., 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., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Plass, J. L., & Jones, L. C. (2005). Multimedia learning in second language acquisition. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (1st ed., pp. 467–488). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.030

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Polat, M., & Erişti, B. (2019). The effects of authentic video materials on foreign language listening skill development and foreign language listening anxiety at different levels of English proficiency. *International Journal of Contemporary Educational Research*. https://doi.org/10.33200/ijcer.567863

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The*

Language Learning Journal, *47*(4), 424–438. https://doi.org/10.1080/09571736.2019.1638630

Quaglio, P. (2010). *Television dialogue: The sitcom "Friends" vs. natural conversation* (Paperback ed). Benjamins.

Reynolds, B. L., Cui, Y., Kao, C.-W., & Thomas, N. (2022). Vocabulary acquisition through viewing captioned and subtitled Video: A scoping review and meta-analysis. *Systems*, *10*(5), 133. https://doi.org/10.3390/systems10050133

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

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Sadowska, A. (2015). Learning English vocabulary from film audio description: A case of Polish sighted students. *Roczniki Humanistyczne*, *63*(11), 101–123. https://doi.org/10.18290/rh.2015.63.11-7

Schwarzer, G. (2023). *General package for meta-analysis* (6.2-1) [R]. https://github.com/guido-s/meta/

Scopus—Sources. (n.d.). Retrieved April 7, 2023, from https://www.scopus.com/sources.uri?zone=TopNavBar&origin=searchbasic

Shen, D. C. (1991). *The effects of captioning on listening comprehension of English as a second language in a computer-based interactive videodisc system* [Doctoral dissertation]. University of Northern Colorado.

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning*, 24.

Terrell, T. D. (1993). Comprehensible input for intermediate foreign language students via video. *IALLT Journal of Language Learning Technologies*, *26*(2), 17–24. https://doi.org/10.17161/iallt.v26i2.9499

Tsai, F.-H. (Will). (2010). *Integrating feature films with subtitles to enhance the listening comprehension of students attending college in Taiwan* [PhD Thesis]. Alliant International University.

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. https://doi.org/10.1111/lang.12343

Vanderplank, R. (2010). *Déjà vu* ? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, *43*(1), 1–37. https://doi.org/10.1017/S0261444809990267

Viechtbauer, W. (2022). *Metafor* (3.8-1) [R]. https://www.metafor-project.org/doku.php/metafor

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). Routledge. https://doi.org/10.4324/9781315883472

Webb, S., & Chang, A. C.-S. (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. https://doi.org/10.1177/1362168814559800

Weyers, J. R. (1999). The effect of authentic video on communicative competence. *The Modern Language Journal*, *83*(3), 339–349.

Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition, 42*(3), 599–624. https://doi.org/10.1017/S0272263120000029

Appendices

Appendix A: Effect Size and Standard Error Calculation

To calculate a within-group effect size, first, Cohen's *d* was calculated:

$$d = \frac{M_{diff}}{s_{av}} = \frac{M_{post} - M_{pre}}{s_{av}}$$ (Cumming, 2012)

Where s_{av} was calculated with the following formula:

$$s_{av} = \sqrt{\frac{s_{pre}^2 + s_{post}^2}{2}}$$ (Cumming, 2012)

Cohen's *d* is known to produce an effect size that is biased, overestimating the effect for smaller samples (Borenstein & Hedges, 2019). For this reason, the correction factor *J* was used to convert *d* to Hedges' *g*, which reduces this bias. The correction factor was first calculated:

$$J = \left(1 - \frac{3}{4*(n-1)-1}\right)$$ (Borenstein & Hedges, 2019)

After the correction factor has been calculated, Hedges' *g* can be calculated by simply multiplying the correction factor by Cohen's *d*.

$$g = J * d$$ (Borenstein et al., 2009)

This yields Hedges' *g* which corrects for the overestimation produced by Cohen's *d*.

