Studies in Second Language Acquisition (2026), **48**, 571–595
doi:10.1017/S0272263126101612

STATE OF THE SCHOLARSHIP

The effects of audiovisual input on second language learning: *A meta-analysis*‡

Dru Sutton and Stuart Webb

Faculty of Education, University of Western Ontario, London, Canada
Corresponding author: Dru Sutton; Email: dsutton6@uwo.ca

(Received 09 January 2025; Revised 09 July 2025; Accepted 27 January 2026)

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) and assessed the effects of audiovisual input on language learning using a within-group (pre-post) meta-analytic approach. Fifteen moderator variables were included in the analysis. Results showed that a) there was a small effect of audiovisual input on L2 learning (g = .89); 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).

Keywords: meta-analysis; audiovisual input; viewing; video

Introduction

Audiovisual input is a form of multimodal input (i.e., input that uses multiple senses) 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 Video Home System (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

‡This article has been updated since original publication. A notice detailing the change has been published.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

(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 (e.g., subtitles or 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., 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. However, researchers have repeatedly advocated the value and use of audiovisual input with and without captions for L2 learning (e.g., Peters & Webb, 2018; Rodgers & Webb, 2011; Webb, 2015; Webb & Nation, 2017; Webb & Rodgers, 2009a, 2009b). Moreover, studies show that not all learners may use captions (Kam et al., 2020). For example, while watching television there are often no captions available. Some streaming platforms, such as YouTube, may have limited availability of captioning. Despite the lack of captioning, learners may still engage in viewing of audiovisual input without captions due to interest or the desire to learn. Other learners may intentionally turn off captions to challenge themselves by relying only on their listening skills. 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). Investigating the extent to which audiovisual input without on-screen text contributes to L2 learning may help inform learners and educators on what benefits or drawbacks turning off on-screen text may have.

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 previously conducted meta-analyses by Montero Perez et al. (2013) and Kurokawa et al. (2024) examined the effects of captioning on vocabulary learning, to the best of our knowledge this is the first meta-analysis conducted to assess the effects of audiovisual input alone. Quantifying the extent to which audiovisual input without on-screen text contributes to L2 learning would allow us to inform learners and teachers how much can be learned with captions turned off. A moderator analysis may also reveal how different approaches to learning through audiovisual input may improve learning and what areas of L2 learning may see the greatest improvements. 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, and listening proficiency). This was done by taking a within-group approach, comparing the learning gains between pre- and post-tests without factoring in between-group differences.

This meta-analysis focuses specifically on the effects of audiovisual input without on-screen text on L2 learning outcomes. The dependent variables analyzed included vocabulary, grammar, pronunciation, speaking, and listening proficiency, as these represent core linguistic competencies. These domains were chosen based on their relevance to the type of input (i.e., the potential to be learned) and their pedagogical significance in L2 learning. The independent variable was operationalized as

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

audiovisual materials that present simultaneous auditory and dynamic visual content without accompanying textual support (e.g., videos, films, or TV programs with no on-screen text). Studies involving captioned input were included only if data from non-captioned conditions could be extracted. Videos with minimal, nontranscriptive on-screen text (e.g., TED Talks) were included, as the text did not serve as a primary source of input.

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., 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) was not considered within the scope of this study unless it was explicitly stated that dynamic imagery was used. Reading studies that included imagery (e.g., Pellicer-Sánchez, 2022) were also not included. Since many studies compare on-screen text to unenhanced audiovisual input, we converted the unenhanced audiovisual input groups' pre-test to post-test scores into relative gains in the following literature review. This approach allows us to interpret results based on within-group change, without relying on between-group comparisons, which are of limited relevance to the context of this study.

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% and 11.29% of unknown words on meaning recall and meaning recognition tests, respectively. Meanwhile, Peters (2019) investigated the effects of viewing an English documentary on L1 Dutch speakers' 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 pretreatment 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

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

(–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 the Level Two 1994 National Spanish Exam, which is a standardized test created by the American Association of Teachers of Spanish and Portuguese. 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. Nonetheless, research indicates that viewing audiovisual input can have 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).

