Effectiveness of Subtitles in L2 Classrooms: A Meta-Analysis Study

Hind M. Alotaibi, Hassan Saleh Mahdi and Deema Alwathnani

Abstract: The use of subtitles in language-learning environments has received considerable attention. Nevertheless, conclusive findings have yet to be reached regarding the effectiveness of integrating subtitles into L2 classrooms. This meta-analysis examined the overall effectiveness of subtitles to enhance learners' L2. A total of 26 experimental studies conducted between 2010 and 2022 were analyzed to examine the performance of L2 classroom learners based on integrating subtitles as a teaching/learning tool, compared to learners who did not use subtitles. A meta-analysis investigated how effect sizes varied depending on L2 skills (listening, speaking, reading, writing, grammar, vocabulary, and pronunciation), institutional level (school, university, and language institute), participants' native language (L1), language proficiency, and implementation mode. The results indicated that integrating subtitles into L2 classrooms have a medium effect size ($d = 0.69$), and the effect sizes of different variables affecting the use of subtitles in language learning are reported. Several pedagogical implications are discussed in light of the findings, which can help promote the successful integration of subtitles in L2 classrooms.

Keywords: audiovisual translation (AVT); subtitles; meta-analysis; L2 learning; pedagogy

1. Introduction

Audiovisual translation (AVT) is “the translation of recorded audiovisual material” [1] (p. 2). This term denotes the transfer of verbal language of audiovisual media. Generally, this term is used interchangeably with “screen translation” and “multimedia translation,” which include subtitling, dubbing, and audio description. Subtitling is the “translation of the spoken source language text of an audiovisual product—generally film dialogs—into a written text, which is superimposed onto the image of the original product, usually at the bottom of the screen” [2] (p. 63).

Integrating subtitles in L2 teaching and learning began decades ago, with several studies investigating the impact of using subtitles on students’ L2 learning. In some studies, subtitles were effective in enhancing L2 skills, especially vocabulary learning [3,4], while other studies have reported contradictory results. Such inconclusive outcomes draw attention to further investigation in this area and examination of the overall effect size of various variables related to the impact of subtitles on L2 learning. A scoping review and meta-analysis of 34 studies synthesized the effects of different types of video captioning (in L2) and subtitling (L1) on L2 vocabulary acquisition [3]. The results revealed that these studies examined productive knowledge more than receptive knowledge. One meta-analysis found that viewing any sort of captioned or subtitled video positively affected L2 vocabulary acquisition. Similarly, a meta-analysis of 10 studies conducted between 2010 and 2016 examined the impact of captions and subtitles on EFL learners’ vocabulary comprehension [4]. The results indicated an overall positive effect of captions and subtitles on vocabulary comprehension among adult EFL learners. The focus of these reviews and meta-analyses was the use of subtitles to learn L2 vocabulary comprehension. Other L2 skills have not yet been fully examined in meta-analyses.
Meta-analyses conducted in the area of audiovisual translation and their impact on language learning are limited, focusing mainly on the use of subtitles to learn L2 vocabulary comprehension; hence, thus far, the other L2 skills have not been fully examined. Additionally, key factors affecting the impact of subtitles on L2 learners—such as the participants’ L1, language proficiency level, institutional level, gender, and language skills—have been ignored in previous meta-analyses. Therefore, a more comprehensive analysis covering all the variables is clearly needed. This study aimed to explore subtitles’ impact on L2 learning by conducting a meta-analysis of primary experimental studies. This meta-analysis examined the effects of several variables that might impact the effectiveness of using subtitles in L2 classrooms, such as L2 skills (listening, speaking, reading, writing, grammar, vocabulary comprehension, and pronunciation), participants’ institutional level (school, university, and language institute students), their native language (L1), their L2 proficiency level, and subtitle implementation mode (viewing and/or creating subtitles). The findings are expected to shed some light on subtitles’ effectiveness on L2 learning and provide suggestions to promote the successful integration of such tool in L2 teaching and learning.

The following section explores the literature on subtitling in L2 contexts and its impact on various L2 skills.

2. Literature Review
2.1. Audiovisual Translation

AVT has gained popularity in the field of language teaching and learning, particularly in classroom and teacher-mediated settings. Presumably, AVT facilitates the learning experience, and motivates and entertains the classroom. According to the definition of [5], AVT focuses on the practices, processes, and products involved in—or resulting from—the transfer of multimodal and multimedia content across languages and/or cultures. AVT can take the form of subtitling, dubbing, or audio descriptions. According to Henrik Gottlieb [6], subtitling is “the rendering in a different language of verbal messages in filmic media, in the shape of one or more lines of written text, presented on the screen in synch with the original verbal message” (p. 86); conversely, in dubbing, information is transmitted via the same semiotic channels in the source and target texts. Thus, subtitling refers to the conversion of oral elements in audiovisual materials into written text. Although subtitles and captions are occasionally used interchangeably, many researchers make a distinction between the two. Subtitles involve translating the dialog’s original language into another language; that is, they are interlingual. However, captions are intralingual because the original dialog is converted into text in the same language.