In order to weight studies, the meta package (Schwarzer, 2023) requires that a study's standard error is calculated. To do so, variance must first be calculated for each study. Variance was calculated as the following:

$$V_d = \left(\frac{1}{n} - \frac{d^2}{2n}\right) 2(1 - r)$$ (Borenstein & Hedges, 2019)

After calculating the variance of Cohen's *d* for the study, it should be multiplied by the previously calculated correction factor, *J*. This yields V_g, which can be used to calculate the standard error by taking its square root:

$$SE_g = \sqrt{V_g}$$ (Borenstein & Hedges, 2019)

One problematic factor in calculating the study's standard error is that the equation for variance requires that the correlation between the pre-test and post-test is known. Unfortunately, in applied linguistics this correlation is rarely reported. To circumnavigate this issue researchers can borrow correlations from similar studies to approximate the study's variance (Borenstein et al., 2009; Borenstein & Hedges, 2019). For the purposes of this study, correlations were taken from included studies where a correlation was reported (or could be calculated based on reported data) in conjunction with several that were provided via correspondence with authors. A total of 11 correlations were coded with an average of $r = 0.758$. As recommended by Borenstein and Hedges (2019), a sensitivity analysis was conducted to determine if this estimate was robust. The sensitivity analysis consisted of increasing and decreasing the averaged correlation by 0.10 (see Appendix B). The change in correlation had little effect on the overall effect size.

Appendix B: Sensitivity Analysis

One study was excluded from the meta-analysis after initially being coded. Polat & Erişti (2019) was initially included in the meta-analysis but contributed four abnormally large effect sizes (g = 11.73, g = 5.98, g = 16.48, and g = 7.09). The effect size of a study was not generally a factor in the exclusion of studies, however as can be seen in Table B1, Polat & Erişti (2019) had a large influence on the overall effect size, as well as a disproportionate influence on the Q-value.

Table B1, Influence Analysis

Analysis	k	g	95% CI	p	I^2	95% CI	Q
Main	75	1.01	[0.77, 1.26]	.000	94.6%	[93.8, 95.3]	1369.29
Incl. Polat & Erişti	79	1.20	[0.83, 1.56]	.000	95.4%	[94.8, 96.0]	1709.49

A subgroup analysis was also conducted to check if there was a significant difference between studies which utilized a borrowed correlation or not. The results showed that studies that used borrowed correlations had a slightly more conservative overall ES (g = 0.98, 95% CI [0.73, 1.23] versus g = 1.49, 95% CI [0.32, 2.66]), but that there was no significant difference (p = .30) between studies which borrowed correlations and those that did not. Although the magnitudes of the effect sizes are different, the confidence interval of the non-borrowed correlation group is large. Furthermore, a sensitivity analysis conducted by adjusting the borrowed correlations by ±0.1 also showed only slight variation in the overall pooled effect and I^2 value (see Table B2). Both the sensitivity analysis and subgroup analysis suggest that the estimation provided by the average of the correlations was robust.

Table B2, Borrowed Correlation Sensitivity Analysis

Model	*g*	95% CI	*p*	I^2	95% CI	*Q*
Main Analysis (*r* = .758)	1.01	[0.77, 1.26]	.000	94.6%	[93.8, 95.3]	1369.29
Correlation + 0.1 (*r* = .858)	1.04	[0.79, 1.29]	.000	96.7%	[96.2, 97.1]	2224.69
Correlation - 0.1 (*r* = .658)	0.99	[0.75, 1.23]	.000	92.7%	[91.5, 93.8]	1013.43

Appendix C: Study-Related Moderator Variables Coding and Results

Indexed

Studies were coded based on whether or not they were indexed in Scopus Sources (*Scopus - Sources*, n.d.) or Web of Science Master Journal List (Clarivate, 2023). The included studies did not all aim to answer the same research questions (e.g., audiovisual was often the control group in a caption-focused study) and therefore publication of these studies was not dependent on the effect sizes calculated in this study making the use of a funnel plot unreliable for testing publication bias. The purpose of this moderator variable was to determine if there was a significant difference between indexed and unindexed studies. Although this cannot tell us whether a publication bias exists, it can help us determine if there is a difference in the effect sizes of studies across journals. If the study was published in one of the aforementioned indexers, it was considered indexed.

Publication Type

Studies were also coded based on their type of publication. Studies were coded as follows: journal, doctoral thesis, master's thesis, conference, or book chapter.

Study Region

The region where studies were conducted was also coded. Regions were coded with two letters to represent their geographic area (ME = Middle East & Northern Africa; NA = North America; EU = Europe; OC = Oceana; AS = Asia; and SA = South America).

