Visual Language Input Improving Vocabulary Learning in College Students: A Comparison of the Effects of Audiovisual, Visual, and Orthographic Input

by
Ananya Arcot

A thesis presented to the Honors College of Middle Tennessee State University
in partial fulfillment of the requirements for graduation from the
University Honors College

Spring 2025

Thesis Committee:

Dr. Meghan Wendelken, Thesis Director

Dr. Kathryn Blankenship, Second Reader

Dr. Mary Evins, Committee Chair

Visual Language Input Improving Vocabulary Learning in College Students: A Comparison of the Effects of Audiovisual, Visual, and Orthographic Input

by
Ananya Arcot

APPROVED:

_____
Dr. Meghan Wendelken, Thesis Director
Professor, Health & Human Performance

_____
Dr. Kathryn Blankenship, Second Reader
Professor, Health & Human Performance

_____
Dr. Mary Evins, Thesis Committee Chair
Professor, History

Acknowledgments

Firstly, I would like to thank my family for supporting me during college and being there for me. This has certainly not been the easiest thing I have ever done, and I was so unfamiliar with research before I embarked on my thesis. Still, my family believed I could take on whatever I wanted.

I want to thank my thesis director, Dr. Meghan Wendelken, for advising, teaching, and inspiring me throughout my thesis. I would also like to thank my research assistants for helping and encouraging me, and my second reader, Dr. Kathryn Blankenship, for guiding and teaching me.

Thank you to all my professors from the Speech-Language Pathology & Audiology program. They have been highly dedicated to helping and caring about every student.

Lastly, thank you to my committee chair, Dr. Mary Evins, for advising me and giving me invaluable tips for my writing.

Abstract

Subtitles have been popularized recently with the rise of streaming services and short-form media, and evidence suggests that subtitles increase vocabulary learning. The study compares three conditions of watching a documentary—audiovisual (AV), visual and text (VT), and audiovisual and text (AVT)—to determine the highest incidental vocabulary acquisition, correlations between existing vocabulary skills and vocabulary learning, and perspectives of the documentary based on the three conditions. For the study, 44 college students were recruited, and a series of tests were given before and after participants watched a 25-minute selected documentary clip. Results show that participants gained the most scores under the VT condition. The AV condition shows the most significant score difference when comparing effect sizes. The relationship between prior knowledge and gain scores was positive but weak. Subjects under the VT condition were less likely to watch the documentary in the same condition than participants under different conditions.

Table of Contents

List of Tables

List of Figures

CHAPTER I

Subtitles in the Media Impacting the Input of Language Modalities

Introduction

Subtitles have become a staple in television shows and modern media. TikTok, YouTube, and Netflix shows are all popular mainstream services for which their use of subtitles has affected the modern media consumer and can potentially change language learning. When media are consumed, three channels of input modality are processed: auditory, visual, and audiovisual inputs. Auditory input and visual input, or listening and watching, have been proven to be great ways to learn language (Balci et al., 2020; Harji et al., 2010; Lertola, 2019; Malitesta et al., 2023; Matielo et al., 2015). Researchers have only recently investigated the effects of on-screen text, whether subtitled (i.e., target-language texts) or captioned (i.e., same-language subtitles), paired with audiovisual input. Evidence suggests that incorporating subtitles into lessons increases language comprehension (Zheng et al., 2022).

Further, many studies have found that pairing subtitles with video or audio input results in more efficient learning of a second language (Caruana, 2021; Teng, 2022; Zhang & Zou, 2022). One possible reason for this is that subtitles provide visual language input rather than the individual having to rely on processing only auditory input when hearing the second language spoken. However, very little is known about the effects of subtitles on language processing when watching media in one's native language. Therefore, this study aims to understand if providing visual linguistic input via subtitles improves vocabulary learning in one's native language.

As previously discussed, when watching television, language input can be provided in multiple modalities and combinations of those modalities, such as audiovisual, visual + text, and audiovisual + text. The audiovisual modality combines listening to the words and sounds played with watching the paired images on screen. The visual + text mode is the images displayed on-screen paired with the subtitles, usually on the bottom of the screen, without the listening/audio component (e.g., muted volume). Lastly, the audiovisual + text mode is the audiovisual component with subtitles on the screen.

There is evidence to support that when on-screen text is paired with the audiovisual mode of input, there is an increase in learning gains (Alotaibi et al., 2023; Kanellopoulou et al., 2019; Mohsen, 2016; Vandergrift, 2007; Wei & Fan, 2022). Documentaries, animations, sitcoms, news clips, etc., have been used to study on-screen language acquisition. However, most studies use documentaries because they "contain more imagery in close proximity to target words than narrative TV genres" (Peters et al., 2019, p. 1). Peters et al. (2019) evaluated the impacts of three different documentary viewing conditions on vocabulary learning for participants learning English as a second language (L2). The study included 142 participants in fifth grade or sixth grade who were English as a Foreign Language (EFL) learners and whose first language (L1) was Dutch. All participants had been receiving formal English instruction in school. Participants were randomly assigned to one of the following documentary viewing conditions: captions (English), L1 subtitles (Dutch), and no subtitles (the control group). All three of the conditions had the audiovisual modality. A list of 36 target vocabulary items was measured for their frequency and concreteness, and the participants within the three

conditions were given a vocabulary test with the target words listed. Results reported that the group with captions learned the most vocabulary terms in English when given written input in the English language. These results support the idea that providing visual input when learning new linguistic terms can increase language learning compared to receiving only auditory input for novel linguistic information.

A similar study including 26 L1 English speakers learning Russian as an L2 was conducted with participants viewing two to three popular Russian comedies under the following conditions: VAC (visual, audio, and captions), VA (audiovisual), and VC (video + captions) (Sydorenko et al., 2010). The English speakers were tested on Russian vocabulary (L2). A written recognition and aural recognition test were then conducted, followed by a translation test, a word knowledge test, and a final questionnaire. The results showed that the VAC group gained the highest scores, and the VA group was the second highest, showing that visual input via captions can increase language acquisition.

Most studies evaluating vocabulary learning through subtitles have focused on participants acquiring vocabulary in their second language (L2) (Black, 2021, p. 1). However, Zheng et al. (2022) "examined whether adding video and subtitles to an audio lesson facilitates its comprehension and whether the comprehension depends on participants' cognitive abilities…" (p. 1). Participants watched 16 lessons covering ecology, astronomy, geography, and chemistry and were tested on vocabulary before and after the lessons were shown. The lessons were watched in the following four conditions: audio, audio + video, audio + text, and audio + text + video. In addition, memory capacity, planning ability, and multitasking tasks were given to analyze comprehension performance. The results indicated "a significant difference in comprehension accuracy

between the audio + video and audio + text + video conditions" (Zheng et al., 2022, p. 9). There was approximately a 5-point difference in mean comprehension scores, with the audio + text + video condition being higher than the audio + text condition. Therefore, we have some initial evidence that subtitles can promote language learning in an individual's native language, not just when learning a second language.

The process of overshadowing should also be considered when thinking about the concept of input modalities and how they impact language learning. Overshadowing happens when one mode of input interferes with or hinders the processing of another mode of input. For example, verbal overshadowing is a psychological process where describing a previously seen face weakens the recognition of the face (Dodson et al., 1997). Auditory overshadowing is when auditory input overshadows visual input, meaning that the individual will preferentially process the auditory signal. Visual overshadowing is when visual input overshadows auditory input, meaning that the individual preferentially processes visual information over auditory information (Robinson & Sloutsky, 2007). Infants mostly experience auditory overshadowing, while young children experience visual and auditory overshadowing. However, adults primarily experience visual overshadowing (Robinson & Sloutsky, 2007, p. 1). This means that when receiving input in two modalities simultaneously, the visual mode of input would hinder or interfere with the auditory mode of input. When applying this to language processing while watching television, this suggests that adults would preferentially process visual linguistic input (i.e. subtitles) over auditory language input, and the visual input may interfere with or hinder the auditory processing. Under extended processing conditions, familiar audio input is less likely to interfere with the processing of visual

input (Robinson & Sloutsky, 2007). Watching a documentary would be considered an extended processing condition. Therefore, it is likely that adults would primarily attend to and process the visual input modality.