The use of subtitles in L2 learning began decades ago, and an overall positive impact on various language skills has been demonstrated. One early study [7] by Vanderplank (1988) highlighted the effectiveness of teletext captions for English as a Foreign Language (EFL) students. Since then, several studies have demonstrated that captioned and subtitled materials can improve language learners’ receptive skills and cultural awareness [7–12]. Several scholars have considered subtitles as valuable resources to promote bilingualism and multilingualism [13–18]. They have acknowledged the potential of subtitles in language learning contexts because of their role as a rich and motivating pedagogical tool, and encouraged educators to integrate subtitles into their instruction to promote L2 skills and cultural awareness. Borghetti and Lertola [19] argued that several opportunities for intercultural learning occur in subtitle tasks, even when teacher mediation is limited, which can promote learner autonomy. However, some challenges are associated with integrating subtitles into the L2 classroom, such as cost, maintenance, inconvenience, and fear of technology [18]. Furthermore, the successful integration of subtitles requires careful planning and adequate selection of audiovisual tasks. The primary limitations of subtitling tasks in L2 learning are the lack of ready-made materials and technology-related issues [20]. Additionally, the design and production of tasks can be time-consuming, and selecting materials with appropriate language and content for subtitling tasks can be challenging.
In the subsequent section, we explore some variables affecting subtitle integration in L2 classrooms related to this meta-analysis.

2.2. Variables Affecting Subtitle Integration in L2 Classrooms

Several variables influence subtitles’ effective use in language learning contexts, six of which are discussed in this meta-analysis—namely, context, language skills, institutional level, native language, language proficiency, and the way subtitles are employed.

The context wherein a language is learned can be broadly categorized into second languages (SL) and foreign languages (FL). FL is defined as a language that is not used outside the classroom in everyday communication [21]. Foreign languages are typically taught as school subjects to communicate with foreigners or read printed materials in the L2. By contrast, SL is broadly defined as “any language learned after one has learnt one’s native language” [21] (p. 514). However, in contrast to FL, the term refers to a language that extensively used in a particular country or region, though several language users may not be native speakers. Subtitles can be used to learn a language in both FL and SL contexts [22–29]. However, prior studies have yielded inconsistent results. Several studies have reported positive effects in FL settings [25,26] and SL settings [23], whereas others have reported contradictory results i.e., with FL learners [27–29] and SL learners [22], suggesting that further research is required to investigate the effect of context (EFL or ESL) on the use of subtitles in L2 classrooms.

The combination of different modes—such as text, audio, video, and images—significantly impacts learning. According to the cognitive theory of multimedia learning [30], learners draw connections between aural and visual information as long as temporal closeness exists; therefore, learning improves when information is processed both orally and in writing. Several researchers have investigated subtitles’ effects on language productive skills (writing and speaking) [31,32] and receptive language skills (listening and reading comprehension) [25–28]. Moreover, the impact of subtitles on other L2 skills—such as vocabulary comprehension, grammar, and pronunciation—has been assessed. However, other studies have reported inconsistent results.

One study [31] assessed the English speaking and writing skills of 52 Korean EFL college students—by asking them to watch American television episodes using either Korean or English subtitles—and found that viewing subtitles had no significant impact on production skills. However, Valizadeh [32] reported contradictory results. The researcher investigated the impact of L1 subtitles, L2 captions, and no subtitles on Turkish EFL learners’ abilities to comprehend and speak about movies. The findings revealed that captions helped students acquire greater content than subtitles and no subtitles.

In terms of vocabulary learning, Pujadas and Muñoz [33] explored the effects of extensive viewing of a TV series on L2 vocabulary learning using captions and subtitles. The results revealed that participants learnt vocabulary in all conditions. However, the subtitles group outperformed the caption group in meaning recall.

The comparison between captions and subtitles was also the focus in Pujadas and Muñoz [34]. Students were divided into the following two groups: One watched a TV series with L1 subtitles and the other with L2 captions. Reportedly, L1 subtitles enhanced content comprehension more than L2 captions; prior vocabulary knowledge helped promote their success. These findings seemingly contradict the results of other studies wherein captions were demonstrably more effective than subtitles [35,36]. Further research is required to compare the effects of subtitles and captions on different language skills (reading, listening, speaking, writing, vocabulary comprehension learning etc.) and to determine the sizes of these effects.

Learners’ native language influence how they acquire a second language [37–39]. For example, the linguistic distance between the L1 and L2 has been shown to be a significant factor in L2 acquisition [39]. Such interference is caused by varied factors, such as similarities and differences between L1 and L2 structures, orthography, and consonant clusters. The similarities between L1 and L2 seem to facilitate the acquisition of L2. In contrast, larger
distances from the L1 the L2 correlate with lower degrees of L2 learnability [39]. Exploring the literature over the past decades on the use of subtitles in L2 classrooms has revealed that subtitles are used to enhance L2 learning among students in various native languages. Some studies [27, 40–48] have examined how the learning performance of Persian EFL students can be improved by utilizing subtitles. Chinese students learning English as an FL were participants in some studies [49–52]. Different studies with students speaking other native languages, including Arabic [25, 28], Brazilian [53], Dutch [54], Italian [55], Indonesian [56], Norwegian [23], Spanish [29], Thai [57], Turkish [32], and Urdu [22], have been conducted in other EFL environments. Few studies have been conducted on native English speakers learning other languages, such as Dutch [26] and Italian [24]. However, these studies have yielded different outcomes. Thus, a deeper analysis of the effect size of the participants’ L1 on successfully implementing subtitles in L2 classrooms is clearly needed.

