Investigating the Interaction of Visual Input, Working Memory, and Listening Comprehension

Kerry Pusey

Northern Arizona University, Flagstaff, Arizona, United States of America

Karen Lenz

Northern Arizona University, Flagstaff, Arizona, United States of America

Abstract

This study investigated the effect of visual input on L2 listening comprehension within the context of a North American intensive English program. The interaction between visual input and working memory (WM) was also investigated, with the aim of clarifying what role visual input, together with WM, plays in L2 listening tests. The study compared two groups of upper-intermediate L1 Chinese and Arabic ESL students. All participants ($N = 24$) took a WM test and were divided into two groups to take a listening comprehension test under two treatment conditions: one with video and one with audio-only texts. Results indicated that the presence of visual input had a significant negative effect on listening comprehension, while working memory had no significant effect. Additionally, no interaction was found between WM and the presence or absence of visual input. This paper concludes by discussing further research questions and implications for L2 listening assessment.

Listening in a second language (L2) has been described as an arduous task: comprehension of speech requires the simultaneous processing of phonological, syntactic, semantic, and pragmatic information (Flowerdew & Miller, 2005). The act of listening, moreover, does not typically occur in isolation. Listeners usually receive visual input, such as observations of kinesic behavior and contextual information (Gregersen, 2007; Kellerman, 1992). In light of this fact, teachers began using video in L2 listening classrooms in the mid-1970s due to its ability to contextualize language and increase motivation (Flowerdew & Miller, 2005). However, while the use of video has now become standard practice in many L2 classrooms, it is not always used in testing situations. This discrepancy begs the question of what effect, if any, the use of video has on listening comprehension test scores.

Another dimension of L2 listening is working memory (WM). Unlike aspects of language ability such as reading and writing, the aural channel through which listening is accomplished is typically more ephemeral in nature; the input listeners receive disappears after a speaker has finished speaking. This is particularly true in many academic contexts, where listening is often a one-way, *transactional* process (Buck, 2001; Morley, 2001; Peterson, 2001) that requires a high level of fluency and possibly a high WM capacity, especially at the discourse level (Juffs &

Harrington, 2011). Thus, the question arises as to whether or not differences in WM capacity impact performance on tests of listening comprehension.

The present study sought to investigate the relationships among visual input, WM, and listening comprehension. Such insight may influence test design, including selection of item types, testing conditions, scoring procedures, and training of raters. It would also alert test developers as to whether certain students (e.g., those with low WM) are being unfairly disadvantaged (i.e., questions of bias and construct-irrelevant variance). Investigating these relationships furthermore raises important questions about the construct definition of listening in academic settings (Ockey, 2007; Wagner, 2008) and the degree to which target language use (TLU) tasks (Bachman & Palmer, 2010) should dictate the format of L2 listening tests. In addition to informing test developers, a better understanding of the variables that affect listening comprehension may help shape curricular objectives, focus listening instruction, and guide learners in strategy selection and use.

Background

In light of the complexity involved in L2 listening described above, much scholarly work has been directed at developing pedagogical techniques to help students improve their listening comprehension skills. The following is a brief list of some of the more widely endorsed practices for L2 listening pedagogy:

- Teachers should activate students' background knowledge and promote purposeful listening in order to help learners "fill in the gaps" when microlevel comprehension problems occur (e.g., misunderstanding a single word or suffix) (Flowerdew & Miller, 2005).
- The rate of speech and length of listening passages used should be closely monitored (Buck, 2001), especially when teaching learners at lower proficiency levels.
- Sociocultural and other specialized knowledge should be taught prior to engaging in listening tasks (Buck, 2001).
- Listening passages should be neither too long nor too short (three to five minutes may be a good rule of thumb) and should be played more than once in order to minimize memory effects (Buck, 2001; Vandergrift & Goh, 2012). This will vary, however, depending on the target listening objectives of the course.
- A variety of texts, tasks, and types of listening opportunities should be provided in order to develop a broad range of listening skills (Buck, 2001).