Research has also examined the effects of audiovisual input on L2 grammar acquisition (e.g., Cintrón-Valentín & García-Amaya, 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 TV show. Utilizing the same audiovisual material, Pattemore and Muñoz (2022) found gains from pre-test to 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 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.

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). Previous meta-analyses found age influenced L2 learning, with Kurokawa et al. (2024) reporting the greatest gains with primary students, while de Vos et al. (2018) found university

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

students made the most gains. The contrasting results suggest further investigation of this variable would be valuable.

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). In their meta-analysis on captioned audiovisual input, Kurokawa et al. (2024) found that there was no statistically significant difference; however, intermediate learners appeared to benefit more than other proficiency levels. The present meta-analysis may allow for a better understanding of whether proficiency moderates L2 learning through unenhanced audiovisual input.

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. Nonetheless, the extent to which the learning target affects learning gains has not been assessed previously and may provide insight into which areas of L2 learning audiovisual input may be most effective in improving.

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., 2022a; Puimège & Peters, 2019), or the effects of viewing several videos 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 (i.e., the target audience is L1 users). In contrast, educators may use language-focused input that has a pedagogical focus aimed

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

at explicitly helping viewers acquire some aspect of a second language (i.e., the target audience is L2 users). While language-focused videos are deliberately designed for L2 learners, non-language–focused videos are not created with L2 learners in mind. 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).

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. While no studies have examined how length of input affects learning, studies vary in the amount of input participants are exposed to. For example, Peters and Webb (2018) had participants watch a single documentary, while Muñoz et al. (2021) had them watch 24 episodes of a series. Meaning-focused vocabulary learning studies indicate that gains may be larger when learners are exposed to greater amounts of written input (e.g., Webb & Chang, 2015). Thus, it would be useful to examine whether viewing time moderates learning through exposure to audiovisual input.

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., 2022a; Weyers, 1999) or productive which involve using the target language (e.g., Majuddin, 2020; Nguyen & Boers, 2019). Tests may also be categorized as aural/oral or written tests. Input-modality – test-modality congruency, which builds upon Transfer Appropriate Processing theory (Morris et al., 1977), 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). However, recent findings in the context of audiovisual input (Kurokawa et al., 2024) report minimal differences between test modalities. Given these mixed results, it is important to examine whether test type moderates the observed effect sizes, as accurate measurement of learning gains is essential for understanding the impact of different treatments.

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?

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

Methodology

Literature search

First, an intensive search of several databases was conducted to identify all relevant literature. The databases searched were: PsycINFO, 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 included studies.

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 (pronunciation, grammar acquisition, vocabulary, speaking, listening). 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 were 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 posttreatment test was conducted then the scores of the tests were averaged if they were conducted at the same time. If 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 participants' 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.

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

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 above-mentioned 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 effect sizes (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) were excluded from analysis. Similarly, Wisniewska and 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. A PRISMA diagram (Page et al., 2021) providing more information about the literature search and screening can be found in Supplementary Materials.

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 Supplementary Information. Following Li (2016), all subgroups with a sample size smaller than $k < 3$ were excluded so that contrasts could be more meaningfully made. As Borenstein et al. (2009) note, small sample sizes provide limited statistical power and interpretive reliability. Furthermore, all studies that did not report data required for inclusion in a given moderator analysis were excluded from that analysis. The coding sheet is available for download in Supplementary Materials.

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

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

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 (beginner, A1–A2; intermediate, B1–B2; intermediate-advanced, B2–C1; advanced, C1–C2). Where levels were reported but not based on standardized tests, the researchers deferred to the judgments of the authors or teachers involved in assessing participants' levels if there was justification for this classification (e.g., a structured curriculum with a level test at the start of the term). 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 75 effect sizes, 71 of these targeted English learning while 4 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), multiword 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 multiword 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

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

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 (e.g., Cintrón-Valentín & García-Amaya, 2021, which used videos designed to teach verb tenses) and "n" for being non-language–focused videos (e.g., Peters & Webb, 2018, which used a BBC documentary on economics). 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 videos that were produced and focused on informing about a specific topic with a non-language focus were 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. These categories are based on the approximate length of a typical TV show (around 20 minutes). To account for shorter and longer videos, a 10-minute buffer was added in each direction. This approach also helped achieve more balanced sample sizes across groups.