The participants’ L2 proficiency levels affect their performance in subtitle-enhanced tasks. Studies conducted over the last few decades on the impact of subtitles on students’ L2 learning have involved participants with various proficiency levels—beginner, intermediate, and advanced. Most studies recruited students with intermediate-level language proficiency (e.g., [27, 29, 41, 42, 44, 45, 51–53]). The results of some studies have indicated that the use of subtitles enhances participants’ L2 skills among intermediate learners. Other studies have demonstrated limited impacts [41, 42]. Students with advanced L2 abilities were involved in one study [25] that reported a positive effect of integrating subtitles on students’ listening comprehension and encouraged their use among advanced-level L2 learners. Some studies [24, 26, 55] involving beginner students reported inconsistent results. For instance, in Lertola’s study [24], the impact of subtitles on L2 beginner learners was not significant. The researchers attributed their findings to cognitive load theory and the limited nature of working memory [30, 58], suggesting that input must be carefully tailored according to the learners’ proficiency level to “encourage the activation of both aural and visual processing mechanisms” [56] (p. 616). In other words, the nature of the tasks and proficiency of L2 learners influences the impact of subtitles on L2 learners. Numerous researchers [56, 58, 59] have agreed that the positive effect of subtitles on L2 learning is subject to carefully selected level-appropriate materials used in the activities. Danan [18] argued that for beginner students, the material must be carefully adapted to their level and should include familiar phrases that can be activated and reinforced by audiovisual elements. Although participants’ proficiency level is, seemingly, a critical factor in successfully integrating subtitles in L2 classrooms, meta-analysis studies conducted previously in this area have overlooked this aspect.

Most studies investigating the impact of subtitles on L2 have involved students viewing subtitled materials. Few studies have examined the effects of students’ involvement in generating subtitles on their L2 learning. Technological advances have made multimedia editing applications increasingly accessible to students. These apps are now affordable, user-friendly, and provide various features—such as editing, adding texts, and sharing. Consequently, these apps have become increasingly popular in technology-enhanced L2 classrooms, wherein students can work on project-based multimedia activities. However, experimental studies exploring the potential of allowing L2 learners to create subtitles are limited. Alabsi [25] investigated the effect of adding subtitles to videos using video editing apps on EFL students’ listening comprehension and found that such tasks significantly improved students’ listening comprehension ability, allowing them to break down a clip, pause to consider its meaning, and translate each phrase in a way that matches the image and is meaningful. Alabsi argued that subtitling could be a simple but influential tool for producing impactful multimedia listening tasks [25:1196], and concluded with a call for further research in this area by comparing the impact of subtitling tasks on students’ L2 learning.

Exploring the literature revealed variations in studies investigating the impact of subtitles on L2 learning; notably, the results seem inconclusive. However, systematic
reviews and meta-analyses in this field are scarce and outdated. All meta-analyses studies tackling this topic seem to focus on one or two L2 kills (mostly vocabulary), overlooking other key variables, such as the participants’ L1 proficiency level, academic level, and L2 skill investigated. This research gap calls for a recent, more profound, and comprehensive analysis. Thus, the present meta-analysis aimed to investigate the effects of integrating subtitles as a pedagogical tool in L2 teaching and learning settings by answering the following questions:

RQ1. What is the overall effect size of integrating subtitles as a pedagogical tool for L2 learning?

RQ2. What is the impact of different variables on the effect of using subtitles on L2 learning?

The research methodology and procedures are described next.

3. Methodology

A meta-analysis was conducted to explore the impact of subtitles on L2 learning. Meta-analysis involves the integration of quantitative results based on several studies on a specific topic to obtain a quantitative effect size using multiple calculations and findings [60]. Such studies are usually conducted to provide evidence supporting or rejecting specific hypotheses. This study followed the steps suggested by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). The meta-analysis was conducted in the following four stages: (1) literature search, (2) application of inclusion criteria, (3) coding, and (4) calculation of effect size. These stages are described in the following subsections.

3.1. Literature Search

To identify appropriate studies for analysis, a search was conducted using the Educational Resources Information Center (ERIC), Web of Science, SCOPUS, and Google Scholar. Overall, 909 articles matched the following key terms: audio-visual translation, subtitles, subtitling and language learning, multimedia and language learning, and subtitles and L2. The abstracts of these studies were screened to determine whether they could be included in the analysis. The search had no time limit, and any study published up to 14 September 2022, was included. Several studies were excluded because they did not fulfill the criteria for this meta-analysis. Finally, 26 articles [23–30,32,42–48,50–57] were included in the analysis.

3.2. Study Inclusion Criteria

The inclusion criteria were created to identify suitable studies for this analysis. Each study fulfilled the following criteria: (1) It was experimental or quasi-experimental in nature. (2) Integrating subtitles as a tool was an essential variable. (3) Participants in the experimental group used subtitles as a tool for learning L2. (4) Means, standard deviations, and the number of participants in each group were reported.

The process of inclusion of the studies involved the following steps: (1) The total number of studies obtained in the search was 909 studies, then 13 studies were removed because they were duplicates. (2) The titles and abstracts of the remaining 896 studies were checked to ensure that they were related to the topic under investigation. This step led to excluding 809 studies because they were not directly related to the topic. (3) The criteria of inclusion mentioned above were applied on the remaining 87 studies. In total, 26 studies were found to be eligible for inclusion in the analysis. The process of identifying relevant studies included in the meta-analysis is presented in Figure 1.
3.3. Coding of Study Characteristics

Several variables affect the use of subtitles in L2 learning. In this meta-analysis, six variables were examined.