Some of these concepts also relate to the use of listening strategies. Language learning strategies are often divided into the following three types: cognitive, metacognitive, and socioaffective. In regard to L2 listening, cognitive strategies are generally conceived of as online "tactics" used during the process of listening, such as guessing from context and taking notes. Metacognitive strategies typically refer to advanced planning techniques, including making an outline and predicting what one will hear in a listening passage. Lastly, socioaffective strategies involve using available resources to help clarify gaps in comprehension (for example, asking for clarification and working with others).

Aside from strategies, many scholars have emphasized the benefits of including visual input (especially video) in language classes. Reasons for doing so include increasing student engagement (Brinton, 2014), exposing students to pragmatic aspects of the target language (Washburn, 2001), and aiding in listening comprehension (Flowerdew & Miller, 2005; Kellerman, 1992).

In addition to the support for using visual input in L2 pedagogy, much of the literature on L2 listening supports the use of visual input (specifically video texts) in language testing for reasons such as content and construct validity, as well as student perceptions and preferences. Feak and Salehzadeh (2001) argue for the use of video in listening placement tests in an English for Academic Purposes setting, noting that “nearly all of the listening encounters of our students—lectures, class discussions, seminars, lab experiments, and office hours — are accompanied by dynamic visual input” (p. 481). They analyzed students’ perceptions of the fairness and authenticity of a video listening assessment and found that few students found the test to be unfair or unnatural. Progosh (1996) incorporated video texts in course materials at a Japanese tertiary institution after conducting a needs analysis and learning that students were interested in studying English through film. After administering tests and surveying students, Progosh found that 92% of learners preferred the use of video texts over audio-only texts in tests of listening comprehension. Ockey (2007) studied student perceptions of the usefulness of video texts in comparison with still images. In this study, most participants found the video texts more helpful than still images, though this finding varied according to text and participant. Several participants also commented on the benefit of specific visual cues present in the video texts (e.g., lip movements, hand motions, facial and body gestures), while the still images were only found to aid in establishing the initial context of the listening passages.

In addition to issues of validity and student perceptions, research has also investigated the effect of video on students’ performance, but has yielded mixed results. In several studies, Wagner (2010, 2013) has examined the effect of visual input on test taker performance and has generally found a positive effect of video texts on listening comprehension test scores. On the other hand, Suvorov (2009) investigated the role of visual input and text type (dialogue and lecture) and found that while the test format (i.e., absence or presence of visual input) did not affect scores on a dialogue task, the use of video did have a negative effect on the lecture task, noting that context visuals may have been distracting, rather than facilitative, in this task. Other studies (e.g., Brett, 1997; Coniam, 2001; Londe, 2009) have yielded results that have shown no clear indication of an advantage or disadvantage to using video as opposed to audio-only texts.

These results point to the need for further research examining the relationship between visual input and listening comprehension, and also other factors (e.g., individual differences). One possible moderating variable is WM. It has been argued that WM span and L2 learning have a positive relationship (Atkins & Baddeley, 1998; Biedroń & Szczepaniak, 2012; Majerus, Poncelet, Van der Linden, & Weekes, 2008; Szmalec, Brysbaert, & Duyck, 2012). Therefore, if WM contributes to other aspects of language learning, it may affect learners’ listening comprehension as well (cf. Cross, 2011, p. 47), especially in the absence of visual input that normally accompanies face-to-face communication (e.g., while taking an audio-only listening test). However, the interaction of WM and visual input during listening comprehension tasks has not, to the researchers’ knowledge, been studied.

The Current Study

In light of these observations, the current study will address the following research questions:

- RQ 1: Does the presence of visual input affect listening comprehension among university-level L2 learners?
- RQ 2: Does working memory affect listening comprehension?
- RQ 3: Is there an interaction between working memory and the presence of visual input on listening comprehension?

