

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.

## **Context Visuals in L2 Listening Tests: The Effects of Photographs and Video vs. Audio-Only Format**

**Ruslan Suvorov**  
Iowa State University

The present study, conducted at a large public Midwestern university in the United States in Spring 2008, investigated the influence of context visuals, namely a single photograph and video, on test-takers' performance on a computer-based listening test developed specifically for this study. The test, consisting of six listening passages and 30 multiple-choice questions, was administered to 34 international students from three English listening classes. In particular, the study examined whether test-takers performed differently on three parts of the listening test: a part with a single photograph, a video-mediated part, and an audio-only part. In addition, the study explored whether the participants performed significantly better on the part of the listening test with their preferred type of visual input than on the other parts of the test. The results indicated that while there was no significant difference between the scores for photo-mediated and audio-only parts of the listening test, the scores on the video-enhanced part of the listening test were significantly lower. The comparison of participants' scores on the part of the listening test with their preferred type of visual input with the scores on the other parts of the test yielded mixed findings.

### **INTRODUCTION**

Visuals have been used in L2 teaching and testing for a number of decades (Coniam, 2001; Ginther, 2001, 2002; Ockey, 2007). However, the role of visual support in assessing L2 learners' listening comprehension is not well understood. Research has investigated the role of visuals in L2 listening comprehension, but the little research on the effect of visual aids in L2 listening tests has been inconclusive (Buck, 2001; Ockey, 2007), and sound theoretical perspectives on this issue remain absent (Ginther, 2001; Gruba, 1999).

Researchers tend to agree that compared to an audio medium alone, video is more authentic in terms of displaying for the listener the context, discourse, paralinguistic features, and culture that should be relevant to comprehension (e.g., Coniam, 2001). However, in spite of the putative merits of video, empirical research comparing audio and video modes in listening tests has been inconclusive (Buck, 2001; Chung, 1994; Coniam, 2001). While some studies showed that visuals can improve students' performance on listening tests (e.g., Ginther, 2002), others evinced no facilitative effect, or in some cases even a detrimental effect, of visuals on test-takers' listening comprehension (e.g.,

---

Coniam, 2001; Gruba, 1993; Ockey, 2007). In view of the readily available authoring possibilities for listening tests today, test developers are regularly faced with the choice of including visuals in listening tests, and therefore, further research is needed to better understand the role of visual support in L2 listening tests.

This study investigated the role of visual support, operationalized as a single photograph and a video, in an ESL listening test. In particular, the study examined whether a difference appeared in students' performance on the three parts of the listening test: one part accompanied by photographs, one with video, and one without any visual aids. This investigation began with a review of what constitutes listening comprehension.

### **Views of Listening Comprehension**

Numerous definitions of listening comprehension have been proposed by researchers (Brett, 1997; Rubin, 1995). While early definitions considered linguistic sound as the means of conveying information in spoken discourse (Lado, 1961), later views of listening have focused on both verbal and non-verbal stimuli (Coakley & Wolvin, 1986; Rubin, 1995). For example, Gruba (1997) and Ockey (2007) describe the process of listening comprehension as affected by the rate of speech, prosody, accent, phonology, and hesitations, as well as rhetorical signaling cues and the listeners' background knowledge. Some researchers (e.g., Buck, 2001; Wagner, 2007) argue that listening comprehension should include not only verbal input but also non-verbal components such as lip movements of the speaker, body movement, gestures, and facial expressions because in most real-life communicative situations, verbal information is often accompanied by visual information.

Visual information is believed to have a number of advantages for listening comprehension:

1. Seeing the situation and the participants increases situational and interactional authenticity, which in some cases may aid comprehension (Buck, 2001; Wagner, 2007).
2. Body language, facial expressions, and gestures of the speaker can provide additional information (Buck, 2001; Coniam, 2001; Ockey, 2007; Rubin, 1995).
3. With visual input, the listener can more easily identify the role of the speaker and the context of the situation (Baltova, 1994; Gruba, 1997; Rubin, 1995).
4. Visual elements can activate the listener's background knowledge (Ockey, 2007; Rubin, 1995).