Context refers to the linguistic role played by a target language in the linguistic environment wherein the learners are situated. Context is categorized into Foreign Language (FL), that is, not spoken in the learner’s immediate environment, and Second Language (SL), that is, spoken in the learner’s immediate environment. Overall, 22 studies in this meta-analysis were conducted in the FL context. Only four primary studies have been conducted in an SL context. The target languages were English, Dutch, Chinese, and Italian. However, English was the dominant target language in 23 primary studies.

Language skill(s) refer to language components or constructs under investigation, including listening (7 studies), vocabulary comprehension (13 studies), reading comprehension (2 studies), speaking (2 studies), and writing (1 study), and only one study investigated a combination of these L2 skills.

Institutional level refers to the participants’ academic stage and was categorized into the following three types: schools (primary or secondary), universities, and others. In this analysis, four primary studies were conducted in schools, that is, any kind of education before college. Sixteen primary studies were conducted among university...
students. University refers to the type of education that occurs in colleges and universities, where students attend different courses to obtain higher education degrees in different disciplines. The five primary studies in the category ‘others’ involved language institutes in which learners of any age attended short FL/SL courses to improve their language proficiency.

Native language refers to the participants’ native tongue or first language L1. The L1 examined in the studies included in this meta-analysis were Persian (9 studies), Chinese (4 studies), Arabic, and English (2 studies each), and one study was found for Brazilian, Chinese, Dutch, Indonesian, Italian, Norwegian, Spanish, Thai, Turkish, and Urdu.

Language proficiency refers to participants’ L2 proficiency as described by researchers in primary studies. They were classified into the following four categories: beginner (4 studies), intermediate (14 studies), advanced (1 study), and three studies involving learners with mixed proficiency. In four other studies, the proficiency levels of the participants were not reported.

Method of implementation refers to how subtitles were implemented and categorized into subtitle-viewing and subtitle-making. Subtitle-viewing refers to the type of study wherein participants were asked to view the material with subtitles created by someone else, while subtitle-making refers to the type of study wherein participants were instructed to create subtitles to the videos. Only one primary study involved subtitle-making, whereas 24 studies investigated subtitle-viewing.

All aforementioned variables were examined and coded for use as moderators.

3.4. Effect Size Calculation

In this meta-analysis, an effect size based on Cohen’s $d$ was used to represent the impact of subtitles on language learning. Cohen’s $d$ is defined as the difference between two means divided by the standard deviation of the data and measures the difference between two independent means in terms of standard deviation units [61]. Cohen’s $d$ was calculated using basic descriptive statistics (sample size, mean scores, and standard deviations of control and treatment groups). This meta-analysis applied the following standard benchmarks [62] to interpret effect size: small effect size, $d = 0.20$; moderate effect size, $d = 0.50$; and large effect size, $d = 0.80$. Moreover, this study followed a random-effects model. This is the case in numerous studies in the field of language learning because the effect size varies from one study to another [61]. This study followed Higgins and Thompson [63] to interpret this statistic: $I^2 = 25\%$ (small heterogeneity), $I^2 = 50\%$ (medium heterogeneity), and $I^2 = 75\%$ (large heterogeneity). Data were analyzed using Comprehensive Meta-Analysis (CMA [64]).

3.5. Publication Bias

We employed the trim-and-fill procedure developed by Duval and Tweedie to assess the effect of publication bias on the overall effect size [65]. The trim-and-fill method removes the most extreme small-effect studies from the positive side of the funnel plot using an iterative process. With each iteration, the effect size was recalculated until the funnel plot became symmetrical under the new effect size [65]. Based on the outcomes of this method, we found two studies to be outliers on the negative side (i.e., [49,66]) which were removed from the publication bias calculations. As presented in Figure 2, the overall effect size decreased slightly. These studies were added when calculating the effect sizes because the change was slight and not significant.
4. Results

Analyses of these studies were conducted in various categories. The overall effect of subtitles on language learning is discussed, and variables that may affect the use of subtitles in language learning were analyzed.

4.1. Overall Effect of Subtitles on Language Learning

This meta-analysis included 26 studies and yielded 34 effect sizes. The overall results of these studies are summarized in Table 1.

Table 1. Overall effect size of the impact of subtitles on L2.

<table>
<thead>
<tr>
<th>N</th>
<th>k*</th>
<th>Confidence Intervals</th>
<th>p-Value</th>
<th>Q-Value</th>
<th>Df</th>
<th>I²</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>34</td>
<td></td>
<td>0.371</td>
<td>1.021</td>
<td>0.000</td>
<td>341.516</td>
<td>33</td>
</tr>
</tbody>
</table>

* k = number of effect sizes calculated.