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, and 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

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

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 respond orally to either written or auditory input. Studies that used mixed tests (i.e., they used each type of modality at different points) were coded as mixed. 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. Like the previous moderator variable, this moderator analysis was concerned with each individual test type and therefore did not average effect sizes across tests.

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 to clarify categorization, all disagreements in coding were resolved and the remaining studies were checked to ensure consistency.

Data analysis

Pooling of effect sizes and moderator analyses were conducted using the R packages meta (Schwarzer, 2023) and metafor (Viechtbauer, 2022). The meta package requires precalculated 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 most studies. Nonetheless, there were still many studies that used an audiovisual treatment group, and the within-group approach may have helped garner a large sample of primary studies since comparison groups could be included based on their pre- and post-test scores alone.

Calculating effect size

This study used Hedges' *g* as the measure for effect size. Hedges' *g* applies a correction to Cohen's *d*, which helps correct for an overestimation of Cohen's *d* for studies with small

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

sample sizes (Borenstein & Hedges, 2019; Cohen, 2013). For the formulae for calculating effect sizes, and standard errors, please see Supplementary Materials.

As within-group meta-analyses tend to overestimate the size of the effect, Plonsky and Oswald (2014) suggest that effect sizes are interpreted as d = .60 being small, d = 1.00 being medium, and d = 1.40 being a large effect. Since Hedges' g simply adds a correction factor to Cohen's d, Hedges' g may be interpreted using the same benchmarks.

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). Borenstein et al. (2009) argue that this approach is good for meta-analyses in educational fields because populations vary considerably (e.g., in age, education, L1). 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 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 that a random-effects model should be compared against 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).

Results

The analyses revealed a significant effect of audiovisual input on language learning, g = 0.89 (95% CI [.69, 1.09]), which is a small-to-medium effect following the benchmarks proposed by Plonsky and Oswald (2014). This indicates that audiovisual input has an overall small 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 0.868 (CI 95% [.842, .891]). This value suggests that 86.8% 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).

A funnel plot was created in order to assess the potential presence of a publication bias. The funnel plot suggests a potential small studies effect, with studies that have

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

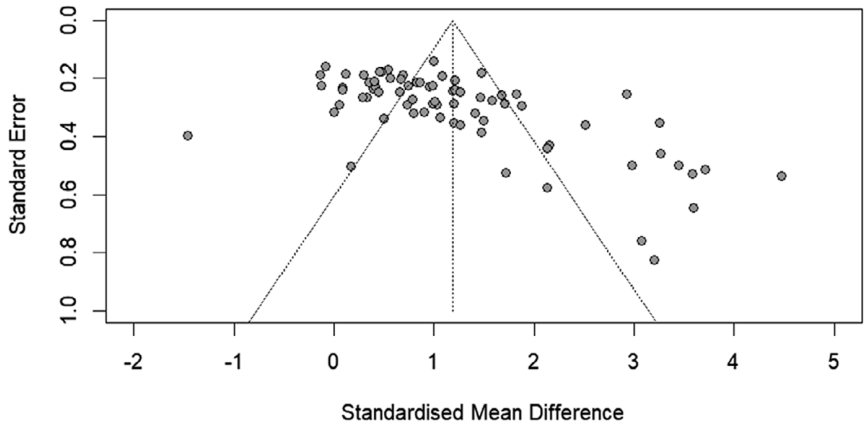


Figure 1. Funnel plot.

greater ES also having greater standard errors. An Egger's test (Egger et al., 1997) was conducted and revealed significant asymmetry (intercept: 5.421, 95% CI [3.69, 7.15], t = 6.14, $p < .00$). This may be attributed to a publication bias or to the real heterogeneity among studies. These findings are also complicated by the increased variance found in within-groups comparisons (Plonsky & Oswald, 2014). Additionally, a moderator analysis on publication status (see Supplementary Materials) was conducted and showed no significant effect of publication status on effect size ($p = .66$).