Results for Study-Related Moderator Variables

Table C1, Moderator Analysis for Study-Related Variables

Variables	*k*	*g*	95% CI		*p*	I^2	*Q* tests	
			LL	UL			*Q*	*p*
Publication Type							0.04	0.98
Journal	49	1.02	0.70	1.35	0.00	.95		
Doctoral Thesis	19	0.98	0.54	1.41	0.00	.94		
Master's Thesis	4	0.97	-0.18	2.12	0.08	.86		
Indexed							0.17	0.68
Yes	38	1.06	0.69	1.44	0.00	.95		
No	37	0.96	0.63	1.29	0.00	.94		
Region							6.38	**0.04**
M.E.	14	1.54	0.46	2.62	0.01	.97		
E.U.	22	0.68	0.43	0.93	0.00	.92		
A.S.	34	1.12	0.77	1.46	0.00	.95		

Publication Type

The *Q*-test revealed that there was no significant difference between the types of publication ($Q = 0.04$, $p = .98$). Studies published in journals had a medium effect ($g = 1.02$, 95% CI [0.70, 1.35], $p = .00$). The pooled effect from doctoral theses was small, but significant ($g = 0.98$, 95% CI [0.54, 1.41], $p = .00$), while master's theses had a small, but insignificant, effect ($g = 0.97$, 95% CI [-0.18, 2.12], $p = .08$).

Indexed

There was no significant difference between studies that were in indexed publications versus non-indexed publications (Q = 0.17, p = .68). Indexed publications had a medium effect of g = 1.06 (95% CI [0.69, 1.44], p = .00), while non-indexed publications had a small effect (g = 0.96 (95% CI [0.63, 1.29], p = .00). These results suggest that there may not be a bias in findings that are published in indexed journals, however they may not be as reliable as the results from a funnel plot, and therefore, should be interpreted cautiously.

Study Region

The Q-test examining differences between the region a study was conducted reached significance (Q = 6.38, p = .04). Studies conducted in the Middle East and Northern Africa had a large effect (g = 1.54, 95% CI [0.46, 2.62], p = .01), but with a large confidence interval. Studies conducted in Asia had a medium effect (g = 1.12, 95% CI [0.77, 1.46], p = .00). There was a small effect for studies that were conducted in Europe (g = 0.68, 95% CI [0.43, 0.93], p = .00).

Discussion of Study-Related Moderator Variables

Publication type and whether or not a publication was indexed in in Scopus Sources (*Scopus - Sources*, n.d.) or Web of Science Master Journal List (Clarivate, 2023) were also investigated. In neither case was there a differential effect between groups. Although this cannot tell us with certainty that publication bias is not present, it does suggest that whether literature was considered to be of "higher quality" or not made little difference in the overall effects found. It also suggests that effects across different study types (theses, journal articles, etc.) did not differ in their effect sizes.

There was no significant difference between regions. Furthermore, there is a large degree of overlap in confidence intervals which suggests that the true ES may be closely aligned.

Appendix D: Meta-Analyzed Studies

Aldera, A. S., & Mohsen, M. A. (2013). Annotations in captioned animation: Effects on vocabulary learning and listening skills. Computers & Education, 68, 60–75. https://doi.org/10.1016/j.compedu.2013.04.018

Aldukhayel, D. M. (2022). Comparing L2 incidental vocabulary learning through viewing, listening, and reading. Journal of Language Teaching and Research, 13(3), 590–599. https://doi.org/10.17507/jltr.1303.15

Alsadoon, R. (2021). Closed-captioning and the learning of grammatical structures for EFL learners. Zarqa Journal for Research and Studies in Humanities, 21(3), 649–661. https://doi.org/10.34028/1069-021-003-014

Alshumrani, H. (n.d.). L2 Incidental vocabulary learning and retention through different modalities of audio-visual input [Doctoral dissertation]. https://eprints.soton.ac.uk/434623/

Ayand, F. K., & Shafiee, S. (2016). Effects of English and Persian subtitles on oral fluency and accuracy of intermediate Iranian EFL learners. International Journal of Applied Linguistics and English Literature, 5(3). https://doi.org/10.7575/aiac.ijalel.v.5n.3p.133