Because visual input typically accompanies the TLU task of listening in academic settings (Bachman & Palmer, 2010), it seems reasonable that the use of video would have a positive effect on listening comprehension. It also seems reasonable that students with higher WM scores will score higher on tests of listening comprehension. Finally, it is hypothesized that the positive effect of visual input on listening comprehension will be more prominent for learners with lower WM capacity.

Methods

Participants

Participants consisted of 24 first language (L1) Chinese and Arabic students enrolled in an intensive English program at a North American university. The majority of students in each class were male, and their ages ranged from approximately 18-24. The Intensive English Program (a pseudonym used here to refer to the program, henceforth referred to as the IEP) is composed of six levels (Level 1 being the lowest), and students are placed into levels based on their performance on a placement test consisting of listening, speaking, reading, and writing sub-sections. Two sections of Level 4 students participated in this study. Because students were all placed in the same IEP level, it was assumed that participants had relatively equal levels of English language ability. These particular participants were selected, in part, because at the time of the study, one of the researchers was teaching one of the Level 4 sections, thus making administration more convenient. (Participants therefore constituted a convenience sample.) Also, compared to other sections of IEP students, both sections of Level 4 students had not been involved in many research projects over the semester. Thus, taking class time to collect data was deemed less burdensome for Level 4 students and their teachers.

Measures

The variables in this study include listening comprehension, the presence of visual input, and WM. Listening comprehension was operationalized as participants' ability to respond correctly to a series of multiple-choice main idea and detail questions on an academic listening comprehension test (titled the Sociology Listening Test [SLT]). The presence and absence of visual input were operationalized as the use of video texts or audio-only texts (respectively) on the SLT. Theoretically, WM refers to "the mental processes responsible for the temporary storage and manipulation of information in the course of on-going processing" (Juffs & Harrington, 2011, p. 138). For the purpose of this study, WM was operationalized as the extent to which one can correctly repeat a series of numbers that are presented in increasingly numerous sets, as represented by scores on a digit span task (DST) and subsequent placement into low, middle, or high WM groups (WMGs).

As noted above, this study employed the SLT under two treatment conditions: one using video and one using audio-only texts. The test itself was identical in the two treatment conditions; the only difference between treatments was the presence or absence of visual input (i.e., the video texts), with a slight difference in audio quality (see below). The scripts used for both the video and audio texts consisted of three "lecturettes" (cf. Wagner, 2010) related to the content that students were studying during the time of administration (introductory topics in sociology). These scripts were chosen to increase the ecological validity of the instrument and lessen any potential negative impact on the participants (Brown, 2012). The speaker in the video texts was one of the researchers, video-recorded from the waist up, speaking at a normal pace in a lecture style for approximately three to five minutes per text. The audio for the audio-only treatment group was recorded using high-quality audio equipment; the audio used for the video treatment group was the audio recorded with the video camera, which was judged by the researchers to be of comparable quality.

As can be seen in the Table of Specifications (see Appendix A, Table A1), the SLT was composed of 20 multiple-choice main idea and detail questions, which were operationalized as questions about the overall message or questions about specific facts mentioned in the listening passage. In both cases, questions were about information that was either explicitly stated or “unequivocally implied” (cf. Buck, 2001; Wagner, 2010). The use and format of this instrument was chosen because it followed the same format used for large-scale listening assessments at the IEP, which students were familiar with from previous listening tests (thus increasing ecological validity [Brown, 2012]), and was therefore believed to have a positive impact on participants. Three subtopics were covered that related to the theme that students were studying at the time of administration (sociology). Each question was worth one point and was scored dichotomously (i.e., one point for a correct answer and zero points for an incorrect answer). The scale of possible scores was thus 1-20 points. Internal consistency was calculated for the test using Cronbach’s alpha ($\alpha = .59$).