Adults receive linguistic input in both visual and auditory modalities when watching television. Therefore, given what we know about overshadowing, adults may preferentially process the visual modality of information over the auditory input, especially when watching television for extended periods. Using subtitles while watching television would provide a visual, linguistic modality to supplement the visual imagery/symbols on the screen. Therefore, it is possible that using subtitles can support vocabulary and language learning for adults when watching television in their native language.

It is also important to discuss differences in intentional vs. incidental vocabulary learning, which will play a significant part in this study. A study evaluating incidental vocabulary acquisition does not inform participants that there will be a vocabulary test, while a study evaluating intentional vocabulary acquisition informs participants that there will be a test before the experiment (Reynolds et al., 2022). While watching television, individuals are engaging in incidental vocabulary learning. Therefore, this study will evaluate participants' incidental vocabulary learning while watching a documentary under different conditions.

Learning how on-screen interaction and different input modes affect human language learning is pivotal to developing Speech-Language and Hearing Science. Social media and television are highly consumed by individuals daily and are typically watched in one's native language. It is possible that watching television or consuming modern

media (e.g., TikTok) under different conditions (e.g., auditorily, visually, or with subtitles) can impact the efficiency with which the linguistic information is processed. Gaining a better understanding of this information is critical to allow media platforms to provide their consumers with an ideal viewing condition (audio, video, or text). Additionally, this information can determine what viewing conditions might best support individuals with language processing impairments.

Thesis Statement

Visual input modalities show more vocabulary learning than aural modalities when consuming media. Prior studies related to the current one have shown that subtitles greatly affect learning a second language, demonstrating the impact of visual learning in adults. The significance of visual input is highlighted in the current study, displaying necessary evidence.

CHAPTER II

Methodology: Viewing a Documentary Using Different Input Modalities

To further study the problem, using university students as the study population, this study compares three conditions of watching documentaries—audiovisual (AV), visual and text (VT), and audiovisual and text (AVT)—to determine the highest incidental vocabulary acquisition, the correlation between existing vocabulary skills and vocabulary learning, and the differing perceptions of the documentary based on the three conditions. More precisely, the study examines which condition will result in the highest incidental vocabulary acquisition in young adults, how the vocabulary skills of participants will influence their ability to learn new vocabulary, and if there are any possible biases in watching the documentary under the three conditions. Thus, three research questions are being asked:

1. When watching a documentary in one's heritage language of English, which of the following conditions (audiovisual, visual + text, or audiovisual + text) might result in the highest incidental vocabulary acquisition among young adults ages 18-26?

    a. Audiovisual condition: images on screen paired with audio input

    b. Visual and text: images on screen paired with subtitles

    c. Audiovisual and text: images on screen paired with audio input and subtitles

Hypothesis: Based on previous research, higher incidental vocabulary acquisition will be presented under the audiovisual + text condition, followed by the audiovisual and visual and text conditions.

2. Is there a relationship between participants' vocabulary skills and vocabulary learning when watching a documentary across all three conditions?

Hypothesis: There will be a positive correlation between a participant's vocabulary skills and vocabulary learning. The higher the Peabody Picture Vocabulary Test – 5th Edition (PPVT-5; Dunn & Dunn, 2007) score, the more efficient the language learning will be; therefore, the higher the outcome scores from the post-test.

3. What are the participants' perceptions of the documentary while watching with different input modalities? More specifically, how do different input modalities impact young adults' perceptions of a documentary in the heritage language containing novel/unfamiliar vocabulary concepts?

Hypothesis: There will be biases presented in young adults when different conditions are displayed. Participants will find the documentary containing unfamiliar concepts and novel words easier to follow and more enjoyable when presented in the AVT condition.

Participants

Forty-four college students between the ages of 18 years and 36 years were recruited for initial data collection. Participants were primarily recruited from the Speech-Language Pathology and Audiology program under the Department of Health and Human Performance. Inclusionary criteria were the following: (a) the participant's primary or first language must be English; (b) participants must be unfamiliar with the content, language, and vocabulary of the documentary that is presented as evidenced by a score of <25% on the pretest; (c) participants must be enrolled as a college student; and (d) participants must have no diagnosis of speech, language, hearing, or reading impairments.

Subjects were recruited via social media posts (Instagram and iMessenger), flyers sent by email, in-class announcements, and word of mouth. Flyers were also attached to bulletin boards in the Alumni Memorial Gym and the MTSU Student Union. Additionally, many students from the Health and Human Performance department were recruited through this research study's connection with the department. A 10-dollar Amazon gift card was given to each participant upon completion of the study to increase the incentive to participate. The subjects were aware of the incentive before signing up for participation. Two research assistants helped with data collection and scoring.

Materials

Two ninth-generation iPads were used for participants to watch the documentary and respond to survey questions to gather demographic information (Appendix A) on

their perceptions of the documentary (Appendix B). The demographic survey was given to all participants and determined one's age, gender identity, subtitle preference (whether the participant uses subtitles or not), hours a week one watches TV, and exposure to other languages. The PPVT-5 (Dunn & Dunn, 2007) was administered to measure participants' receptive vocabulary skills. The PPVT-5 is a well-established, standardized assessment used to measure receptive vocabulary skills in individuals ranging in age from 2 years 6 months to 90+ years.

The pretests and post-tests were written, and participants were given a paper copy to complete. These were administered to measure vocabulary knowledge before and after watching the documentary. The pretest and post-tests consisted of two subtests: a fill-in-the-blank (FIB) subtest and a written definition (DEF) subtest. The FIB subtest consisted of 10 items. Each item was worth 1 point, and participants could earn a maximum score of 10 on the FIB portion. The DEF subtest consisted of 10 vocabulary words, and participants could earn a maximum score of 30 on the DEF subtest. Thus, participants could earn a maximum score of 40 points on the pretest and post-test. Scoring procedures are described below.

Procedures

Data Collection

Participants came to the campus speech-language-hearing clinic in the Alumni Memorial Gym (AMG) for the data collection procedures. First, participants completed an informed consent form stating their participation rights. Then, they were given a brief explanation about the pretest after walking into the research room. After that, they

completed the pretest, consisting of the FIB and DEF subtests. The pretest took approximately 10 minutes to complete. Participants then watched the 25-minute pre-selected clip of the documentary "Your Brain: Who's in Control?" (Alvarado et al., 2023) in a designated research room in the SLPA clinic. Each participant was randomly assigned to one of the three viewing conditions: audiovisual and text (AVT; n = 17), visual and text (VT; n = 14), and audiovisual (AV; n = 13). The designated groups determined the input by which each participant watched the documentary.

After the documentary was shown, every participant completed the post-test, and fifteen minutes passed. The fifteen-minute time between watching the documentary and the post-test was implemented to give equal waiting time to each participant (Irvin & Blankenship, 2022). Subjects completed the demographic and perceptions/biases surveys during this wait time. The PI or a research assistant also administered the PPVT-5 to the participants. If time remained after completing the surveys listed, they were asked to wait until the fifteen minutes had passed before taking the post-test.