On the other hand, because listening involves "making sense" of the received input (whether that input is aural or visual), the process draws upon the listener's cultural and educational background knowledge (Rubin, 1995, p. 151). Therefore, if the visual input does not fit into the listener's cultural expectations or background knowledge, one might

hypothesize that the visual input could confuse the listener and impede listening comprehension. Even though visual information seems to play an important role in oral communication, it is not clear exactly how listeners make use of various visual clues available in the process of communication. It has been suggested, however, that visuals can also be distracting or misleading when there is little or no relationship between what is said and what is shown (Rubin, 1995). In other words, it is not clear whether visuals can always be assumed to help comprehension.

Furthermore, the role of visual information in construct definitions of L2 listening ability in assessment is not clear and is often ignored by researchers (Buck, 2001; Gruba, 1993; Wagner, 2007). Many test developers avoid using video in listening tests because of the possible construct irrelevant variance that it might produce in test scores (Progosh, 1996). The concern is whether listening tests with visuals measure what they purport to measure, i.e., listening comprehension of language, or whether they measure some other aspects that may affect test-takers' scores such as interpretation of non-linguistic meaning in visuals. Thus, a number of researchers claim that use of visuals in listening assessment requires careful analysis of the validity, usability, and reliability (Gruba, 1997; Ockey, 2007; Wagner, 2007).

Buck (2001), for example, argues that in L2 listening tests aiming to measure test takers' ability to comprehend aural input rather than their ability to engage in interaction, test developers should be advised to avoid the use of video and to present a still picture of the participants and the scene instead. The majority of researchers, however, argue for the inclusion of non-verbal components in the construct definition of L2 listening ability, claiming that non-verbal information is an integral part of interpersonal communication in many real-life situations (e.g., Ockey, 2007; Progosh, 1996; Wagner, 2007). They assert that the exclusion of non-verbal information from listening tests might threaten their validity (Progosh, 1996; Wagner, 2007), and therefore, the listening construct in most cases needs to include the ability to obtain information from visual clues and even the ability to take notes (Ockey, 2007).

The agreement on whether visuals should be included in or excluded from the construct definition of L2 listening ability can possibly be reached if we allow for the existence of different construct definitions of L2 listening ability. If the purpose of a listening test is to measure students' ability to comprehend academic lectures in the context of a university, where the students are present in an auditorium and can both hear and see a professor, then we can argue for the inclusion of visual information in the construct definition of L2 listening ability that is being measured. However, if the purpose of a listening test is to measure students' ability to understand phone conversations that are oral-only communicative situations in which interlocutors do not exchange any visual information, we can argue that the listening test and the construct definition of L2 listening ability being measured by such a test must exclude any visual information. Therefore, a decision to include or exclude visuals from the construct definition of L2 listening ability measured by a listening test should depend on the purpose of the listening test and the communicative situation presented in the test. Leaving aside the fact that many listening

tests are intended to assess a construct that is relevant across more than one context, a question for test developers who wish to include visuals in listening test is what type of visual to include.

### **Types of Visuals**

Not all visuals are the same and accordingly they may have different (either facilitative or debilitating) effects on listening comprehension. The two main types of visuals defined in second language studies are context (or situation) visuals and content visuals (Bejar et al., 2000; Ginther, 2002). Context visuals provide information about the context of the verbal exchanges, such as the participants, the setting, and text type (e.g., a photo that depicts a man and a woman talking to each other in a classroom). The main purposes of context visuals are to set the scene for the verbal exchange and to indicate a change of speakers in a conversation (Ginther, 2002). Content visuals depict important content of the verbal interaction (e.g., a photo of Leonardo DaVinci's Mona Lisa accompanying a lecture on this painting). They can be classified into four groups: content visuals replicating the audio stimulus, content visuals illustrating the audio stimulus, content visuals organizing information in the audio stimulus, and content visuals supplementing the audio stimulus (Bejar et al., 2000). Bejar et al. (2000) maintain that the first three types of content visuals may facilitate the comprehension of the oral stimulus, whereas the last type of content visuals may make it harder.