The study also employed a test of WM. The DST is one of the more widely used measures in studies involving WM (Juffs & Harrington, 2011) and was used to categorize participants into low, middle, and high WMGs based on the percentiles in which students scored. The test used in the study was the Cambridge Brain Sciences Digit Span Test, freely available online at <http://www.cambridgebrainsciences.com/browse/memory/test/digit-sp>. For this test, participants were presented with strings of increasingly numerous digits, which they had to memorize and repeat by typing them in the correct order. Scores could range from 0 to infinity (theoretically), and correspond to the length of the number string correctly repeated. (The Cambridge Brain Sciences website reports that the average adult score is around 7.) To the authors’ knowledge, no reliability estimate exists for this test.

Procedures

Research design. For this study, a postpositivist stance was adopted, with the aim of gathering evidence about the true nature of phenomena that exist in the world (i.e., the hypotheses about the research questions guiding the study). Although both researchers were teaching students in the IEP at the time of the study, every effort was made to maintain distance from the participants and to remain impartial, unbiased, and objective during every stage of the study. The methodology was deductive and involved quantitative statistical analysis. This method of analysis and overall methodological approach is aligned with other studies that have investigated similar phenomena (e.g., Wagner, 2010, 2013).

This quasi-experimental study employed a between-groups design, where comparisons were drawn between independent groups (Hatch & Lazaraton, 1991). Groups were assigned randomly via a coin toss. The SLT scores of the two groups were compared and conclusions were drawn regarding the effect of the presence of visual input (i.e., the videos’ texts) on listening comprehension. All participants also took the DST in a computer lab on campus. Additional comparisons were made between SLT scores and WM scores in order to determine if WM had an effect on listening comprehension and if there was an interaction between WM and the presence of visual input on listening comprehension. For all comparisons, an alpha level was set at .05.

Administration. Data was collected in two stages during regular Level 4 IEP classes. During the first stage, the DST was administered during students’ Computer-Assisted Language Learning class in a computer lab at the University. Students were assisted in signing up for a Cambridge Brain Sciences account and shown (through explanation and modeling) how to complete the DST. During the second stage, the SLT was administered during regular Level 4 IEP Listening

and Speaking classes. One section of Level 4 students took the audio-only version of the SLT and the other section took the video version.

Results

Research Question 1 investigated whether or not there were any differences in students' SLT scores when a video text was used as opposed to an audio-only text. In order to test whether or not there was a statistically significant difference between groups, the Mann-Whitney U statistical procedure was selected to compare the two groups' scores. This procedure was chosen because, after analyzing histograms of the two groups' scores, it was clear that the data were not normally distributed.

Descriptive statistics of listening comprehension scores are presented in Appendix B. Results of the Mann-Whitney U test are presented in Table 1. The observed z statistic of -2.39 can be found in the far right column. Because the observed value of -2.39 exceeds the critical value of ± 1.96 , the null hypothesis that there is no difference in listening comprehension scores between students who received video input during the test and students who received audio-only input was rejected. Effect size was calculated using η^2 and was found to have a value of .25, indicating a low / moderate degree of practical significance.

Table 1

Difference in Listening Comprehension Scores Between Video and Audio Groups

Group	n	Mean rank	z
Video	14	9.64	- 2.39*
Audio	10	16.50	

Note. Mann-Whitney U, $z_{\text{critical}} = 1.96$; * $p < .05$; $\eta^2 = .25$.

Research Question 2 investigated whether or not there were any differences in scores on the SLT among students with varying levels of WM, as determined by results of a DST. Based on the results of the DST, students were placed into one of three WMGs: low, middle, or high. Descriptive statistics of WM scores are presented in Appendix C. In order to test whether or not there was a statistically significant difference in listening comprehension scores among these three groups, the Kruskal-Wallis test was selected. Kruskal-Wallis was chosen because, after analyzing histograms of the three groups' WM scores, it was clear, once again, that the data were not normally distributed. Results of the Kruskal-Wallis test indicated no statistically significant differences among WMGs. (The observed chi-square statistic of 3.10 did not exceed the critical value of 5.99.)