Scoring

Scoring procedures for the FIB subtest and the DEF subtest were determined prior to scoring by the research team, which included the primary investigator and two research assistants. The FIB subtest was worth 10 points, as there were 10 FIB items, each worth 1 point. The DEF subtest was worth 30 points, as there were 10 items, and participants could earn up to 3 points for each stimulus item. The items on the DEF portion of the test were scored using a scale from 0-3. Zero points for one definition meant none of the linguistic concepts used to describe that vocabulary term in the documentary were mentioned in the participants' definition. A score of one meant one concept was

mentioned, and two meant two concepts were mentioned. A score of three meant three or more concepts were mentioned in the definition. See Appendix D for a summary of scoring criteria for the DEF subtest. Thus, participants could earn a total of 40 possible points on the vocabulary tests used for the pretest/post-test.

After the initial scoring, each research team member scored 10 to 15 out of the 44 participants. After the initial scoring, the researcher and at least one research assistant scored all the pretests and post-tests, and a 100% agreement was reached after discussing discrepant scores. The scorers were blind to the pretest and post-test statuses, eliminating bias and establishing internal reliability and consistency in scoring procedures.

Data Analysis

Data for the following variables was entered into the Statistical Packages for Social Sciences (SPSS) for analysis for each condition: (a) FIB pre-test score, (b) FIB post-test scores, (c) FIB gain scores, (d) DEF pre-test scores, (e) DEF post-test scores., (f) DEF gain scores, (g) total gain scores, (h) PPVT scores, (i) Likert scale scores from the perceptions survey.

Research question 1 evaluated which viewing condition (i.e., AVT, VT, AV) was most effective for vocabulary learning. The inclusion criteria stated that participants needed to score <25% on the pre-test for their data to be used in this study. 31 out of 44 participants scored above 25% on the TOTAL pre-test. 37 participants scored above 25% on the pre-test's FIB (fill-in-the-blank portion). Therefore, the data analysis did not include FIB gain scores and total scores. However, the DEF (definition portion) gain scores were used for data analysis. A repeated measures ANOVA was used to determine

if there were significant differences in vocabulary learning across the three viewing conditions (Zheng et al., 2022). An Independent Samples T-test was conducted to compare gain scores between two conditions rather than all three.

Additionally, effect sizes were calculated to compare means between two groups using the online Effect Size Calculator (Cohen's D for T-Test) (https://www.socscistatistics.com/effectsize/default3.aspx). Hedge's g was used as an effect size measure in this study. Hedges' g, like Cohen's d, is a standardized measure of the difference between the means of two groups in terms of the sample standard deviations (Ellis, 2010). Hedges' g was used for this study because it was designed to correct Cohen's d to reduce statistical bias when sample sizes are small (Lin & Aloe, 2021) and when sample sizes vary (Taylor & Alanazi, 2023). When using Hedges' g, an effect size of 0.15 implies a small effect, an effect size of 0.40 suggests a moderate effect, and an effect size of 0.75 implies a large effect (Brydges, 2019).

The second research question aimed to identify if a relationship was present between the PPVT-5 scores and vocabulary learning. Correlation analyses were run using the Pearson Coefficient Correlation Calculator (https://www.socscistatistics.com/tests/pearson/default2.aspx). PPVT and gain scores for each condition were entered to determine if a significant relationship was present.

Research question 3 sought to describe participant perceptions of the documentary based on the viewing condition they were assigned to. This question was analyzed using the perceptions survey and averaging the answers from each question measured using a Likert scale (Sydorenko et al., 2010). The results were then organized by viewing

condition or group and analyzed. The hypothesis is that there will be differing perceptions between the input modalities.

CHAPTER III

The Potency of Visual and Orthographic Modalities

Results

A total of 44 participants were initially recruited for this study. One participant was removed because they did not answer the pretest DEF questions. Additionally, three participants scored above the average range on the PPVT-5, creating outlier scores, and therefore were eliminated from the data analysis, resulting in 40 participants. The mean PPVT = 5 score for condition 1 was 96.46 (*SD* = 9.32). The mean PPVT-5 score for condition 2 (VT) was 96.60 (*SD* = 10.75). The mean PPVT-5 score for condition 3 was 102.00 (*SD* = 9.30). PPVT-5 scores ranged from 78 to 115 for all viewing conditions. Condition one held the minimum score, and condition three held the maximum score. A summary of this information can be found in Table 4.

Participants were also removed if they scored <25% on the pre-test. 38 participants scored above 25% on the FIB subtest, and 19 participants scored <25% on the total vocabulary pretest (FIB + DEF). As a result, the FIB subtest and total scores were not included in the data analysis. However, the DEF subtest scores were used to measure vocabulary learning. Eight participants scored >25% on the DEF pretest and were not included in the data analysis. After eliminating these participants, there were 32

participants whose data was analyzed. Data was analyzed for 13 participants in condition 1 (AVT), 10 participants in condition 2 (VT), and 9 participants in condition 3 (AV).

Research question one aimed to evaluate if one viewing condition (i.e., AVT, VT, AV) was most effective in teaching vocabulary to college students when viewing a documentary. This was evaluated using the gain scores from the DEF subtest. Condition one (AVT) obtained a mean DEF gain score of 8.3077 (*SD* = 1.70219; Range 5-10). The DEF pretest mean for condition one was 3.8462 (*SD* = 1.90815), and the DEF post-test mean was 12.1538 (*SD* = 2.5115). Condition two (VT) obtained a mean gain score of 9.8000 (*SD* = 3.11983; Range 5-14) on the DEF subtest. DEF pretest mean for condition two was 2.6000 (*SD* = 1.77639), and DEF post-test mean was 12.4000 (*SD* = 3.6878). Condition three (AV obtained a mean gain score of 9.5556 (*SD* = 2.18581; Range 6-12) on the DEF subtest. DEF pretest mean for condition three was 5.0000 (*SD* = 1.41421), and DEF post-test mean was 14.5556 (*SD* = 2.8333). See Tables 1 and 2 for a summary of this data.

A repeated measure ANOVA was used to determine if differences between mean gain scores across viewing conditions were statistically significant. The ANOVA revealed no statistically significant differences between the groups (p-value = 0.276). To compare the means of just two groups to each other, an independent samples t-test was run. The p-values when comparing conditions one and two and conditions one and three were approaching significance, as they were 0.078 and 0.074, respectively. For conditions two and three, the p-value was 0.424.

Additionally, effect sizes were calculated using Hedge's g to determine differences between the means among groups while considering standard deviations.

According to Brydges (2019), when using Hedge's g, a value of 0.15 indicates a small effect, 0.40 indicates a moderate effect, and 0.75 indicates a large effect. When comparing conditions one and three, the effect size was calculated to be 0.6532, indicating a moderate effect. When comparing conditions two and three, the effect size was calculated to be 0.0898, indicating little to no effect. See Table 1 for a summary of p-values and effect sizes.

Table 1. Significance and effect size compared between two conditions

Viewing conditions compared	p-value (significance)	Effect Size (Hedge's g)
1 vs. 2	0.078	0.6182
1 vs. 3	0.074	0.6532
2 vs. 3	0.424	0.0898

Table 2. Definition gain score means, ranges, and standard deviations

Viewing condition	Number of Participants Assigned	Definition Gain Score Mean	Definition Gain Score Range	Definition Gain Score Standard Deviation
1 (AVT)	13	8.3077	5.00-10.00	1.70219
2 (VT)	10	9.8000	5.00-14.00	3.11983
3 (AV)	9	9.5556	6.00-12.00	2.18581

Table 3. Definition pretest and post-test means and standard deviations

Viewing condition	Number of Participants Assigned	Definition Pretest Mean	Definition Pretest Standard Deviation	Definition Posttest Mean	Definition Posttest Standard Deviation
1 (AVT)	13	3.8462	1.90815	12.1538	2.5115
2 (VT)	10	2.6000	1.77639	12.4000	3.6878
3 (AV)	9	5.0000	1.41421	14.5556	2.8333

Table 4. PPVT means and standard deviations

Viewing condition	Number of Participants Assigned	PPVT Mean	PPVT Standard Deviation
1 (AVT)	13	96.4615	9.31500
2 (VT)	10	96.6000	10.75174
3 (AV)	9	102.0000	9.30054

Research question two aimed to determine how prior knowledge of vocabulary affects vocabulary gains. This was done by gathering PPVT scores, which showed prior vocabulary knowledge, and comparing them with definition gain scores. The significance between PPVT score means and definition gain scores were weak, with an R-value of 0.2213. However, a positive trend was observed. Figure 1 shows a general positive correlation between PPVT scores and DEF gain scores (the y-axis represents PPVT scores, and the x-axis represents post-test gain scores). This indicates that as PPVT scores increase, DEF gain scores also increase.