Baltova, I. (1999). The effect of subtitled and staged video input on the learning and retention of content and vocabulary in a second language [Doctoral dissertation, University of Toronto]. https://hdl.handle.net/1807/13234

Bellalem, F., Neddar, B. A., Bouagada, H., & Djellou, D. B. l. (2018). The use of subtitled movies for vocabulary acquisition in ESP settings: Insights from an experimental study in

Algeria. Arab World English Journal, 9(3), 3–16. https://doi.org/10.24093/awej/vol9no3.1

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. PLoS ONE, 11(6). https://doi.org/10.1371/journal.pone.0158409

Chai, J., & Erlam, R. (2008). The effect and the influence of the use of video and captions on second language learning. New Zealand Studies in Applied Linguistics, 14(2), 25–44.

Charles, T. J., & 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. 173–198). Peter Lang.

Cintrón-Valentín, M. C., & García-Amaya, L. (2021). Investigating textual enhancement and captions in L2 grammar and vocabulary: An experimental study. Studies in Second Language Acquisition, 43(5), 1068–1093. https://doi.org/10.1017/S0272263120000492

Csapó-Sweet, R. M. (1997). “Sesame Street”, English vocabulary and word usage of Hungarian ESL students. Communications, 22(2), 175–190. https://doi.org/10.1515/comm.1997.22.2.175

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

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

Ismahane, A. (2014). The effects of watching subtitled English movie clips on enhancing students' vocabulary learning [Master's Thesis]. Larbi Ben M'Hidi University.

Karaka, A., & Sariçoban, A. (2012). The impact of watching subtitled animated cartoons on incidental vocabulary learning of ELT students. Teaching English with Technology, 12(4), 3–15.

Kim, H.-S. (2015). Using authentic videos to improve EFL students' listening comprehension. International Journal of Contents, 11(4), 15–24. https://doi.org/10.5392/IJoC.2015.11.4.015

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. https://doi.org/10.1017/S0272263120000108

Liversidge, G. B. (2000). The role of closed captioning in second language acquisition [Doctoral dissertation]. Temple University.

Majuddin, E. (2020). Incidental and intentional acquisition of multiword expressions from audio-visual input: The effects of typographically enhanced captions and repetition [Doctoral dissertation, Victoria University of Wellington]. http://hdl.handle.net/10063/9126

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. Studies in Second Language Acquisition, 43(5), 985–1008. https://doi.org/10.1017/S0272263121000036

Matielo, R., & Collet, T. (2013). The effects of interlingual and intralingual subtitles on vocabulary learning by Brazilian EFL learners: An exploratory study. Intercâmbio, 27, 83–99.

Matielo, R., Oliveira, R. P. de, & Baretta, L. (2018). Intralingual subtitles, interlingual subtitles, and L2 vocabulary: Developments from an exploratory study. Acta Scientiarum. Language and Culture, 40, 36570. https://doi.org/10.4025/actascilangcult.v40i1.36570

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

Mousavi, F., & Gholami, J. (2014). Effects of watching flash stories with or without subtitle and reading subtitles on incidental vocabulary acquisition. Procedia - Social and Behavioral Sciences, 98, 1273–1281. https://doi.org/10.1016/j.sbspro.2014.03.543

Nguyen, C., & Boers, F. (2019). The effect of content retelling on vocabulary uptake from a TED Talk. TESOL Quarterly, 53(1), 5–29. https://doi.org/10.1002/tesq.441

Nguyen, C.-D. (2017). Fostering incidental vocabulary uptake from audio-visual materials: The role of text comprehension [Doctoral dissertation]. Victoria University of Wellington.

Pasban, M. A., Forghani, M., & Nouri, A. (2015). The effects of using English captions on Iranian intermediate EFL students learning of phrasal verbs. Journal of Language Teaching and Research, 6(2), 373. https://doi.org/10.17507/jltr.0602.18

Pattemore, A., & Montero Perez, M. (2022). Learning of Spanish subjunctive through original version audio-visual input. International Conference of the Spanish Society for Applied Linguistics (AESLA), Gran Canaria, Spain.