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.

Age

Age did not significantly affect the extent to which audiovisual input promoted learning ($Q = .33$, $p = .57$). A small effect was found for adult learners ($g = .88$), with a relatively narrow CI (95% CI [.66, 1.09]) suggesting a precise estimate of the ES. For young learners, the medium effect ($g = 1.06$) was associated with a larger CI (95% CI [.29, 1.83]), indicating greater uncertainty which was likely due to the sample size for young learners being relatively low ($k = 6$). The substantial overlap between groups' CIs supports the interpretation that age may not moderate the effectiveness of audiovisual input.

Language proficiency

There was not a significant difference ($Q = 6.04$, $p = .11$) between participants' language proficiency and the effects of viewing audiovisual input on L2 learning. For intermediate learners ($g = .99$, 95% CI [.53, 1.44]) and advanced learners ($g = .93$, 95% CI [.69, 1.17]) there was a small effect, bordering on medium. However, there was no effect for intermediate-advanced learners ($g = .52$, 95% CI [.17, .80]). Results for beginner learners were not precise, making it difficult to determine a positive or negative effect

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

(95% CI [–1.00, 3.48]). The substantial overlap among the CIs of all three groups reinforces the *Q*-test finding that differences between groups were not significant.

First language

The *Q*-test indicated significant differences between first languages ($Q = 17.38$, $p = .02$). There was a medium effect found for Vietnamese ($g = 1.19$, 95% CI [.85, 1.53]), with a narrow CI suggesting a precise estimate of the effect. A small effect was found for Dutch ($g = .72$, CI 95% [.37, 1.07]) and Chinese ($g = .96$, 95% CI [.29, 1.62]), with Dutch having a more precise estimate based on a narrower CI. Results from mixed L1s ($g = .52$, 95% CI [.28, .77]) and Korean ($g = .54$, 95% CI [.12, .95]) showed no effect, and wider relative CIs, indicating more uncertainty. Arabic, Japanese, and Spanish/Catalan produced very wide CIs that spanned zero, suggesting a possible null effect.

Target language

The *Q*-test result examining differences between the target language subgroups was not significant ($Q = 1.21$, $p = .27$). English as a target language had a small effect ($g = .87$, 95% CI [.66, 1.08]), with a narrow CI suggesting an accurate estimation of ES. There was a large and significant effect for non-English target languages ($g = 1.27$, 95% CI [.15, 2.38]); 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. The CIs of both groups overlapped, with interval of English fully nested within that of non-English target languages, indicating there may be no difference between groups.

Learning target

Results from the *Q*-test suggest that there was no significant difference between learning targets ($Q = 8.18$, $p = .23$). Audiovisual viewing had a medium effect on the learning of single words with a narrow CI ($g = 1.07$, 95% CI [.76, 1.38]). In the learning of listening skills ($g = .93$, 95% CI [.34, 1.52], $p = .01$) and mixed single- and multiword units ($g = .89$, 95% CI [.17, 1.61]), a small effect was found with wide CIs, making the magnitude of the true effect unclear. No effect was found for speaking, and the CI approached zero, suggesting the possibility of a null effect ($g = .47$, 95% CI [.02, .93]). Learning targets of multiword units, grammar, and pronunciation all had relatively large CIs spanning zero. This is likely due to the small sample sizes of these categories.

Multiple vs. single sessions

There was no significant difference between single or multiple treatment sessions ($Q = .04$, $p = .84$). There was a small and significant effect of multiple sessions ($g = .92$, 95% CI [.51, 1.33], $p < .00$), and single sessions ($g = .87$, 95% CI [.65, 1.10], $p < .00$), and both estimates had relatively narrow and overlapping CIs suggesting comparable effects.

Language-focused video

There was no significant difference between whether the audiovisual input was language-focused or not ($Q = 2.60$, $p = .11$). The effects of viewing language-focused audiovisual input were significant and produced a large effect ($g = 1.61$, 95% CI [.30,

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

2.93]) 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 = .83, 95% CI [.63, 1.04]). The large CI for language-focused audiovisual input introduces considerable uncertainty, making it difficult to determine potential group differences.