The relationship between total gain scores (DEF + FIB) and PPVT scores was also measured. The correlation, shown in Figure 2, is weak (R-value = 0.0895), and though the data points are scattered, they display a generally positive trend like Figure 1. A positive trend suggests that total gain scores increased as PPVT scores increased.

Finally, the correlation between PPVT scores (y-axis) and pretest scores (x-axis) was analyzed and showed an R-value of 0.2331. This means the correlation is positive, but its significance is weak. See Figure 3 for the graph showing this significance. Again, this positive trend indicates that as PPVT scores increased, so did pretest scores.

Figure 1. Correlation between PPVT scores and definition gain scores

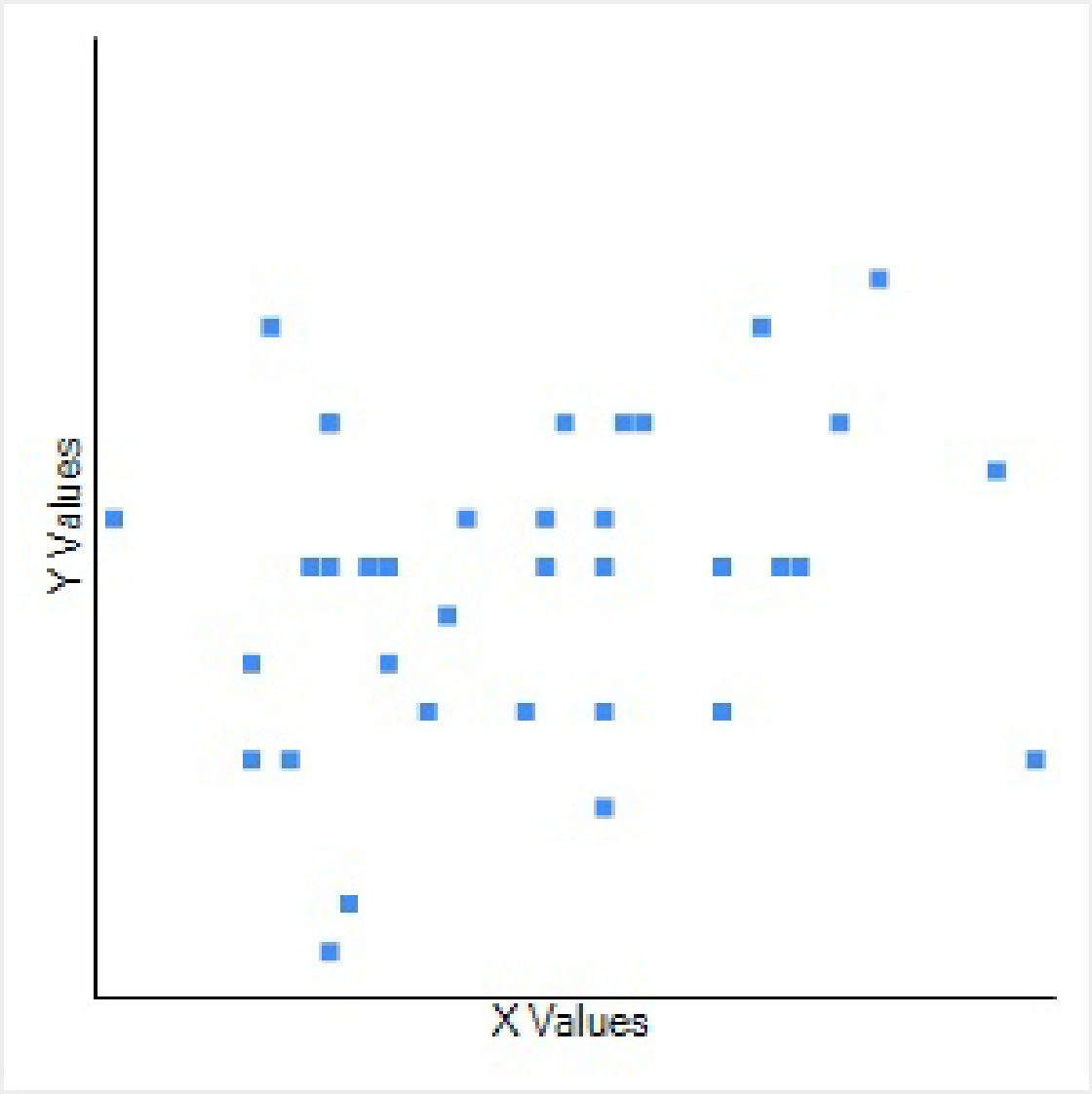


Figure 2. Correlation between PPVT scores and total gain scores

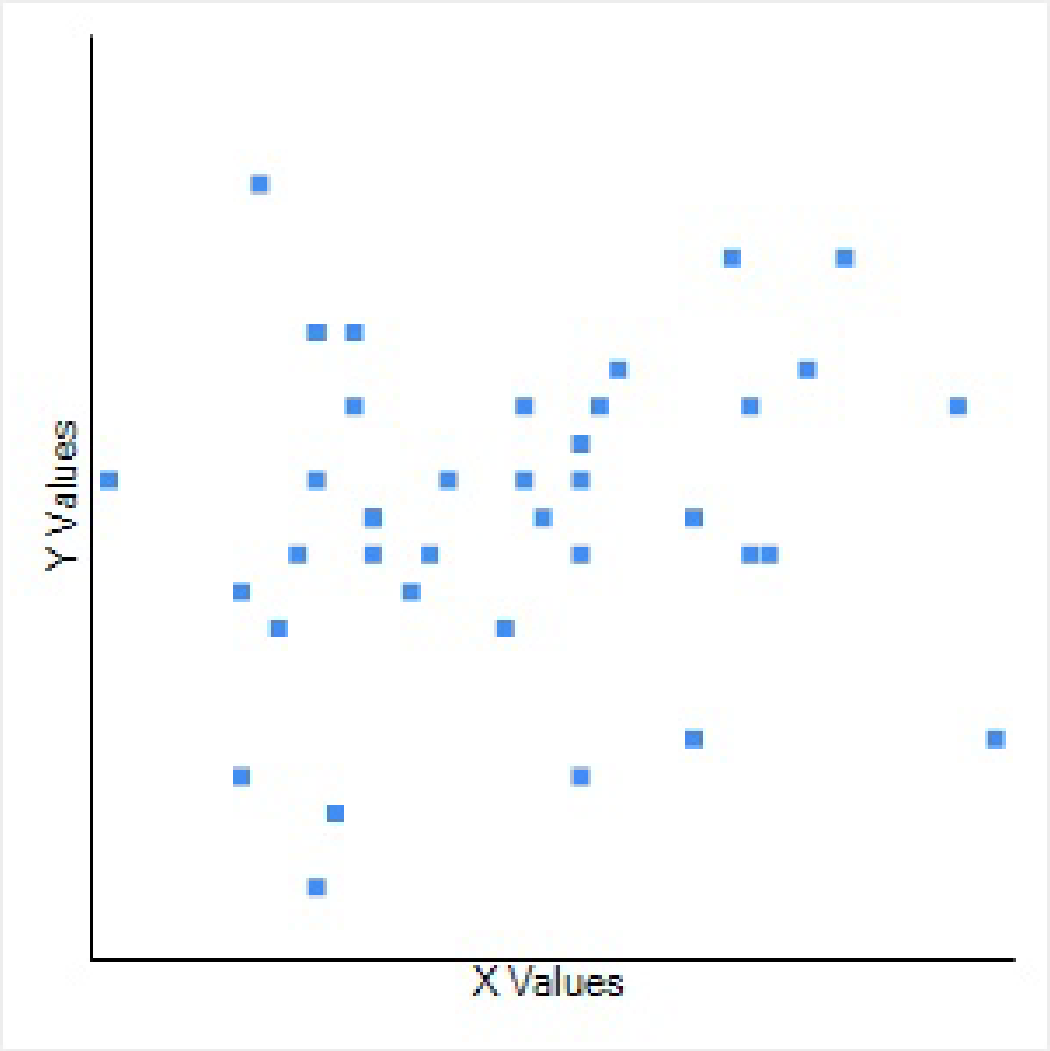
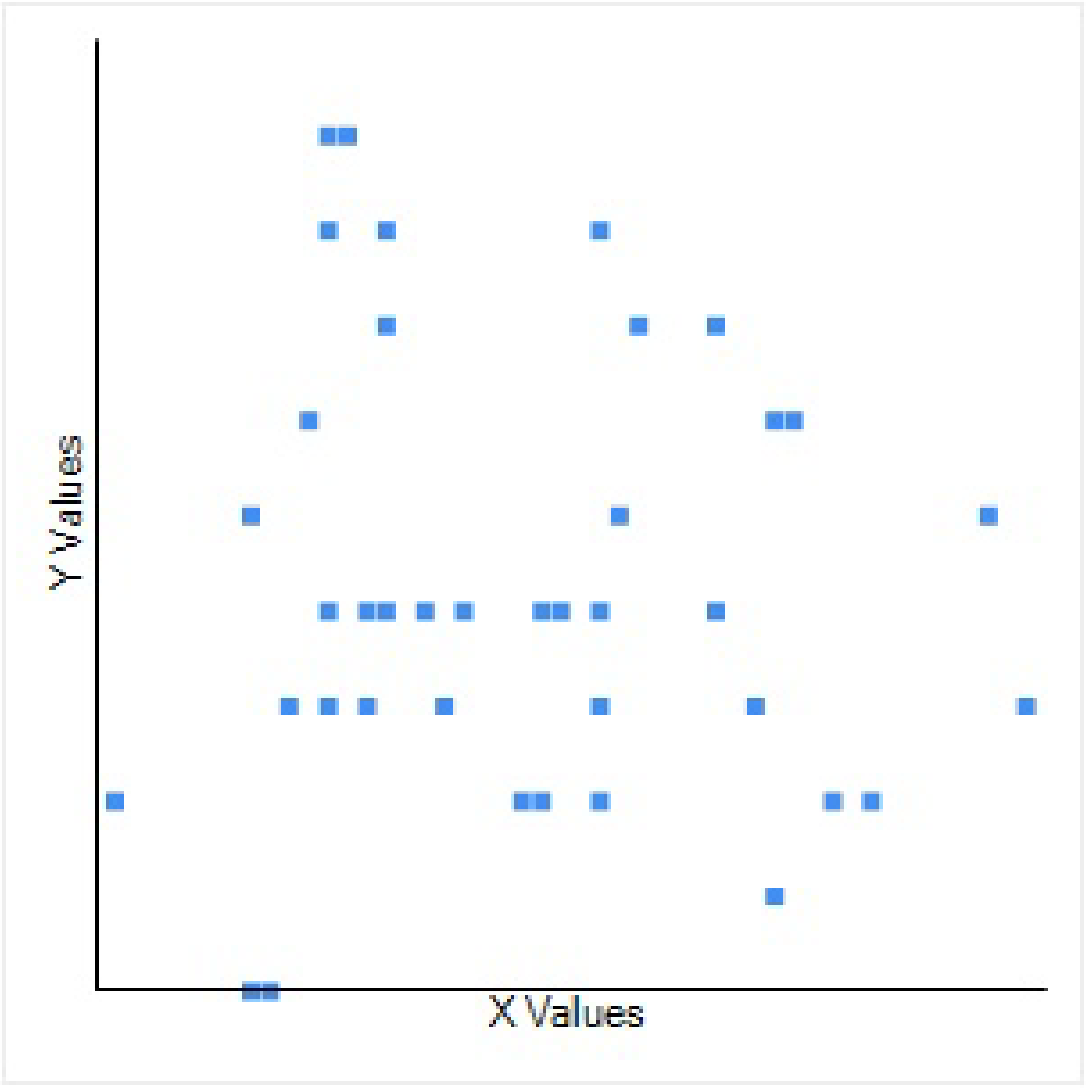
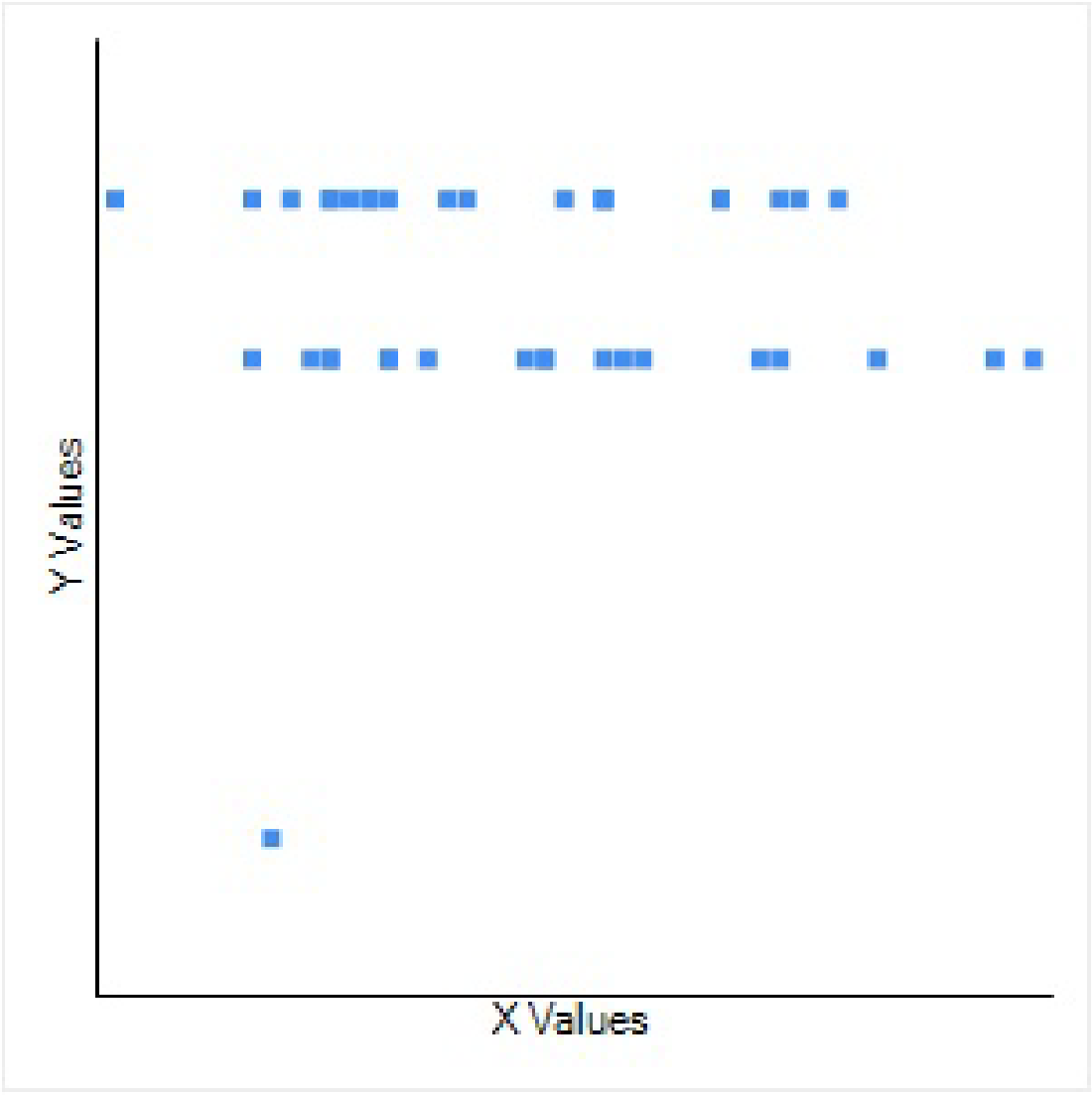