Research Question 3 investigated whether or not there was an interaction between students' digit span scores and the use of video vs. audio texts on the SLT. This question was analyzed descriptively using a means plot of the two independent variables: WMG and the presence or absence of visual input (see Appendix D). Based on the descriptive analysis of the means plot, it was concluded that there was no clear interaction between the two independent variables, thus suggesting that the statistical conclusion drawn from the Mann-Whitney U test (i.e., that there was a statistical difference between groups based on the presence or absence of visual input) can be said to explain, in part, differences between the mean scores of the groups.

Discussion

In this study, it was hypothesized that the use of video would have a positive effect on listening comprehension (RQ 1). However, contrary to this hypothesis, the audio input group's scores were statistically significantly higher than the video input group's scores. One possible explanation for this finding is that the audio texts were played through higher quality speakers than the video texts. (The audio for the video group had to be played through external speakers, rather than the television speakers, due to technical difficulties.) While the researchers judged the audio quality as being comparable in each treatment, this slight difference may have impacted the results. The two classes were also given the SLT at different times of the day (morning and late afternoon), which may have affected performance on the test. The visual input group took the test in the late afternoon and may have been somewhat fatigued from a full day of classes. Additionally, although both groups of students were in the same IEP level, there may have been preexisting differences in listening proficiency between the two groups, which could have affected the results as well.

It was also hypothesized that students with higher WM scores would score higher on the SLT (RQ 2). However, the results of the Kruskal-Wallis test did not support this hypothesis, as there were no statistically significant differences in listening comprehension scores among the three WMGs. In fact, the high working memory group received the lowest scores. This somewhat unexpected result may be due to factors such as small sample size and the validity of the WM test, as the DST measured visual, as opposed to aural WM. The authors suggest that future research be conducted with larger sample sizes and different tests of WM.

Finally, it was hypothesized that if visual input had a positive effect on listening comprehension, this effect would be more prominent for learners with lower WM capacity (RQ 3). However, the plot of means (see Appendix D) did not support this hypothesis. No lines intersected, indicating no clear interaction between WM and visual input on listening comprehension. In fact, the relationship appears to be the opposite of that which was hypothesized: Students with low WM scored higher on the SLT when visual input was *not* present. On one hand, it seemed reasonable to hypothesize that visual input could aid students with lower WM by providing them with additional (i.e., visual) strategies for remembering content. On the other hand, this input may have also been overstimulating for some learners with low WM, making it difficult to cope with the cognitive demands of simultaneously attending to aural, visual, and written input (i.e., the multiple-choice questions from the SLT).

Limitations

There are several threats to the internal and external validity of this study. As previously noted, the slight difference in audio quality between the video and audio texts poses a threat to the study's internal validity. Another threat is the fact that the two classes were given the SLT at different times of the day. There are also concerns about the reliability of both measures used in this study. The Cambridge Brain Sciences Digit Span Test does not, to the researchers' knowledge, provide published information about the test's reliability. In addition, after calculating Cronbach's alpha for the sample population, the SLT was found to have a fairly low internal reliability of .59. Finally, because this study used intact groups, rather than random samples, the findings should not be generalized beyond the sample. Therefore, the results of this study should be interpreted with these limitations in mind.

Conclusion

Despite the limitations mentioned above, the results of the current study have important theoretical and practical implications. In theoretical terms, this study has further problematized

the construct of L2 listening in academic contexts, aligning with studies such as Suvorov (2009). In practical terms, this may imply that L2 listening assessment that lacks a visual component does not necessarily hinder student performance; rather, visual input may in fact be detrimental to certain students (for example, those with high WM, though this claim is extremely tentative and merits further investigation). On the other hand, such tests may be underrepresenting the construct and disadvantaging those test takers who are more adept at utilizing visual input content (cf. Ockey, 2007; Wagner, 2008). These findings thus indicate that the construct of L2 listening may be in need of a critical update, and call into question how listening is conceptualized in current models of language ability (cf. Bachman & Palmer, 2010).