Audiovisual input type

There was a significant difference between the effects of different types of audiovisual input (Q = 25.28, p = .00). There were medium effects found for documentaries (g = 1.21, 95% CI [.45, 1.97]) and TED Talks (g = 1.23, 95% CI [.91, 1.56]). While both CIs suggest precise estimates, the CI for TED Talks was slightly narrower. Results for language-focused videos are the same as the previous subgroup analysis. There was no effect found for movies (g = .53, 95% CI [.32, .74]) or TV series (g = .51, 95% CI [.19, .83]). The estimate for TV series suggests greater variation in gains than for movies. Mixed video types had a small effect (g = .63, 95% CI [.08, 1.18]); however, the CI was wide and all ESs came from the same primary study. Therefore, these findings should be interpreted cautiously. Two types of audiovisual input had insignificant p-values and wide confidence intervals that spanned zero: animation (p = .28) and lectures (p = .12).

Viewing time

Viewing time had no significant effect on language learning (Q = 2.98, p = .23). A medium effect was found for short (g = 1.03, 95% CI [.23, 1.83]) and medium viewing times (g = 1.03, 95% CI [.70, 1.36]), while long viewing times had a small effect (g = .62, 95% CI [.22, 1.01]). All CIs suggest relatively precise estimates but have substantial overlap, suggesting that outcomes between these groups may be comparable.

Viewing plus

There was not a significant difference between whether or not viewing was supplemented with other activities (Q = 2.32, p = .13). Viewing plus yielded a medium effect (g = 1.09, 95% CI [.72, 1.46]), while viewing alone had a small effect on language learning (g = .76, 95% CI [.53, 1.00]). Despite relatively narrow CIs, there was a large degree of overlap, reinforcing the Q-test finding that there is little difference in effect.

Written vs. oral/aural tests

There was no significant difference found between written and aural/oral tests (Q = .93, p = .63). There was a medium effect found for aural/oral tests (g = 1.02, 95% CI [.76, 1.28]) and a small effect found for written tests (g = .87, 95% CI [.58, 1.16]). The mixed test group had a lower estimated effect, but with a larger CI (g = .79, 95% CI [.29, 1.30]). The substantial overlap of CIs aligns closely with the findings of the Q-test.

Productive vs. receptive tests

There was no significant difference between whether a test was receptive, productive, or mixed (Q = 2.87, p = .24). There was a small bordering on medium effect for receptive

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

tests (g = .99, 95% CI [.60, 1.38]) and a small effect for productive tests (g = .90, 95% CI [.71, 1.09]). CIs for both groups indicate precise estimates, with a high degree of overlap suggesting there is little difference in the effect measured by the different test types. The mixed group had a substantially smaller estimated effect, with a very large CI (g = .59, 95% CI [.10, 1.09]).

Discussion

This meta-analysis aimed to quantify the extent to which viewing audiovisual input contributes to L2 learning. Results yielded a Hedges' g of 0.89, 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 small effect. These results suggest that there was a clear and significant improvement in L2 learning after viewing audiovisual input. However, it is important to note that the high heterogeneity estimates (I^2 = 86.8%) indicate that there is substantial variation across studies. 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 using a meta-analytic approach. We hope the findings increase acceptance of audiovisual input as a useful 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).

Since the within-groups meta-analytic approach is unable to account for any potential testing effect, two separate analyses were conducted using a subset of the studies included in this meta-analysis. The first was a between-groups meta-analysis of those studies which did include a test-only condition, in which the control group completed pre- and post-tests without exposure to any L2 audiovisual input (k = 15). In this analysis, an effect size of g = .65 (95% CI [.32, .98]) was found. Under Plonsky and Oswald's (2014) area-specific guidelines, this constitutes a small effect (approaching medium, g = .70). Although the CI indicates a lack of precision in this estimate, it also suggests that audiovisual input is effective to some extent. The second analysis was conducted to estimate the extent to which a test effect may exist. This analysis used a within-groups approach and looked at only test-only conditions' pre-post contrasts (k = 19). The effect size for this analysis was g = .15 (95% CI [–.08, .38]), and the CI was relatively narrow, suggesting a precise estimate. Interpreted together, the ES and CI suggest there is likely little to no testing effect. The magnitude between the within-groups and between-groups meta-analyses is of the same degree (i.e., small approaching medium). While the test effect cannot be accounted for in the within-groups analysis, the similar magnitude of both analyses suggests that this is an accurate estimation of the pooled effect.