Figure 3. Correlation between PPVT scores and pretest scores



The third research question determined perceptions of the documentary and how they changed between the three conditions and different PPVT scores (prior vocabulary knowledge). Figure 4 shows the relationship between PPVT scores and how easily participants could understand the documentary. The PI chose to investigate this under the assumption that increased vocabulary prior to watching the documentary would result in the viewer having an easier time understanding the language in the documentary. The significance of correlation is weak (R-Value = 0.1313). The graph shows a similar understanding (constant relationship) of the documentary among different PPVT scores.

Figure 4. Correlation between PPVT scores and understanding of the documentary



The perceptions of the documentary across viewing conditions were measured using Likert scales to rate different statements. Likert scale values represented the following: 5 = strongly agree, 4 = agree, 3 = neutral, 2 = disagree, 1 = strongly disagree. The mean Likert scale scores for six statements in each viewing condition were taken from the perceptions survey. These mean scores are presented in Table 5.

Table 5. Mean Likert scale scores of perceptions survey statements

Survey Statements:	Condition 1 (AVT):	Condition 2 (VT):	Condition 3 (AV):
1. Enjoyed documentary	4.35	4.57	4.62
2. Documentary was easy to understand	4.47	4.29	4.54
3. Easy to maintain attention	3.88	3.93	4
4. Would watch again in the same condition assigned	3.71	2.71	3.62
5. Would recommend documentary to a friend	3.82	4	3.92
6. Learned something new	4.65	4.57	4.69

Discussion

Vocabulary Learning Between Viewing Conditions

Effect size calculations revealed little difference between conditions two (VT) and three (AV). This suggests that when the participants were exposed to novel language terms with only one input modality—auditory or visual—there was no difference in their vocabulary learning. However, the effect size for condition one (i.e., AVT) was substantially different from conditions two and three. Interestingly, this group had the lowest mean gain score and the lowest range. Therefore, these results show that the multimodal language input—visual and auditory—resulted in less vocabulary learning than groups who received only one input modality. This could be due to input overshadowing, when one mode of input interferes with or hinders the processing of another mode of input (Dodson et al., 1997). Adults tend to use visuals to overshadow aural input, which supports the results demonstrating the impact of visuals (images or subtitles) on learning vocabulary.

Participants who watched the documentary in the VT condition made the strongest gains in vocabulary learning. The VT group also had the highest PPVT scores, though the differences between conditions were insignificant. Again, overshadowing may have played a role in these results, since adults, in this case, college students, overshadow aural input with visuals. Previous research has also shown the positive effects of orthographic input, which could have assisted the VT condition in vocabulary learning. Participants watching a documentary in the VT condition may have benefited from completing a written test because the condition and the tests included similar input modalities (written text). Had the participants completed a test where they had to produce

the information verbally, the results may have changed. Additionally, there may have been more participants with a background in neuroscience in one condition than in another, skewing the results of vocabulary learning.

The p-value of the difference in gain scores between all viewing groups was 0.276, indicating no statistically significant differences between vocabulary learning and the viewing conditions. This weak correlation could be because the repeated measures ANOVA measured differences between all conditions, which, collectively, were not varied. Additionally, a smaller sample size could have resulted in this lack of significance. However, when looking at effect sizes between two conditions instead of three at once, there was a considerable difference between condition one and conditions two and three.

Total gain scores were the points gained from the pretest to the post-test. For total gain scores, group two (VT) produced the highest average (M = 14.10), and group one (AVT) produced the lowest average (M = 1.92). This means group two gained the most, and group one gained the least. These results do not support the hypothesis that the group with all three input modalities (audio, visual, and orthographic) would yield the highest vocabulary gains, and the visual and text group would produce the lowest gains. It could be that participants in group two were so focused on using the subtitles to understand the vocabulary in the documentary that they showed better outcomes in the post-test than the participants who had the audio. Again, since the pretest and post-tests were written tests, group two may have had an advantage in reading the subtitles rather than listening to the audio. Group two also did not have audio turned on, forcing them to focus more on subtitles. Group two had an average DEF gain score of 9.80, the highest of the three

conditions. Group one had an average DEF gain score of 8.31, the lowest of the three, and group three had an average DEF gain score of 9.56. These results line up with the total gain scores, meaning the group with visual and text (no audio present) had the highest definition and total gain scores, and the group with audio, visual, and orthographic modalities had the lowest definitions and total gain scores.

Relationship Between Language Skills and Language Learning

Though there is a general positive correlation (as PPVT scores, or vocabulary knowledge, increase, gain scores increase), the R-value of 0.2213 indicates a weak correlation. The weakness of correlation is also shown in the graph in the results section, displaying scattered data points. The weak correlation could be because the gain scores were too similar across all or two of the three conditions. PPVT scores determine prior vocabulary rather than how well participants learn new information, which may be another reason a strong correlation was absent.

Perceptions of the Documentary Between Viewing Conditions

For the statement, "I found it easy to maintain my attention to the documentary while it was playing," condition one had the lowest average Likert scale score (3.88), while group three had the highest (4). One thing to note is that group three had the highest Likert scale scores for the first three statements (see Table 4). This could mean that the audiovisual modality (without text) can slightly increase enjoyment, understanding, and attention, although the differences are too minimal to say anything yet. Attention was maintained similarly across all conditions.

The results of this statement ("I would watch the documentary in the same condition again") yielded the most substantial differences between conditions out of all

statements. Group two had the lowest average Likert scale score (2.71), and group one had the highest (3.71). Group three was not far from group one, with an average of 3.62. So, the participants who watched the documentary with no audio but with the visual and text modalities would be least likely to watch the documentary in the same condition again. Participants with all three modalities (AVT) would most likely watch the documentary in the same condition again. This could be because audio was not present, therefore making it harder to want to watch the documentary. In Zheng et al. (2022), eye movement patterns were considered among viewing conditions, and "a more centralized (looking mainly at the screen center) eye movement pattern predicted better comprehension as opposed to a distributed pattern (with distributed regions of interest)." A distributed pattern may have made it more difficult to keep watching. It could have also had the opposite effect: making it hard to look away from the subtitles and the video. For the statement, "I have learned something new from this documentary," condition two had the lowest average Likert scale score (4.57), and group three had the highest (4.69). Once more, the difference is too weak to say that any condition affected participant perception. However, this statement had higher scores than all the other statements, which means the likelihood that participants learned something new is higher than in the other statements, according to participants.