The effect of visuals on listening comprehension may depend on their meaning and purpose. Visuals become facilitative when the language learner can interpret their meaning correctly (Chung, 1994), but they can be distracting when they decorate the text and do not convey any meaningful information (Schriver, 1997). Thus, when deciding how visuals affect listening comprehension, it is important to make a distinction between context visuals and content visuals as they provide different types of information to accompany textual or oral input.

### **Visuals in L2 Listening Tests**

In view of the content-context distinction as well as the distinction one might make between video and still images, listening tests can employ five possible modes of input: audio-only, context-only still images, context-only video, content still images, and content video. As the effect of different types of visuals on listening comprehension is not exactly clear (Coniam, 2001), it is important to investigate the role of visuals in L2 listening tests and whether the inclusion or exclusion of different types of visuals (such as images or video) from the listening tests can have an impact on test-takers' scores.

Several studies investigating the use of visuals in listening tests have been carried out during the last two decades (e.g., Coniam, 2001; Gruba, 1993; Jones, 2003). However, relative to the importance of the issue, this research has not been sufficient to provide clear results on the role of visual support in testing L2 listening, and specifically studies are needed comparing audio-only listening tests with tests that include images and video,

as well as comparative studies of different types of visuals (i.e., context and content visuals) and their effect on test-takers' performance (Ginther, 2002; Ockey, 2007).

The few comparative studies that examined the effect of visual support on test-takers' performance on L2 listening tests focused almost exclusively on context visuals. The findings of Ginther's (2002) study that employed both types of visuals suggested that the effect of visuals depended on the text types. Specifically, content visuals in mini-talks were found to be facilitative, whereas context visuals had a debilitating effect in mini-talks, no effect in dialogues/short conversations, and a facilitative effect in academic lectures. Other studies that involved only context visuals (e.g., Coniam, 2001; Gruba, 1993; Ockey, 2007) revealed neither facilitative nor detrimental effects of visuals on L2 learners' performance on listening tests. Thus, the results of the existing research on the use of visual aids in L2 listening tests appear to be inconclusive, especially with regards to context visuals that were found to be both facilitative and detrimental for test-takers' performance.

In order to better understand the effect of visuals on test-takers' performance on L2 listening tests, it is important to know how (and if) test-takers make use of the visual information presented to them during L2 listening tests. Wagner's (2007) study, which examined the amount of time test-takers made eye contact with the video monitor during an L2 listening test, suggested that in order to determine whether video can be a distraction or not, researchers need to observe the extent to which test-takers watch the video during the listening test. If some test-takers spend little, if any, time watching the computer screen during a listening test, then visual support has little or no effect on their test scores.

Finally, individual differences among test-takers, including their preferences of visual aids, might have an impact on their performance on L2 listening tests. Progosh's (1996) study, for example, looked at students' preferences of visual support for listening comprehension and found that 91.9% of the ESL students preferred video listening tests to audio-only tests. But does it mean that students who prefer video-mediated listening tests to audio-only listening tests will actually benefit from the visuals aids and receive higher test scores? Further research is needed to address this issue.

This study addressed the following three research questions:

1. Is there a statistically significant difference among types of visual input – namely a single photograph, video, and audio-only format – in an L2 listening test in terms of their effect on L2 test-takers' performance?
2. Is there a statistically significant difference between text types – namely a dialogue and a lecture – in an L2 listening test in terms of their effect on L2 test-takers' performance? If there is a difference, does the effect of visuals on test-takers' performance depend on text types?
3. Do test-takers who prefer a particular type of visual aid perform statistically

significantly better on the part of the test with their preferred type of visual than on the parts of the listening test with other types of visual input?

## **METHODOLOGY**

The questions were addressed in a quantitative research project that used a within-subjects experimental design. The use of a within-subjects design was chosen rather than a between-subjects design to avoid error variance associated with individual differences of test-takers.