Future studies should investigate whether scores on different types of listening test tasks (e.g., short-answer, note-taking) are affected by visual input. Brindley (1998), for example, discusses the processing demands of multiple-choice listening tests, noting that this test format requires test takers to listen to aural input, read test questions and possible answers, retain those answers in working memory, and match them with aural input. Therefore, it is possible that in multiple-choice tests, where the processing load is already heavy, additional visual stimuli may be distracting, rather than facilitative. However, this may not be the case in more open-ended listening tasks, where the test taker has more freedom to focus on the visual content, rather than the content of the test itself.

A qualitative component to future studies is also recommended. In particular, a mixed methods design (Johnson & Onwuegbuzie, 2004) may help clarify how students engage with visual input during testing situations and other TLU domains related to academic listening (e.g., lectures, groups discussions, labs). In the current study, both researchers noticed that test takers tended to look at their tests, rather than the video, during the first viewing of each video text. Although a few studies (e.g., Cross, 2011; Ockey, 2007; Wagner, 2008) have in fact included qualitative data on how students interact with visual input during listening comprehension tests, more research is needed in order to gain a deeper understanding of the various TLU domains that students engage with and the response types elicited in those domains (i.e., questions of situational and interactional authenticity [Bachman & Palmer, 2010; Buck, 2001]). This type of qualitative data may inform discussions of how performance on different task types is affected by visual input.

This study also has practical implications for teachers and administrators. As previously noted, visual input accompanies most of the aural input students receive outside the classroom. Therefore, it is recommended that visual input be used in both classroom activities and assessment. However, the authors suggest that teachers consider the following questions when creating or selecting classroom assessments:

- Do listening assessment tasks reflect the types of tasks that students (a) have been exposed to in the classroom and (b) will encounter outside of the classroom?
- In real life, will students be exposed to visual input for the task they are being tested on? In some tasks (e.g., taking notes during a phone call), visual input may not be necessary; in other tasks (e.g., summarizing a news program), visual input is necessary.
- What is the processing load of the listening assessment tasks? Will students be required to attend to *too much* visual input (i.e., will they be required to simultaneously attend to video texts, read test questions, and consider possible responses)?

It is hoped that the findings of this study have contributed to the growing body of research on L2 listening assessment. The researchers also hope to have provided some impetus for further research within this particular area of applied linguistics.

Author Note

Kerry Pusey, Program in Intensive English, Northern Arizona University, Flagstaff, Arizona, United States of America; Karen Lenz, English Department, Northern Arizona University, Flagstaff, Arizona, United States of America.

Kerry Pusey is now at the Insituto de Idiomas, Universidad del Norte, Barranquilla, Colombia. Karen Lenz is now at the Department of Academic English/ESL, University of California, Irvine, Irvine, California, United States of America.

This study was conducted as part of our M.A. TESL coursework at Northern Arizona University. We would like to acknowledge Dr. Joan Jamieson for her guidance and encouragement throughout this project.

Correspondence concerning this article should be addressed to Kerry Pusey, Insituto de Idiomas, Universidad del Norte, Km. 5 Vía Puerto Colombia, Barranquilla, Colombia. E-mail: kpusey@uninorte.edu.co