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. Kurokawa et al. (2024) reinforce the benefits of using captioned audiovisual input, but similar to Montero Perez et al. (2013), only tell us the extent to which captions are useful compared to uncaptioned audiovisual input. 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 both

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

Montero Perez et al. (2013) and Kurokawa et al. (2024) found captions to contribute to greater learning than unenhanced audiovisual input.

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, 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). 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. The effect sizes between the intermediate ($g = 1.19$, 95% CI [.63, 1.75]) and advanced ($g = 1.11$, 95% CI [.74, 1.39]) groups 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 they had not yet passed the threshold into advanced proficiency; 2) the smaller sample size skewed results. Disregarding this group, there appeared to be a more pronounced effect for beginner learners, with the effect size decreasing as proficiency increased. One possible explanation is that learners with higher proficiency levels were shown videos that were substantially more difficult (e.g., more difficult subject matter, faster rate of speech, more abstract content). Overall, these results suggest that learners at both intermediate and advanced levels may also benefit from audiovisual input.

The first language of learners had a significant influence on the effect size. Based on transfer theory (Odlin, 2003), 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, and the CI for English L2s was fully nested in that of non-English L2s, suggesting comparable effects. Although both ESs indicate the presence of some effect, the estimate for non-English L2s was relatively imprecise due to the small sample size ($k = 4$). 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 videos. However, it should be noted that three L2 targets had wide CIs spanning zero, making the estimates for these groups unreliable. The lack of precision for these groups is likely due to the relatively small sample sizes of these groups (grammar, $k = 6$; multiword, $k = 8$; pronunciation, $k = 3$). The estimated effect sizes for all three of these areas were small. However, the wide CIs suggest these results should be interpreted with caution. Although most studies focused on vocabulary learning (single items, $k = 38$; multiword items, $k = 8$; mixed, $k = 5$), there were still a large number of studies focusing on other areas of L2 learning ($k = 24$). For

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

this reason, although individual subgroup estimates may lack precision, 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 indicates that further research in these areas is warranted.

Treatment-related variables also provided some insight into the current state of audiovisual input research. Studies which used multiple ($k = 29$) and single treatment ($k = 46$) sessions had narrow and overlapping CIs, suggesting comparable learning gains. 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. However, since this meta-analysis did not consider delayed post-tests, no definite conclusion can be drawn. 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. However, non-language–focused videos being nested within the CI of language-focused videos suggests a lack of precision. This prevents us from drawing a definite conclusion. The larger effect size for language-focused videos may be due to the focus on learning language features, slower speech rates, and controlled use of language to make the input more comprehensible for learners. However, the wide CI suggests a lack of certainty in the true magnitude of language-focused videos.

Among audiovisual types, there was no effect of movies and TV series on learning, while there were medium 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 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). According to the dual coding theory (Paivio, 2010), this may help alleviate some of the cognitive burden in processing faster speech. Among the other video types, language-focused videos had the largest effect among all the groups, while animations and lectures had very large CIs spanning zero, perhaps because of their sample sizes ($k = 5$, $k = 3$, respectively; see Table 1). 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 postviewing activities to enhance learning and understanding (Webb, 2015).

Viewing time was not a significant factor in language learning. There was a smaller effect of longer viewing than 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

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

Table 1. Moderator analysis.