The quantitative data consisted of item scores for the computer-based listening test from 34 non-native learners of English enrolled in three ESL listening classes at a public Midwestern university in the United States. Additionally, the quantitative data included the participants' responses to a post-test questionnaire concerning their preferences of visual stimuli in the listening test. The independent variables measured throughout the experiment were type of visual input (video-mediated part, photo-mediated part, and audio-only part of the listening test) and text type (dialogue and lecture), as well as the participants' preferences of visuals in the listening test (video, photograph, or audio only). The dependent variables consisted of participants' scores on each of the three parts of the listening test.

### **Participants**

The participants were non-native speakers of English enrolled in three ESL listening classes at a large public Midwestern university: one high-level listening class with students from an intensive English Program (IEP group), which is a pre-university program, and two listening classes with students enrolled in regular university classes. The overall English proficiency level of IEP students was considered to be lower than that of the students in the listening class on the basis of their TOEFL scores.

A total of 34 students participated in the study, including 12 students from the IEP group, 13 students from one listening class, and 9 students from the other listening class. The majority of the students were 18 to 20-year-old native speakers of Chinese. Out of 34 participants, nine were females and 25 were males. Most of the participants had only been in the USA for several months, and only five students had lived in the USA for a year or more.

### **Materials**

The researcher designed a computer-based listening test that consisted of six listening passages (LP) and 30 multiple-choice questions (five questions for each passage) and that lasted for 44 minutes. Figure 1 provides a screenshot of the listening section with a photograph. The online version of the test is available at [http://rsuvorov.public.iastate.edu/Listening\\_Tests/welcome.html](http://rsuvorov.public.iastate.edu/Listening_Tests/welcome.html).