Analysis was performed using the standardized mean difference as the outcome measure, as presented in Table 1 (d). A random effects model was used to fit the data. Table 1 indicates that the overall effect size of subtitles’ impact on L2 was medium (d = 0.69). Additionally, the Q-test for heterogeneity and I² statistics are reported. Overall, 26 studies with 34 effect sizes were included in the analysis. Confidence intervals ranged from 0.371 to 1.021. Therefore, the average outcome differed significantly from zero. According to the Q-test, the true outcomes were heterogeneous (Q = 341.516, p = 0.000), suggesting that subtitles are more effective than learning without subtitles. The forest plot presented in Figure 3 provides a context for the analysis.
4.2. Effect Sizes of Various Variables

Variables that may affect the use of subtitles in language learning were also analyzed. The first variable was the context in which the study was conducted. The context was classified into two categories: SL and FL. The results in Table 2 indicate that the use of subtitles have a medium effect size ($d = 0.63$, CI = [0.526, 0.735]) for studies investigating FL learning. However, the use of subtitles had no significant effect size ($d = 0.083$, CI = [−0.206, 0.372]) for studies conducted in a SL environment. The difference between categories was significant ($p = 0.000$).

Table 2. Effect size of context.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>$k^*$</th>
<th>$D$</th>
<th>Confidence Intervals</th>
<th>$p$-Value</th>
<th>Q-Value</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>22</td>
<td>29</td>
<td>0.632</td>
<td>0.526 – 0.738</td>
<td>0.000</td>
<td>329.324</td>
<td>32</td>
</tr>
<tr>
<td>SL</td>
<td>4</td>
<td>5</td>
<td>0.083</td>
<td>−0.206 – 0.372</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$k^*$ = number of effect sizes calculated.

The second variable was language skills. They were classified into six sub-categories (i.e., listening, speaking, reading comprehension, writing, vocabulary, and mixed skills i.e., two or more language skills). The results presented in Table 3 indicate that the use of subtitles have a large effect size for studies that used subtitles for writing ($d = 3.25$, CI = [2.47, 4.43]), listening ($d = 0.88$, CI = [0.686, 1.088]), and speaking ($d = 0.92$, CI = [0.403,1.43]) and a medium effect size for studies that used subtitles for reading comprehension ($d = 0.66$, CI = [−0.206, 0.372]). However, the use of subtitles had a small effect size for studies that used subtitles for vocabulary ($d = 0.29$, CI = [0.149, 0.445]), and no
significant effect for mixed skills \((d = 0.022, CI = [-0.387, 0.431])\). The differences between the categories were significant \((p = 0.001)\).

Table 3. Effect size of different language skills.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>k*</th>
<th>D</th>
<th>Confidence Intervals</th>
<th>p-Value</th>
<th>Q-Value</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Limit</td>
<td>Upper Limit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening</td>
<td>7</td>
<td>8</td>
<td>0.88</td>
<td>0.686</td>
<td>1.088</td>
<td>0.000</td>
<td>28</td>
</tr>
<tr>
<td>Speaking</td>
<td>2</td>
<td>2</td>
<td>0.92</td>
<td>0.403</td>
<td>1.434</td>
<td>0.000</td>
<td>28</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>2</td>
<td>5</td>
<td>0.66</td>
<td>0.437</td>
<td>0.888</td>
<td>0.000</td>
<td>28</td>
</tr>
<tr>
<td>Writing</td>
<td>1</td>
<td>1</td>
<td>3.25</td>
<td>2.474</td>
<td>4.432</td>
<td>0.000</td>
<td>28</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>13</td>
<td>17</td>
<td>0.29</td>
<td>0.149</td>
<td>0.445</td>
<td>0.000</td>
<td>28</td>
</tr>
<tr>
<td>Mixed skills</td>
<td>1</td>
<td>1</td>
<td>0.022</td>
<td>-0.387</td>
<td>0.431</td>
<td>0.000</td>
<td>28</td>
</tr>
</tbody>
</table>

* k = number of effect sizes calculated.

The third variable was learners’ educational level. Context was classified into the following four categories: schools, universities, language institutes, and mixed. The results in Table 4 indicate that the use of subtitles had medium effect sizes for studies conducted in schools \((d = 0.592, CI = [0.346, 0.838])\), universities \((d = 0.541, CI = [0.541, 0.657])\), and language institutions \((d = 0.687, CI = [0.687, 1.011])\). Regarding mixed institutions, the effect size was large \((d = 0.88, CI = [-0.000, 1.77])\), but only one study was reported. Therefore, the results in this category should be considered cautiously. The difference between categories was significant \((p = 0.000)\).

Table 4. Effect size of institutional level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>k*</th>
<th>D</th>
<th>Confidence Intervals</th>
<th>p-Value</th>
<th>Q-Value</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Limit</td>
<td>Upper Limit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>4</td>
<td>6</td>
<td>0.592</td>
<td>0.346</td>
<td>0.838</td>
<td>0.000</td>
<td>28</td>
</tr>
<tr>
<td>University</td>
<td>16</td>
<td>21</td>
<td>0.541</td>
<td>0.425</td>
<td>0.657</td>
<td>0.000</td>
<td>30</td>
</tr>
<tr>
<td>Language institutes</td>
<td>5</td>
<td>6</td>
<td>0.687</td>
<td>0.363</td>
<td>1.011</td>
<td>0.000</td>
<td>28</td>
</tr>
<tr>
<td>mixed institutions</td>
<td>1</td>
<td>1</td>
<td>0.889</td>
<td>-0.000</td>
<td>1.779</td>
<td>0.000</td>
<td>28</td>
</tr>
</tbody>
</table>

* k = number of effect sizes calculated.