References

- Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, 19(4), 537-552.
<http://dx.doi.org/10.1017/S0142716400010353>
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. New York, NY: Oxford University Press.
- Biedroń, A., & Szczepaniak, A. (2012). Working memory and short-term memory abilities in accomplished multilinguals. *Modern Language Journal*, 96(2), 290-306.
<http://dx.doi.org/10.1111/j.1540-4781.2012.01332.x>
- Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. *System*, 25(1), 39-53. [http://dx.doi.org/10.1016/S0346-251X\(96\)00059-0](http://dx.doi.org/10.1016/S0346-251X(96)00059-0)
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171-191. <http://dx.doi.org/10.1017/S0267190500003536>
- Brinton, D. M. (2014). Tools and techniques of effective second/foreign language teaching. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a Second or Foreign Language* (4th ed., pp. 341-361). Boston, MA: Heinle Cengage.
- Brown, J. D. (2012). Sampling: Quantitative methods. *Encyclopedia of applied linguistics*. Oxford, UK: John Wiley & Sons.
<http://dx.doi.org/10.1002/9781405198431.wbeal1033>
- Buck, G. (2001). *Assessing listening*. New York, NY: Cambridge University Press. Available at <http://dx.doi.org/10.1017/CBO9780511732959>
- Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. *System*, 29(1), 1-14.
[http://dx.doi.org/10.1016/S0346-251X\(00\)00057-9](http://dx.doi.org/10.1016/S0346-251X(00)00057-9)
- Cross, J. (2011). Comprehending news videotexts: The influence of the visual content. *Language Learning & Technology*, 15(2), 44-68. Retrieved from <http://llt.msu.edu/issues/june2011/cross.pdf>
- Feak, C. B., & Salehzadeh, J. (2001). Challenges and issues in developing an EAP video listening placement assessment: A view from one program. *English for Specific Purposes*, 20(Suppl. 1), 477-493. [http://dx.doi.org/10.1016/S0889-4906\(01\)00021-7](http://dx.doi.org/10.1016/S0889-4906(01)00021-7)
- Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. New York, NY: Cambridge University Press.
- Gregersen, T. S. (2007). Language learning beyond words: Incorporating body language into classroom activities. *Reflections on English Language Teaching*, 6(1), 51-64. Retrieved from <http://www.nus.edu.sg/celc/research/books/relt/vol6/no1/51-64gregersen.pdf>
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston, MA: Heinle & Heinle.
- Johnson, R. B., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
<http://dx.doi.org/10.3102/0013189X033007014>
- Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, 44(2), 137-166. <http://dx.doi.org/10.1017/S0261444810000509>
- Kellerman, S. (1992). I see what you mean: The role of kinesic behaviour in listening and the implications for foreign and second language learning. *Applied Linguistics*, 13(3), 239-258.
- Londe, Z. C. (2009). The effects of video media in English as a second language listening comprehension tests. *Issues in Applied Linguistics*, 17, 41-50. Retrieved from <http://escholarship.org/uc/item/0c080191>

- Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition*, 107(2), 395-419.
<http://dx.doi.org/10.1016/j.cognition.2007.10.003>
- Morley, J. (2001). Aural comprehension instruction: Principles and practice. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed., pp. 69-86). Boston, MA: Heinle & Heinle.
- Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, 24(4), 517-537.
<http://dx.doi.org/10.1177/0265532207080771>
- Peterson, P. W. (2001). Skills and strategies for proficient listening. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed., pp. 87-100). Boston, MA: Heinle & Heinle.
- Progosh, D. (1996). Using video for listening assessment opinions of test-takers. *TESL Canada Journal*, 14(1), 34-43. Retrieved from
<http://www.teslcanadajournal.ca/index.php/tesl/article/viewFile/676/507>
- Suvorov, R. (2009). Context visuals in L2 listening tests: The effects of photographs and video vs. audio-only format. In C. A. Chapelle, H. G. Jun, & I. Katz (Eds.), *Developing and evaluating language learning materials* (pp. 53-68). Ames, IA: Iowa State University.
- Szmaliec, A., Brysbaert, M., & Duyck, W. (2012). Working memory and (second) language processing. In J. Altarriba & L. Isurin (Eds.), *Memory, language, and bilingualism: Theoretical and applied approaches* (pp. 74-94). New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139035279.004>
- Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. New York, NY: Routledge.
- Wagner, E. (2008). Video listening tests: What are they measuring? *Language Assessment Quarterly*, 5(3), 218-243. <http://dx.doi.org/10.1080/15434300802213015>
- Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing*, 27(4), 493-513. <http://dx.doi.org/10.1177/0265532209355668>
- Wagner, E. (2013). An investigation of how the channel of input and access to test questions affect L2 listening test performance. *Language Assessment Quarterly*, 10(2), 178-195.
<http://dx.doi.org/10.1080/15434303.2013.769552>
- Washburn, G. N. (2001). Using situation comedies for pragmatic language teaching and learning. *TESOL Journal*, 10(4), 21-26. Retrieved from
<http://onlinelibrary.wiley.com/doi/10.1002/j.1949-3533.2001.tb00045.x/abstract>