PPVT Scores and Participants' Understanding

Graph 4 shows two lines of constant data points in PPVT scores and different perceptions of participants' understanding of Likert scale scores. The R-value is 0.1313, which shows no relationship between PPVT scores and understanding of the documentary, meaning prior knowledge did not affect how well participants understood

the documentary. However, since only one statement about participants' understanding was analyzed, there is insufficient evidence to suggest that PPVT scores do not affect participants' learning. Additionally, using a documentary with denser language should be considered. It may give more accurate results on prior knowledge and how easily participants understood the documentary because several participants in this study had previously taken neurology and anatomy courses, which meant that gain scores were not as significant for these individuals.

Implications

The importance of visual modalities has been highlighted, which supports prior evidence on the impact of subtitles and the concept of overshadowing. Adults, in this case college students, tend to pick up visual modalities before aural ones when viewing media. This provides information for streaming services like Netflix and Hulu that subtitles are a big part of learning while watching media for young adult audiences. Additionally, subtitles and visuals are everywhere in the age of short-form media. Many watch TikTok videos, YouTube Shorts, and Instagram Reels on mute. Subtitles, visual graphics, or any form of visual media have already started to and will change the way adults watch shows, movies, and documentaries in the next few years.

There are also clinical implications that can be considered. As previously mentioned, these results show the significance of visual and orthographic learning. Clinically speaking, teachers and speech-language pathologists can use these results and prior evidence of orthographic vocabulary learning to increase learning in school-age children. Though the study did not involve adolescent participants, subtitles and visual learning

can still be effective in teaching. Educational professionals can include subtitles when showing children videos, shows, and movies to increase language awareness. This increased awareness occurs because they have more than one or two language inputs.

Conclusions

Although evidence suggests that having multiple modalities of viewing input increases vocabulary learning, the results from this study show the opposite effect. The most significant finding was that effect sizes showed a barrier between condition one (AVT) and conditions two and three (VT and AV). The VT condition showed the highest vocabulary gain scores, which supports the idea that visual learning is more impactful than aural learning in most college students. Prior knowledge does increase vocabulary learning (Gensemer et al., 1976), but the data from this study are too weak and scattered to show significance. The results support the hypothesis and past research that vocabulary learning can directly correlate to prior knowledge. Though the viewing conditions display no change regarding most participants' perceptions of the documentary, there were a few notable differences between having more versus fewer input modalities. Individuals with all three inputs were more likely to watch the documentary again in the same condition than those who had two input modalities, and significantly more than participants without audio in the documentary.

Future Directions

There are many routes that this study can take in the future. Firstly, vocabulary learning in an unfamiliar language can be studied, but modifying the limitations of this study can also be done to examine a familiar language. For one, the process of test creation should be altered. Creating and scoring tests with easily understood vocabulary using context clues was not challenging to the participants, which may have affected the outcome of gain scores. Additionally, using a documentary with a denser vocabulary can

be beneficial to gain scores and scoring criteria precision. Almost all participants (except two) were females, so future research can examine how gender impacts vocabulary and include a diverse group of people. Formulating solid scoring criteria before the scoring process begins is ideal and can be applied in future studies. Finally, many studies about language learning use subtitles, which are used in unfamiliar languages, and participants learn entirely new words. This study used participants' L1 or their first language, but for future research, it can be expanded to compare vocabulary learning between a familiar language and an unfamiliar language.

References

Alotaibi, H. M., Mahdi, H. S., & Alwathnani, D. (2023). An effectiveness of subtitles in L2 classrooms: A meta-analysis study. *Education Sciences, 13*(3)*,* 274. https://doi.org/10.3390/educsci13030274

Alvarado, D., Bicks, M., Strachan, & A. L., Sussberg, J. (2023). Your brain: Who's in control? In NOVA, Season 50. PBS. https://www.pbs.org/video/your-brain-whos-in-control-q6suyy/

Balci, L. V., Rich, P. J., & Roberts, B. (2020). The effects of subtitles and captions on an interactive information literacy tutorial for English majors at a Turkish university. *Journal of Academic Librarianship, 46*(1), 1–7. https://doi.org/10.1016/j.acalib.2019.102078

Black, S. (2021). The potential benefits of subtitles for enhancing language acquisition and literacy in children. *Translation, Cognition & Behavior, 4*(1), 74–97. https://doi.org/10.1075/tcb.00051.bla

Brydges, C. R. (2019). Effect size guidelines, sample size calculations, and statistical power in gerontology. *Innovation in Aging, 3*(4), 1–8. https://doi.org/10.1093/geroni/igz036.

Caruana, S. (2021). An overview of audiovisual input as a means for foreign language acquisition in different contexts. *Language and Speech, 64*(4), 1018–1036. https://doi.org/10.1177/0023830920985897

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition, 25*(2), 129–139. https://doi.org/10.3758/BF03201107

Dunn, L. M. & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test – Fourth Edition*. Pearson.

Ellis, P. D. (2010). The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results. Cambridge University Press. https://doi.org/10.1017/CBO9780511761676

Gensemer, I. B., Walker, J. C., & Cadman, T. E. (1976). Using the Peabody picture vocabulary test with children having difficulty learning. *Journal of Learning Disabilities, 9*(3), 179- 181. https://doi.org/10.1177/002221947600900310

Harji, M. B., Woods, P. C., & Alavi, Z. K. (2010). The effect of viewing subtitled videos on vocabulary learning. *Journal of College Teaching & Learning, 7*(9), 37-42. https://doi.org/10.19030/tlc.v7i9.146

Irvin, S. & Blankenship, K. G. (2022). Vocabulary & academic success in university undergraduate students. *Teaching and Learning in Communication Sciences & Disorders, 6*(2), 1–18.

Kanellopoulou, C., Kermanidis, K. L., & Giannakoulopoulos, A. (2019). The dual-coding and multimedia learning theories: Film subtitles as a vocabulary teaching tool. *Education Sciences, 9*(3)*,* 210. https://doi.org/10.3390/educsci9030210

Lertola, J. (2019). Second language vocabulary learning through subtitling. *Spanish Journal of Applied Linguistics, 32*(2), 486–514. https://doi.org/10.1075/resla.17009.ler

Lin, L., & Aloe, A. M. (2021). Evaluation of various estimators for standardized mean difference in meta-analysis. *Statistics in Medicine, 40*(2), 403–426. https://doi.org/10.1002/sim.8781

Malitesta, D., Cornacchia, G., Pomo, C., Merra, F. A., Noia, T. D., & Sciascio, E. D. (2023). Formalizing multimedia recommendation through multimodal deep learning. *Conference Acronym, 3*(3), 1-28. https://doi.org/10.48550/arXiv.2309.05273

Marfo, P., & Okyere, G. A. (2019). The accuracy of effect-size estimates under normals and contaminated normals in meta-analysis. *Heliyon, 5*(6). https://doi.org/10.1016/j.heliyon.2019.e01838

Matielo, R., D'Ely, R., & Beretta, L. (2015). The effects of interlingual and intralingual subtitles on second language learning/acquisition: A state-of-the-art review. *Trabalhos em Linguística Aplicada, 54*(1), 161-182. https://doi.org/10.1590/0103-18134456147091

Mohsen, M. A. (2016). The use of help options in multimedia listening environments to aid language learning: A review. *British Journal of Educational Technology, 47*(6), 1232-1242. https://doi.org/10.1111/bjet.12305

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

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

Robinson, C. W., & Sloutsky, V. M. (2007). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science, 10*(6), 734–740. https://doi.org/10.1111/j.1467-7686.2007.00627.x

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

Taylor, J. M. & Alanazi, S. (2023). Cohen's and Hedge's g. *Journal of Nursing Education, 62*(5), 316-317. https://doi.org/10.3928/01484834-20230415-02.