The fourth variable was the learners’ L2 proficiency levels. This variable is classified into the following four categories: beginner, intermediate, advanced, and mixed. Some studies did not report learners’ L2 proficiency levels and were, thus, categorized under the “not available” (NA) category. The results presented in Table 5 indicate that subtitles had no significant effect size for beginner learners \((d = 0.125, CI = [-0.178, 0.429])\), a small effect size for the mixed levels \((d = 0.38, CI = [-0.086, 0.693])\), a medium effect size for intermediate learners \((d = 0.60, CI = [0.470, 0.729])\), and a large effect size for advanced learners \((d = 1.51, CI = [1.00, 2.02])\). The difference between categories was significant \((p = 0.000)\).

Table 5. Effect size of participants’ L2 proficiency level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>k*</th>
<th>D</th>
<th>Confidence Intervals</th>
<th>p-Value</th>
<th>Q-Value</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Limit</td>
<td>Upper Limit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beginners</td>
<td>4</td>
<td>4</td>
<td>0.125</td>
<td>-0.178</td>
<td>0.429</td>
<td>0.000</td>
<td>29</td>
</tr>
<tr>
<td>Intermediate</td>
<td>14</td>
<td>21</td>
<td>0.600</td>
<td>0.470</td>
<td>0.729</td>
<td>0.000</td>
<td>29</td>
</tr>
<tr>
<td>Advanced</td>
<td>1</td>
<td>1</td>
<td>1.51</td>
<td>1.00</td>
<td>2.02</td>
<td>0.000</td>
<td>29</td>
</tr>
<tr>
<td>Mixed</td>
<td>3</td>
<td>3</td>
<td>0.38</td>
<td>0.086</td>
<td>0.693</td>
<td>0.000</td>
<td>29</td>
</tr>
<tr>
<td>NA</td>
<td>4</td>
<td>5</td>
<td>0.63</td>
<td>0.392</td>
<td>0.886</td>
<td>0.000</td>
<td>29</td>
</tr>
</tbody>
</table>

* k = number of effect sizes calculated.
The fifth variable was the native language used in the subtitles. The studies included in this meta-analysis used 13 languages for their subtitles. The results in Table 6 indicate that subtitles had a large effect size for Persian \((d = 1.044, CI = [0.834, 1.25])\), Indonesian \((d = 3.25, CI = [2.47, 4.43])\), and Turkish \((d = 3.085, CI = [2.43, 3.73])\); a moderate effect size for Arabic \((d = 0.690, CI = [0.244, 1.04])\), English \((d = 0.699, CI = [0.351, 1.58])\), and Norwegian \((d = 0.637, CI = [−0.437, 1.71])\); an extremely small but not significant effect size for Thai \((d = 0.022, CI = [−0.387, 0.431])\), Chinese \((d = 0.397, CI = [0.244, 0.550])\), and Urdu \((d = 0.294, CI = [−0.425, 1.01])\); and a negative effect size for Spanish \((d = −0.319, CI = [−1.76, 0.128])\), Italian \((d = −0.543, CI = [−1.26, 0.076])\), Brazilian \((d = −0.655, CI = [−1.47, 0.176])\), and Dutch \((d = −0.299, CI = [−0.800, 0.203])\) subtitles. Noteworthily, the target languages differed among the studies (e.g., Chinese, Dutch, English, and Italian). The difference between categories was significant \((p = 0.000)\).

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>k*</th>
<th>D</th>
<th>Confidence Intervals</th>
<th>p-Value</th>
<th>Q-Value</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persian</td>
<td>9</td>
<td>9</td>
<td>1.044</td>
<td>0.834 1.258</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>4</td>
<td>10</td>
<td>0.397</td>
<td>0.244 0.550</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arabic</td>
<td>2</td>
<td>2</td>
<td>0.690</td>
<td>0.332 1.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thai</td>
<td>1</td>
<td>1</td>
<td>0.022</td>
<td>−0.387 0.431</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>1</td>
<td>2</td>
<td>−0.319</td>
<td>−0.766 0.128</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>2</td>
<td>2</td>
<td>0.699</td>
<td>0.351 1.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urdu</td>
<td>1</td>
<td>1</td>
<td>0.294</td>
<td>−0.425 1.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norwegian</td>
<td>1</td>
<td>1</td>
<td>0.637</td>
<td>−0.437 1.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>1</td>
<td>1</td>
<td>−0.543</td>
<td>−1.262 0.076</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazilian</td>
<td>1</td>
<td>1</td>
<td>−0.655</td>
<td>−1.47 0.167</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indonesian</td>
<td>1</td>
<td>1</td>
<td>3.25</td>
<td>2.477 4.433</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish</td>
<td>1</td>
<td>2</td>
<td>3.085</td>
<td>2.436 3.733</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>1</td>
<td>1</td>
<td>−0.299</td>
<td>−0.800 0.203</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\* k = number of effect sizes calculated.

The sixth variable investigated the effect size of the subtitles implementation method. This variable was classified into the following three categories: viewing, creating, and mixed. The results presented in Table 7 indicate a large effect size when learners were asked to create subtitles \((d = 1.51, CI = [1.00, 2.02])\), as was mixed \((d = 1.43, CI = [0.185, 1.88])\). However, the effect was medium \((d = 0.523, CI = [0.421, 0.625])\) when learners viewed materials that had already been subtitled by someone else. The difference between categories was significant \((p = 0.000)\).