Appendix A

Table of Specifications

Table A1
Table of Specifications: Sociology Listening Test

Construct	Listening comprehension		# of items	# of points	% of points
Subconstructs					
Content	Details	Main Ideas			
Body Image	1, 2, 4, 6	3, 5, 7	7	7	35%
Regional Differences	8, 10, 11	9, 12, 13, 14	7	7	35%
Youth Subcultures	16, 17, 18, 19	15, 20	6	6	30%
# of items	11	9	20	20	
# of points	11	9			
% of points	55%	45%			100%

Appendix B

Descriptive Statistics of Listening Comprehension Scores

Descriptive statistics of listening comprehension test scores are presented in Tables B1 and B2. In Table B1, SLT scores are categorized by input groups (video or audio), and can be found in the left column. As can be seen in Table B1, the audio input group had a higher mean score (18.60) and lower standard deviation (1.65) than the video input group (with a mean and standard deviation of 16.79 and 1.85, respectively). In Table B2, scores are organized by WMG (low, middle, or high). As shown in Table B2, the low WMG had a higher mean score (18.50) and lower standard deviation (1.52) than the other two WMGs.

Table B1

Descriptive Statistics of Listening Comprehension Scores by Input Group

Group	<i>n</i>	<i>M</i>	<i>SD</i>	95% CI	
				LL	UL
Video	14	16.79	1.85	15.72	17.85
Audio	10	18.60	1.65	17.42	19.78
Total	24	17.54	1.96	16.72	18.37

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

Table B2

Descriptive Statistics of Listening Comprehension Scores by Working Memory Group

Group	<i>n</i>	<i>M</i>	<i>SD</i>	95% CI	
				LL	UL
Low	6	18.50	1.52	16.91	20.09
Middle	7	17.86	1.77	16.22	19.50
High	11	16.82	2.14	15.38	18.25
Total	24	17.54	1.96	16.72	18.37

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

Appendix C

Descriptive Statistics of WM Scores

Table C1 contains descriptive statistics for WM scores. WMGs can be found in the left column. For the low and middle groups, the mean scores were 4 and 5, respectively; however, standard deviations were 0, as there was no variation in scores ($n = 6$ and 7 , respectively). For the high WMG, the mean score was 7.10, with a standard deviation of 1.10 ($n = 10$). Overall, the mean score of the DST was 5.65, with a standard deviation of 1.53.

Table C1
Descriptive Statistics of Working Memory Scores

Group	<i>n</i>	<i>M</i>	<i>SD</i>	95% CI	
				LL	UL
Low	6	4	0	-	-
Middle	7	5	0	-	-
High	10	7.10	1.10	6.31	7.89
Total	23	5.65	1.53	4.99	6.31

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

Appendix D

Means Plot of WMG and Visual Input

In a means plot, if the lines cross, this indicates an interaction between the two independent variables, which would then compromise any claim of a main effect. As can be seen in Figure D1, WMGs (low, middle, and high) are plotted on the x-axis; listening comprehension scores are plotted on the y-axis, and the body of the figure contains the two visual input groups (*yes* and *no*). Looking at the low WMG, one can see that the visual input group had an average score just under 18, while the audio input group had an average score of about 20. For the middle WMG, the visual input group again had an average score just below 18, and the audio input group scored slightly higher, with a mean score just above 18. Finally, for the high WMG, the visual input group had the lowest mean of approximately 15.5, while the audio input group had a mean score of about 18.5. It should be noted that the means plots of the two visual input groups do not intersect.

Figure D1. Interaction between WM and visual input.