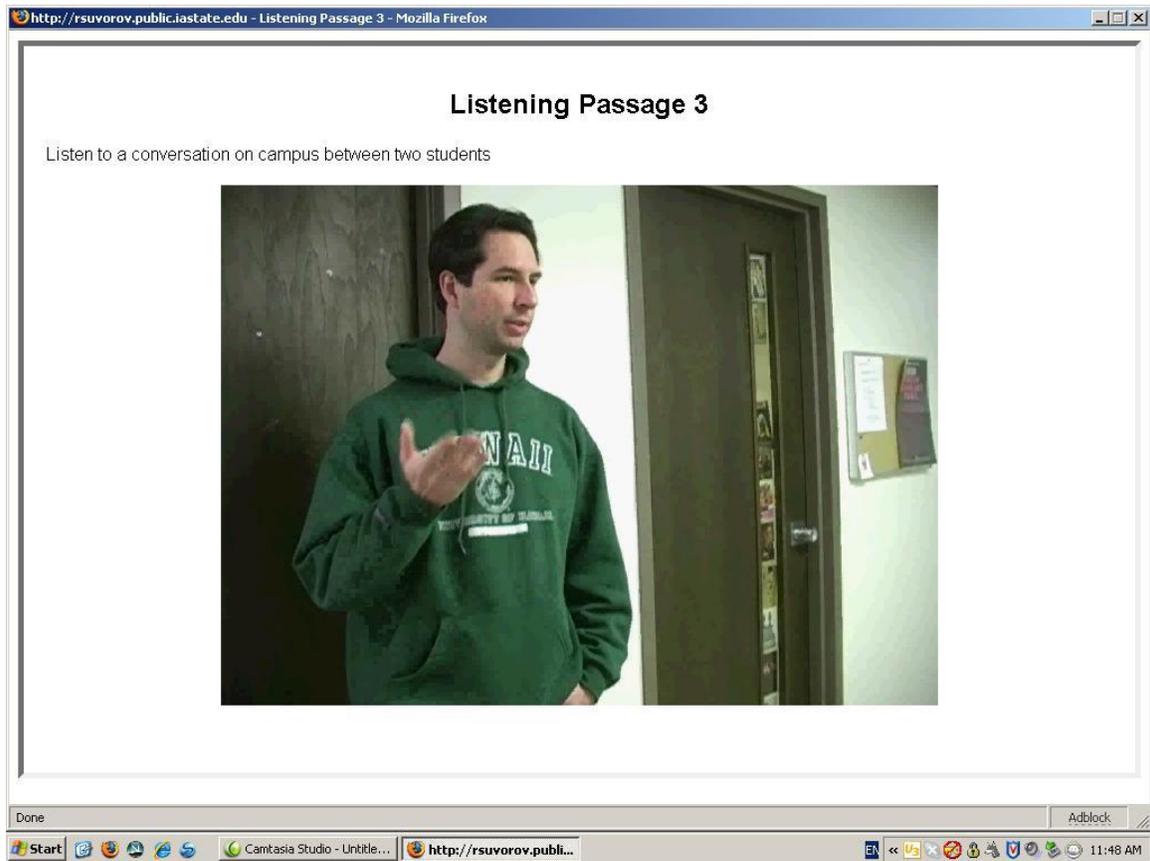


Figure 1. A screenshot of one page of the listening test

Each listening passage utilized one of two text types: a dialogue between two college students or a professor and a student (D) or a short academic lecture given by a university professor (L). In addition, the researcher used one of the three types of visual input in the test—a single photograph, video, or no visuals (i.e. audio-only format).

Table 1. Structure of the Listening Test

Audio-only part (A)		Photograph part (P)		Video part (V)	
Dialogue 1	Lecture 1	Dialogue 2	Lecture 2	Dialogue 3	Lecture 3
LP#1	LP#4	LP#6	LP#2	LP#3	LP#5

*Notes:* LP = listening passage; # - the order in which listening passages appear in the listening test

Table 1 outlines the structure of the listening test. According to the classification of visuals proposed by Bejar et al. (2001) and Ginther (2002), only context visuals were used in the listening test.

The texts of the listening passages, the length of which varied from 2.5 to 3.5 minutes, covered topics in Journalism, Linguistics, Biology, Sports and Nutrition, and History that do not require prior specialized knowledge in those areas. Test-takers had 12 seconds in between questions to answer multiple-choice items on a computer screen. They could only hear the questions and choices and then had to choose the letter that corresponded to the best answer. Taking into account that it was a test of L2 listening comprehension, the researcher decided not to display the content of questions and answers on the screen in written form as reading skills might have come into play and affected test-takers' performance and the validity of inferences about listening ability.

Five types of multiple-choice questions were used in the listening test: which is true (e.g., According to the passage, which of the following is true about pyramids?), exception (e.g., According to the passage, all of the following describe cacti EXCEPT), inference (e.g., What can be inferred about protein?), details (e.g., What is the student's main problem?), and purpose (e.g., Why does the man want to talk with the woman?).

Besides the listening test, a pre-test questionnaire and a post-test questionnaire were created. The pre-test questionnaire, consisting of 14 questions, was used to obtain information about participants' background. The post-test questionnaire that consisted of 15 open-ended questions was used to get the participants' feedback, specifically their opinions about the usefulness of visuals in the listening test. One of these questions, which was used to answer the last research question, asked test-takers to indicate the type of visual input they preferred in the listening test: video, photograph, or audio-only.

## **Procedures**

The appropriateness of the listening test for the proficiency level of the test-takers was evaluated systematically by two professors in the Department of English at the university. Additionally, the researcher conducted a pilot study to check the effectiveness of procedures related to test administration, clarity of instructions and questions, quality of audio and video recording, appropriateness of listening passages, and time constraints of the study. Three international students whose overall profiles were similar to the profiles of the students in the main study participated in the pilot study. The main finding of the pilot study was that due to some technical issues related to the speed of the Internet connection, the listening test should be administered locally from DVDs rather than online.

Before taking the listening test in a computer lab, test-takers filled out a pretest questionnaire. Because listeners, unlike readers, do not have the option of reviewing the information that has been presented to them (Thompson, 1995), the participants were given paper for taking notes during the test. Note-taking allowed test-takers to jot down main ideas or facts from the lectures and dialogues that they could later use for answering

questions. Furthermore, as multiple-choice questions were presented only in an auditory format, some test-takers in this study took the opportunity to write down questions and/or multiple-choice answers as they were listening to them and used them when providing responses on the screen. After participants finished the listening test, they were asked to fill out the post-test questionnaire.