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>k*</th>
<th>D</th>
<th>Confidence Intervals</th>
<th>p-Value</th>
<th>Q-Value</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viewing</td>
<td>24</td>
<td>32</td>
<td>0.523</td>
<td>0.421 0.625</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Making</td>
<td>1</td>
<td>1</td>
<td>1.510</td>
<td>1.00 2.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>1</td>
<td>1</td>
<td>1.435</td>
<td>0.185 1.88</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\* k = number of effect sizes calculated.

A discussion of these findings is presented subsequently.

5. Discussion

This study examined the impact of integrating subtitles into L2 classrooms. We analyzed 26 experimental studies conducted between 2010 and 2022 to investigate the overall effect size of integrating subtitles as a pedagogical tool for L2. Moreover, we examined how effect sizes varied according to the learners’ native language(s), language
skills, educational level, learners’ native language, their language proficiency, and the way subtitles were implemented. In response to the first research question related to the overall effect size of integrating subtitles as a pedagogical tool for L2 learning, the overall effect size of viewing the subtitles in L2 learning, the meta-analysis results revealed that the overall effect size of using subtitles on L2 learning revealed a medium effect size ($d = 0.69$), suggesting that subtitles positively impacts L2 learning, which coincides with previous meta-analyses of the impact of subtitles [3,4]. In Reynolds et al.’s study [3], for instance, the overall effect of subtitled video viewing on vocabulary acquisition was significant ($d = 0.874$), suggesting a positive impact. Nevertheless, noteworthy, such results must be considered with great caution “as the impacts of examined learning conditions may vary depending upon learning outcomes and learners’ proficiency levels” [52] (p. 538). Therefore, a deeper analysis was conducted to examine the effect sizes of various variables.

First, the effect of context on the use of subtitles in language learning was analyzed. The results indicated that the use of subtitles in FL classrooms revealed a medium effect size, while it was not significant in the SL context. One possible explanation for this result is that an FL setting exists inside the classroom, which mandates a degree of consciousness from the individual and the internal processes of learning carried to the learning task. Additionally, the impact of subtitles might be washed out in SL settings, wherein numerous opportunities for input exist.

The second variable examined was the language skill targeted; the findings revealed that subtitles revealed different effect sizes according to the L2 skill under investigation. Large effect sizes were found for listening, speaking, and writing skills, whereas medium effect sizes were found for reading comprehension. Unlike in some prior studies [3,4], the effect size of subtitles on vocabulary comprehension was small. Notably, the results for speaking and writing skills should be considered with caution, because only two studies were included for speaking [32,48] and one study for writing. This indicates that the limited number of studies conducted on speaking and writing might have impacted the effect size, considering the number of studies focusing on listening and vocabulary learning. This also draws attention to the need for further investigating the effect of using subtitles to enhance L2 speaking and writing skills, as studies in this area seem limited.

Regarding learners’ educational level, the results of this meta-analysis revealed that subtitles have a medium effect size for all levels, indicating that the use of subtitles can be effective for learners at all institutional stages. Students in schools, universities, and language institutes can benefit from subtitles in their learning, and educators at various academic levels are encouraged to integrate them into their instruction.

Learners’ L2 proficiency level was one of the variables examined in this meta-analysis. Notably, subtitles had a large effect size for advanced learners and a medium effect size for intermediate learners. However, no significant effect was found in beginner learners. Research suggests that low-proficiency students find it difficult to process multiple modalities simultaneously because of their limited working memory and cognitive capacity [30,38]. Cognitive load usually impedes beginners’ comprehension, particularly when incorporating multiple modalities [32,56]. These findings encourage instructors to consider their students’ proficiency levels and carefully design their tasks to accommodate students’ needs to ensure the successful integration of subtitles in L2 classrooms. However, the results for advanced learners should be interpreted with caution because only one study was conducted on advanced learners [25], which may have influenced the results. Further studies on the impact of subtitles on advanced students’ L2 abilities are needed.

The effects of the learners’ L1 were also examined in the meta-analysis. The results revealed a wide range of effect sizes depending on the participants’ L1, from large (e.g., Persian, Indonesian, and Turkish) to negative (e.g., Spanish, Italian, Dutch, and Brazilian). Some studies reported a medium effect size (e.g., Arabic, English, and Norwegian), whereas others reported a small effect size (e.g., Chinese and Urdu). These results should be interpreted with caution because most native languages were used in only one or two studies. The differences in some effect sizes can be attributed to the linguistic distance
between L1 and L2 e.g., Chinese and English, Urdu and English [37–39]. However, this assumption is contradicted by the large effect sizes for Persian and negative effect sizes for Dutch and Spanish. Therefore, further research is required to determine how linguistic distance influences subtitle use in L2 classrooms.

The last variable explored was the way subtitles were implemented, that is, whether students were asked to view them as opposed to whether they were asked to create them. The analysis results indicate that subtitles exhibit a large effect size when learners work on creating subtitles themselves or when combining the use and creation of subtitles. The effect size was medium when learners viewed only the subtitles without participating in their generation. According to Alabsi [25], the involvement of students in the creation of subtitles may yield significant outcomes for language learning, as it helps motivate learners to listen attentively, recognize unfamiliar words, and observe differences in usage. This is also supported by Vanderplank [7], who stated that viewing multimedia is predominantly a passive activity and that students’ language proficiency depends predominantly on productive practice, which can be promoted by subtitling tasks. However, because only one study wherein learners generated subtitles was included in the meta-analysis, these suggestions must be considered with caution. These findings also indicate that further empirical investigation should be conducted on the effects of subtitle-making on students’ L2 learning.