			95% CI				Q tests	
Variables	*k*	*g*	LL	UL	*p*	I^2	*Q*	*p*
Age							.33	.57
Adult	69	.88	.66	1.09	.00	.88		
Young	6	1.06	.29	1.83	.02	.68		
Language proficiency							6.04	.11
Beginner	3	1.24	–1.00	3.48	.14	.83		
Intermediate	29	.99	.53	1.44	.00	.92		
Intermediate-Advanced	8	.52	.17	.88	.01	.70		
Advanced	12	.93	.69	1.17	.00	.84		
L1							17.38	**.02**
Arabic	5	1.19	–.88	3.27	.19	.95		
Multi	10	.52	.28	.77	.00	.54		
Chinese	11	.96	.29	1.62	.01	.90		
Korean	4	.54	.12	.95	.03	.36		
Japanese	3	.50	–1.74	2.75	.44	.84		
Dutch	6	.72	.37	1.07	.00	.66		
Vietnamese	11	1.19	.85	1.53	.00	.64		
Spanish/Catalan	4	.37	–.50	1.24	.27	.79		
L2							1.21	.27
English	71	.87	.66	1.08	.00	.86		
Non-English	4	1.27	.15	2.38	.04	.89		
Learning target							8.18	.23
Single word	38	1.07	.76	1.38	.00	.86		
Grammar	6	.64	–.49	1.77	.20	.95		
Speaking	5	.47	.02	0.93	.05	.51		
Listening	10	.93	.34	1.52	.01	.90		
Multiword	8	.59	–.01	1.25	.08	.87		
Pronunciation	3	.65	–2.73	4.02	.50	.92		
Single + Multi	5	.89	.17	1.61	.03	.57		
Multiple vs. single session							.04	.84
Single	46	.87	.65	1.10	.00	.84		
Multiple	29	.92	.51	1.33	.00	.90		
Language-focused							2.60	.11
Language-focused	5	1.61	.30	2.93	.00	.84		
Non-language–focused	70	.83	.63	1.04	.03	.95		
Video Type							25.28	**.00**
Animation	5	.88	–1.08	2.85	.28	.92		
Documentary	14	1.21	.45	1.97	.01	.91		
Movies	15	.53	.32	.74	.00	.73		
TV Series	16	.51	.19	.83	.00	.81		
Lecture	3	1.31	–.79	3.41	.12	.87		
TED Talk	12	1.23	.91	1.56	.00	.95		
Various	3	.63	.08	1.18	.04	.16		
Language-focused	5	1.61	.30	2.93	.03	.95		
Viewing Time							2.98	.23
Short	10	1.03	.23	1.83	.02	.83		
Medium	20	1.03	.70	1.36	.00	.82		
Long	24	.62	.22	1.01	.01	.88		
Viewing Plus							2.32	.13
Yes	30	1.09	.72	1.46	.00	.89		
No	45	.76	.53	1.00	.00	.85		
Test Type							.93	.63
Written	63	.87	.58	1.16	.00	.91		
Aural/Oral	45	1.02	.76	1.28	.00	.87		
Mixed	8	.79	.29	1.30	.01	.79		

(*Continued*)

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

Table 1. (*Continued*)

Variables	*k*	*g*	95% CI		*p*	I^2	*Q* tests	
			LL	UL			*Q*	*p*
Test Type							2.87	.24
Receptive	47	.99	.60	1.38	.00	.91		
Productive	64	.90	.71	1.09	.00	.89		
Mixed	5	.59	.10	1.09	.03	.73		

CI = Confidence interval; LL = lower limit; UL = upper limit.

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 effect of viewing plus was medium ($g = 1.09$), while viewing alone had a small effect ($g = .76$). While there is some degree of overlap in CIs, the effects of viewing plus are demonstrably larger than viewing without additional activities. The more pronounced ES of viewing plus may be related to the theory of desirable difficulty (Bjork, 1994), which suggests that learning may be enhanced when greater cognitive effort is required. Supplementing audiovisual input with activities that reinforce L2 learning may increase cognitive effort, which in turn may lead to greater 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 noted 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 there was not a significant difference between these two formats and CI overlapped substantially. However, there was a slightly larger effect found for aural/oral tests than for written tests. This suggests that while both tests can detect learning that has occurred, congruent tests may reveal slightly larger learning gains. However, the input-modality – test-modality effect was not as pronounced as the effect found by Jelani and Boers (2020). These results also aligned with Kurokawa et al. (2024), who found no difference between test types. 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, $g = .99$; productive, $g = .90$). 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.