### **Data Analysis**

The item data from 34 participants were analyzed using SPSS 16.0, and the reliability of the participants' scores on each section of the test was calculated. To answer the first research question, descriptive statistics for each part of the test were calculated and the ANOVA procedure with visual input type as the independent variable was used followed by the Tukey-Kramer test for post-hoc comparison. To answer the second research question, the researcher calculated descriptive statistics for two test sections defined by text type and conducted a one-sample t-test with text type as the independent variable. Then, an ANOVA procedure with a 3x2 factorial design (types of visual input by text types) was conducted followed by the Tukey-Kramer test for post-hoc comparison. To answer the third research question, the ANOVA procedure was used to determine whether the test-takers' scores on the part of the test with their preferred type of visual input (which the participants self-reported in the post-test questionnaire) were statistically significantly higher (at  $\alpha=.05$ ) than their scores on the parts of the listening test with other types of visuals. In this analysis, the independent variable was the participants' scores on the preferred type of visuals in the listening test (i.e., video, photograph, or audio only), and the dependent variable was their scores on each of the other two parts of the listening test.

### **RESULTS**

Internal consistency reliability (KR-20) of the listening test scores was .70, which, considering the relatively small number of participants in this study, is acceptable. The KR-20 of scores on the audio-only part of the listening test (10 items) was .54, the KR-20 of scores on the part with a single photograph (10 items) was .63, and the KR-20 of scores on the video-mediated part (10 items) was .39. The internal consistency reliability of scores on the listening passages with dialogues was .58 (15 items) and the reliability of scores on the listening passages with lectures was .56 (15 items). The results of the ANOVA indicated that there was a statistically significant difference ( $p<.05$ ) among mean scores for the test parts defined by type of visual input and for parts defined by text type. The ANOVA procedure with 3x2 factorial design (type of visual input by text type) showed that test-takers' scores on items associated with the video-mediated lecture were significantly lower than their scores on the other parts of the listening test. Overall, the test-takers' performance on the part of the listening test that contained their preferred type of visual aid was not statistically significantly higher than their performance on the other parts of the test.

### Differences Related to Visuals

The first research question addressed the differences in mean scores on the three parts of the test--audio, photograph, and video. To answer this question, test-takers' scores on the three parts of the test were compared. Table 2 presents descriptive statistics for each part based on the type of input for 34 test-takers.

Table 2. Descriptive statistics for types of visual input (n=34)

Type of visual input	Number of items	Mean	<i>SD</i>
Audio-only	10	6.35	1.98
Photograph	10	6.32	2.25
Video	10	5.06	1.74

The ANOVA procedure followed by the Tukey-Kramer test for post-hoc comparisons found that the difference between mean scores for the audio-only part of the listening test and for the part with photographs was not significant ( $p=.997$ ). However, significant differences were found between the means for audio-only and video-mediated parts ( $p=.006$ ), as well as between the means for photo vs. video-mediated parts ( $p=.008$ ).

On the basis of these results, the answer to the first research question is yes, there was a statistically significant difference in test-takers' scores when listening passages with different types of visuals were used as input. Specifically, the mean score for video-mediated passages was significantly lower than mean scores for audio-only listening passages and listening passages with photographs.

### Differences Related to Text Types

The second research question addressed the differences between text types (i.e., dialogue and lecture) and their effect on test-takers' performance on test items across different types of visuals. Table 3 presents the descriptive statistics for text types.

Table 3. Descriptive statistics for text types (n=34)

Text type	Number of items	Mean	<i>SD</i>
Dialogue	15	10.12	2.47

---

Lecture	15	7.62	2.65
---------	----	------	------

---

A one-sample t-test was run to determine if the mean difference in test-takers' performance on dialogues and lectures was statistically significant. The results revealed that test-takers' performance on dialogues was significantly better than on lectures ( $p < .0001$ ).