Teng, M. F. (2022). Incidental L2 vocabulary learning from viewing captioned videos: Effects from learner-related factors. *System, 105*, 1–12. https://doi.org/10.1016/j.system.102736

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*(3), 191-210. https://doi.org/10.1017/S0261444807004338

Wei, R., & Fan, L. (2022). On-screen texts in audiovisual input for L2 vocabulary learning: A review. *Psychology of Language, 13,* 1-11. https://doi.org/10.3389/fpsyg.2022.904523

Zhang, R., & Zou, D. (2022). A state-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning. *Computer Assisted Language Learning, 35*(9), 2790–2816. https://doi.org/10.1080/09588221.2021.1896555

Zheng, Y., Ye, X., & Hsiao, J. H. (2022). Does adding video and subtitles to an audio lesson facilitate its comprehension? *Learning and Instruction, 77*, 1–13. https://doi.org/10.1016/j.learninstruc.2021.101542

Appendices

Appendix A

Demographic Survey Questions

- What is your age?
    - Participants will type in their age in the text box.
- What is your date of birth? (Year/Month/Day)
    - Participants will type in their date of birth in the text box.
- What is your gender identity?
    - Multiple choice options given: Female, Male, Non-binary
- What is your subtitle preference?
    - Multiple choice options given: ‘I prefer to read subtitles while watching TV.’; ‘I prefer not to turn off subtitles while watching TV.’; ‘I don’t have a preference.’
- How many hours a week do you watch TV (cable, streaming, etc.)?
    - Multiple choice options given: Less than 8 hours, 8-12 hours, 12-16 hours, 16-20 hours, 20-24 hours, 24+ hours
- Do you have a speech, language, reading, or hearing impairment?
    - Participants will type in their answers in the text box.
- How often are you exposed to languages other than English?
    - Multiple choice options given: Never, Rarely, Sometimes, Frequently, Always
- What is your primary language?
    - Participants will type in their answers in the text box.

- Have you watched the following documentary: “Your Brain: Who’s In Control?”
    - Participants will type in their answers in the text box.

Appendix B

Pretest and Post-test Questions

Define:

1. Pre-frontal cortex

2. E.E.G. (Electroencephalogram)-

3. Thalamus-

4. Corpus callosum-

5. Insula-

6. Split-brain phenomenon-

7. Amplitude-

8. Slow brain wave-

9. Trust Game-

10. Oscillation-

Fill in the blanks using the word bank provided (a word bank will be provided at the top of the page):

1. One part of the brain that doesn't wake up during sleepwalking is called the _____.
2. They used a(n) _____, a device that rests on the scalp and detects electrical activity in the brain.

The _____ is a region of the brain that acts as a communication hub.

3. For some people with epilepsy, a seizure in one hemisphere can quickly spread to the other by way of the _____.
4. Signals about having a gut feeling that maybe this isn't a good idea originate from the _____.
5. The _____ suggests that there can be "two minds" inside of a skull.
6. The _____ measures how small or big brainwaves are.
7. One person has some sum of money, and they can choose to invest any amount of that money in their partner in the _____.
8. The brain generates brainwaves, or _____.
9. The _____ measures how fast brainwaves come and go.

Appendix C

Perceptions Survey Statements

- Rate the documentary on a scale of 1-5 (out of five stars)
    - Choose one if you did not like the documentary. Choose five if you liked the documentary.
- I found the content in this documentary easy to understand.
    - Choose ‘strongly disagree’ if you found the documentary difficult to understand. Choose ‘strongly agree’ if you found the documentary easy to understand.
- I enjoyed watching the documentary.
    - Choose ‘strongly disagree if you did not enjoy watching the documentary. Choose ‘strongly agree’ if you enjoyed watching the documentary.
- I found it easy to maintain my attention to the documentary while it was playing.
    - Choose ‘strongly disagree if you believe it was not easy to maintain attention while watching the documentary. Choose ‘strongly agree’ if you feel it was easy to maintain attention while watching the documentary.
- I would choose to watch this documentary in the same condition if I were to watch it again.
    - Choose ‘strongly disagree if you would not watch this documentary in the same viewing condition if you were to see it again. Choose ‘strongly agree’ if you would watch this documentary in the same viewing condition if you were to see it again.
- I would recommend this documentary to a friend.

    - Choose ‘strongly disagree’ if you would not recommend this documentary to a friend. Choose ‘strongly agree’ if you would recommend this documentary to a friend.
- I have learned something new from this documentary.
    - Choose ‘strongly disagree’ if you believe you did not learn something new from this documentary. Choose ‘strongly agree’ if you believe you learned something new from this documentary.

Appendix D

Pretest and Post-test Scoring Criteria

SCALE

3-Participant included 3+ concepts from documentary

2-Participant included 2 concepts

1-Participant included 1 concept

0-Participant included 0 concepts

Written Definitions Answer Key	
Term	**Acceptable Responses/Language**
Prefrontal Cortex	decisions/choices
	Self-awareness
	part of the brain that doesn't wake up during sleepwalking
	part of the brain that controls balance and speech becomes active without this
EEG	set of electrodes
	rests on scalp
	detects electrical activity
	comes in the form of waves
thalamus	central way station
	information processing
	two parts
	communication hub
	processes auditory, visual, and pain or information
corpus callosum	transfer pathway
	communication passes through this
	bundle of fibers
	connects left and right hemispheres
	if severed, seizure can no longer cross to the other side of the brain
	severing can treat epilepsy
insula	brain's thermometer

	signals gut feeling and guilt
	allows us to make decisions to avoid harming someone else
Split-brain phenomenon	two minds
	single brain
	cross-cueing and management of two mental systems into one unified act
Amplitude	small
	big
	waves
Slow brain wave	deep sleep stage
	brain waves
	slow down
Trust Game	cooperative game
	one person has a sum of money and can choose to invest any amount of that money in their partner
	investment grows, then they decide whether to be greedy or generous
	many return due to guilt
Oscillation	brain waves measured by
	frequency
	amplitude

Appendix E

IRB Approval Letter

Office of Research Compliance
2269 Middle Tennessee Blvd.
Sam H. Ingram Bldg (ING) Room 010A
Box 124
Murfreesboro, TN 37132
www.mtsu.edu/irb

Date: June 11, 2024
PI: Ananya Arcot
Department: Middle Tennessee State University, Health and Human Performance
Re: Initial - IRB-FY2024-206
The Effects of Audiovisual, Visual, and Orthographic Input on Language Learning in College Students

The Middle Tennessee State University Institutional Review Board has reviewed and approved by Expedited Review the above referenced research study. The approval is effective starting June 11, 2024.

Decision: Approved
Category: 7. Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies. (NOTE: Some research in this category may be exempt from the HHS regulations for the protection of human subjects. 45 CFR 46.101(b)(2) and (b)(3). This listing refers only to research that is not exempt.)

Findings:
Research Notes:

The following apply to your approved study:

1. In accordance with 45 CFR 46.110 and the regulations for Expedited Review (Common Rule), this project does not expire and continuing review is not required by the IRB.
2. Any unanticipated harm to participants or adverse events must be reported to the Office of Compliance.
3. All modifications to the approved study must be submitted for review through Cayuse IRB for approval before their implementation. Adding new researchers constitutes a modification to the protocol. Per MTSU Policy, a researcher is defined as anyone who handles the data or interacts with participants. Everyone meeting this definition for this project must have completed the required CITI training and received IRB approval prior to becoming actively involved in the project.
4. Closure of the study must be submitted within Cayuse when the study ends or when personal identifiers are removed from the data and all codes and keys are destroyed.
5. All research materials must be retained by the PI for at least three (3) years after study completion and then destroyed in a manner that maintains confidentiality and anonymity.
6. All approval letters and study documents are located within Submission Details in Cayuse IRB.

Sincerely,

The Middle Tennessee State University Institutional Review Board