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

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.

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. Although there has been increased research on grammar learning through audiovisual input (e.g., Cintrón-Valentín & García-Amaya, 2021; Pattemore & Muñoz, 2022), there is a need for 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 and older 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. The present study also revealed little research on the effects of different cognitive abilities on learning through audiovisual input, and few studies that investigated the effects of L1. More research on these areas could help show how audiovisual input can be used more effectively.

Furthermore, there are few studies that investigate audiovisual input with a true control (e.g., Dang et al., 2022a; 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).

A final limitation is that this study addressed the issue of ES dependency by averaging multiple ESs originating from the same study. While this helps manage dependency, it reduces the number of ESs contributing to the pooled estimate, resulting in some data loss. This could be addressed in a future study using a multilevel, multivariate meta-analysis (Norouzian & Bui, 2024), which may also address the high heterogeneity observed in the present study.

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 also found across different L2 learning targets, for learners at different proficiency levels, and across different regions. Findings also suggest that learners may benefit most 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.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0272263126101612.

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

Acknowledgments. This article is based on the first author's MA thesis, which was funded by the Social Sciences and Humanities Research Council (SSHRC) of Canada. We would like to extend a special thanks to the following researchers for providing extra data that helped in conducting this study: Yanxue (Michelle) Feng, Rafael Matielo, Anastasia Pattemore, Elke Peters, Mark Feng Teng, and Danijela Trenkic. Correspondence concerning this article should be addressed to Dru Sutton, University of Western Ontario, Faculty of Education, 1137 Western Road, London, ON N6G 1G7, Email: dsutton6@uwo.ca

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

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.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about Knowing*. The MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001

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.

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

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* Routledge. https://doi.org/10.4324/9780203771587

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

Dang, T. N. Y., Lu, C., & Webb, S. (2022a). 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

Dang, T. N. Y., Lu, C., & Webb, S. (2022b). 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

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

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

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

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* (3rd 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

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

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

Kurokawa, S., Hein, A. M., & Uchihara, T. (2024). Incidental vocabulary acquisition through captioned viewing: A meta-analysis. *Language Learning*, lang.*12697*. https://doi.org/10.1111/lang.12697

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

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. http://hdl.handle.net/10063/9126

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/10.1177/0267683211015797

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

Norouzian, R., & Bui, G. (2024). Meta-analysis of second language research with complex research designs. *Studies in Second Language Acquisition*, *46*(1), 251–276. https://doi.org/10.1017/S0272263123000311

Odlin, T. (2003). Cross-linguistic influence. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (1st ed., pp. 436–486). Wiley. https://doi.org/10.1002/9780470756492.ch15

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. https://doi.org/10.1136/bmj.n71

Paivio, A. (2010). Dual coding theory and the mental lexicon. *The Mental Lexicon*, *5*(2), 205–230. https://doi.org/10.1075/ml.5.2.04pai

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

Pellicer-Sánchez, A. (2022). Multimodal reading and second language learning. *ITL - International Journal of Applied Linguistics*, *173*(1), 2–17. https://doi.org/10.1075/itl.21039.pel

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

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

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

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

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

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

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

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]. ProQuest Dissertations & Theses.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73. http://llt.msu.edu/vol14num2/sydorenko.pdf

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 [Doctoral dissertation, Alliant International University]. San Diego ProQuest Dissertations & Theses.

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

Viechtbauer, W. (2022). *Metafor* (Version 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

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press

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

Webb, S., & Nation, I.S.P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.

Webb, S. & Rodgers, M.P.H. (2009a). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407–427.

Webb, S. & Rodgers, M.P.H. (2009b). The vocabulary demands of television programs. *Language Learning*, *59*(2), 335–366.

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

Cite this article: Sutton, D., & Webb, S. (2026). The effects of audiovisual input on second language learning: *A meta-analysis*. *Studies in Second Language Acquisition,* 48: 571–595. https://doi.org/10.1017/S0272263126101612

https://doi.org/10.1017/S0272263126101612 Published online by Cambridge University Press