6. Conclusions

This meta-analysis aimed to investigate the overall effect size of integrating subtitles as a pedagogical tool for L2 learning and explored the impact of different variables affecting the integration of subtitles in L2 classrooms. The analysis included 26 experimental studies conducted between 2010 and 2022 that examined the performance of L2 learners. The meta-analysis examined how effect sizes varied depending on L2 skills (listening, speaking, reading, writing, grammar, vocabulary, and pronunciation), institutional level (school, university, and language institute students), participants’ native language (L1), language proficiency, and implementation mode. Results showed that subtitles had a medium effect on L2 learning, suggesting they can be an effective pedagogical tool. However, other variables may affect the successful integration of subtitles into L2 teaching and learning environments, which must be taken into consideration when interpreting these general findings. It was interesting to find that learners’ performance when subtitles were used for listening, speaking, and writing improved significantly. In the case of teaching reading comprehension, the effect was moderate. On the other hand, subtitles were found to have a small impact on vocabulary learning. According to these results, the effective use of subtitles in L2 classrooms depends on the language skill being taught.

Subtitles can also be effective depending on the proficiency level of learners in L2. L2 advanced and intermediate learners were more likely to benefit from subtitles than beginners. Because of their limited working memory and cognitive capacity, beginner L2 learners have difficulty processing multiple modalities simultaneously, thereby reducing subtitle effectiveness. Thus, findings suggest educators should consider learners’ proficiency levels when integrating subtitles into their lesson plans and carefully design tasks to accommodate their abilities and cognitive capacity. In addition, the effectiveness of subtitles was examined based on learners’ native languages. The results indicated different effect sizes for different L1s. The use of subtitles proved more effective among Persian, Indonesian, and Turkish speakers with a large effect size. Moderate effect of subtitles on Arab, English, and Norwegian speakers’ performance. A limited impact was reported when subtitles were used with speakers of Urdu and Chinese, whereas no effect was detected for Thai, Italian, Brazilian, or Dutch speakers. There are, however, a limited number of studies in some of these settings, which could affect the findings. These findings highlight a need for further research to explore the potential effects of linguistic distance and orthographic differences between L1 and L2.
Additionally, the results indicated that subtitles proved more effective when integrated into a foreign language learning context than second language learning, where the effect size was negative. These results can be attributed to the fact that FL classrooms require a degree of consciousness from the learners and the internal processes of learning carried into the learning task. This is unlike in SL settings, where the impact of subtitles might be washed out due to the numerous opportunities for input. In order to confirm these conclusions, however, further research is necessary.

In terms of learners’ institutional levels, the analysis revealed that subtitles were effective for students at all levels (including schools, universities, and language institutes). In light of these findings, subtitles can be integrated into L2 classrooms at all levels of education. Lastly, subtitle implementation methods in L2 classrooms may affect their effectiveness. Results suggest that learners may benefit more when they are involved in creating subtitles than when they only watch subtitled materials. Therefore, educators should consider designing tasks to move beyond passive viewing and encourage learners to produce their own subtitles instead of using ready-made materials. This recommendation should be considered with caution, however, since only one study in which students generated subtitles was included in the meta-analysis. Further studies on the effectiveness of integrating subtitling tasks in L2 classrooms is required.

Finally, in considering the findings of this study, it is imperative to keep in mind its limitations, which are described in the following section.

**Limitations and Suggestions for Future Studies**

This meta-analysis has certain limitations. It is important to note that due to the limited number of studies that measured some of the variables, the results should be interpreted with caution. The need for future meta-analyses that incorporate a larger number of studies on each variable is evident. Second, considering the target language used in subtitles is crucial. As a result of the limited number of studies in some languages, this variable has not received sufficient attention in this meta-analysis. The use of subtitles in different contexts and languages needs to be explored through experimental studies. Further research is needed on the effectiveness of integrating subtitling tasks in L2 classrooms by involving learners in producing subtitles rather than viewing ready-made meanings, as studies exploring such areas are scarce.

**Author Contributions:** Conceptualization, H.M.A. and H.S.M.; methodology, H.M.A. and H.S.M.; validation, H.S.M.; formal analysis, H.S.M.; investigation, H.M.A., H.S.M. and D.A.; resources, H.M.A., H.S.M. and D.A.; writing—original draft preparation, H.M.A., H.S.M. and D.A.; writing—review and editing H.M.A., H.S.M. and D.A.; visualization, H.S.M.; supervision, H.M.A. and H.S.M.; project administration, H.M.A.; funding acquisition, H.M.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Literature, Publishing, and Translation Commission, Ministry of Culture, Kingdom of Saudi Arabia, under [59/2022] as part of the Arabic Observatory of Translation, grant number 59/2022.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


3. Reynolds, B.L.; Cui, Y.; Kao, C.W.; Thomas, N. Vocabulary acquisition through viewing captioned and subtitled video: A scoping review and meta-analysis. *Systems* **2022**, *10*, 133. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.