Pattemore, A., & Muñoz, C. (2020). Learning L2 constructions from captioned audio-visual exposure: The effect of learner-related factors. System, 93, 102303. https://doi.org/10.1016/j.system.2020.102303

Pattemore, A., & Muñoz, C. (2022). Captions and learnability factors in learning grammar from audio-visual input. The JALT CALL Journal, 18(1), 83–109. https://doi.org/10.29140/jaltcall.v18n1.564

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. TESOL Quarterly, 53(4), 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. Studies in Second Language Acquisition, 40(3), 551–577. https://doi.org/10.1017/S0272263117000407

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

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. Studies in Second Language Acquisition, 42(3), 525–549. https://doi.org/10.1017/S027226311900055X

Raine, P. (2012). Incidental learning of vocabulary through subtitled authentic videos [Master's Thesis]. University of Birmingham.

Rasizadeh, M. J., & Naeini, J. (2016). The effect of interest-based movie clips on motivation and oral proficiency of intermediate Iranian EFL learners. International Journal of Modern Language Teaching and Learning, 1(6).

Rodzuan, N. A. M. (2017). The effects of movie subtitles on phrasal verb recognition [Master's Thesis]. University of Malaya.

Roohani, A., & Rabiei, S. (2011). Effect of watching movies with bimodal, standard and no-subtitles on incidental learning of formal and informal vocabulary. Journal of Research in Applied Linguistics, 2(2), 100–118.

Sadowska, A. (2015). Learning English vocabulary from film audio description: A case of Polish sighted students. Roczniki Humanistyczne, 63(11), 101–123. https://doi.org/10.18290/rh.2015.63.11-7

Shahid, S. H. & Zuraina Ali. (2017). Effects of video-podcasts on listening comprehension of Saudi EFL learners. European Journal of English Language Teaching, 2(4), 169–194. https://doi.org/10.5281/ZENODO.891143

Shen, C. D. (1991). The effects of captioning on listening comprehension of English as a second language in a computer-based interactive videodisc system [Doctoral dissertation]. University of Northern Colorado.

Sirmandi, E. H., & Sardareh, S. A. (2016). The effect of BBC world clips with and without subtitles on intermediate EFL learners' vocabulary development. Malaysian Online Journal of Educational Sciences, 4(4).

Tsai, F.-H. (Will). (2010). Integrating feature films with subtitles to enhance the listening comprehension of students attending college in Taiwan [Doctoral dissertation]. Alliant International University.

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

Weyers, J. R. (1999). The effect of authentic video on communicative competence. The Modern Language Journal, 83(3), 339–349. https://doi.org/10.1111/0026-7902.00026

Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? Studies in Second Language Acquisition, 42(3), 599–624. https://doi.org/10.1017/S0272263120000029

Wong, P. (2010). The effectiveness of learning vocabulary through English news subtitles: A case study with Hong Kong ESL learners (pp. 991031046169703400, b45167448) [Master of Arts, The University of Hong Kong]. https://doi.org/10.5353/th_b4516744

Xue, X. (2017). Effects of different types of annotations on college students' foreign language learning in the synchronous multimodal computer-mediated communication environment [Doctoral dissertation]. Florida State University.

Yang, H.-C., & Sun, Y.-C. (2013). It is more than knowledge seeking: Examining the effects of OpenCourseWare lectures on vocabulary acquisition in English as a foreign language (EFL) context. Computer Assisted Language Learning, 26(1), 1–20. https://doi.org/10.1080/09588221.2011.624523

Yuksel, D., & Tanriverdi, B. (2009). Effects of watching captioned movie clip on vocabulary development of EFL learners. The Turkish Online Journal of Educational Technology, 8(2).

Curriculum Vitae

Name:	Dru Michael Sutton
Post-Secondary Education and Degrees:	The University of Western Ontario, London, Ontario, Canada 2018-2021 B.A. The University of Western Ontario, London, Ontario, Canada 2021-2023 M.A.
Honors and Awards:	Western Entrance Scholarship 2021-2023 Canada Graduate Scholarship – Master's (CGS-M) 2022-2023 Ontario Graduate Scholarship 2022-2023 Ontario Graduate Scholarship 2023-2024
Related Work Experience:	Research Assistant University of Western Ontario 2020-2021