Due to the evidence for statistically significant difference between text types, an ANOVA procedure with 3x2 factorial design (type of visual input by text type) was used to investigate a potential interaction between type of visual and text type for the 34 test-takers. According to the results of the ANOVA, mean scores for the video-mediated lecture items were significantly lower than mean scores for items on the other five listening passages at the  $\alpha = .05$  level. However, the differences between mean scores for the video-mediated dialogue and other four listening passages with audio-only format and a photograph were not statistically significant (with p-values varying from  $p = .756$  to  $p = .969$ ) and, in fact, students performed slightly better on the video-mediated dialogue than on the audio-only lecture and the lecture with a photograph.

Thus, with regards to the second research question, the data analysis suggested that the effect of visuals on students' performance depended on the text type: while the use of video in dialogues did not have any effect on the scores, the use of video in lectures had a detrimental effect on students' performance.

### **Test-Takers' Preferences vs. Performance on Visuals**

The last research question asked whether the participants of the study performed statistically significantly better on the part of the listening test with their preferred type of visual input than on the other parts of the test.

The results of test-takers' responses to the question in the post-test questionnaire about their preferences of visuals in the listening test are given in Table 4.

Table 4. Test-takers' preferences of visuals in the listening test

---

Preferred type of visual input
--------------------------------

---

	Audio-only	Photograph	Video	Total
Number of test-takers	15	7	12	34

The ANOVA procedure was used to determine whether there was a statistically significant difference ( $p < .05$ ) between the test-takers' scores on the part of the test with their preferred type of visual input and their scores on each of the other two parts of the listening test. The results of the ANOVA revealed that overall the test-takers did not perform better on the part of the test with their preferred type of visual stimulus than on the other parts of the listening test. However, the test-takers who preferred the audio-only part of the listening test scored statistically significantly higher ( $p = .008$ ) on this part than on the video part of the listening test. It is possible that the test-takers who preferred the audio-only format of the listening test were auditory learners who could get easily distracted by visuals and, therefore, performed significantly better on the audio-only part of the listening test.

Thus, on the basis of these results, the answer to Research Question #3 was inconclusive: although overall the test-takers who preferred a particular type of visual aid did not perform statistically significantly better on the part of the test with their preferred type of visual than on the parts of the listening test with other types of visual input, the participants with a preference for the audio-only format performed significantly better on the audio-only part of the listening test than on the video part.

## LIMITATIONS

These results provide some interesting findings pertaining to the use of visuals in listening tests. However, they need to be interpreted in view of the limitations in the research. The first major limitation of this study was revealed by item analysis of the 30 multiple-choice questions used in the listening test. The results of this item analysis indicated that four out of 30 items (one item from the audio-only part, one item from the photo-mediated part, and two items from the video-mediated part of the listening test) had negative item-total correlations, which means that students who did poorly on the test overall tended to answer these four questions correctly. The deletion of these items would have raised the internal consistency reliability (KR-20) of the listening test up to .71-.72. Another way to deal with these four items would be to modify them to improve the overall internal consistency reliability of the test.

The authenticity of the listening test used in this study is also a concern. As the audio texts used in this study were designed and produced by the researcher for the study, they might lack authenticity, which according to Gruba (1997) is important in listening tests

and, thus, authentic texts (i.e., the texts with a high “degree of correspondence of the characteristics of a given language test task to the features of a T[arget] L[anguage] U[se] task” (Bachman & Palmer, 1996, p. 23)), are generally preferable.

Finally, some researchers argue that multiple-choice tests are less effective than short-answer tests or tests requiring an extended answer (Hearst, 2000). Unlike short-answer tests, multiple-choice tests “lend themselves to test-taking strategies, which do not evaluate the student’s understanding of the question” (Hearst, 2000, p. 31). Therefore, as the listening test designed for this study consisted only of multiple-choice questions, the results of the study could have been different, had the listening tasks required constructed responses.

## IMPLICATIONS FOR FURTHER RESEARCH

The main finding of this study was that the type of visuals used in the ESL listening test affected test-takers’ scores. The magnitude of the impact of visuals on students’ performance depended on the types of visuals used. The use of a single photograph in one part of the listening test as compared to the part of the listening test without any visual support did not make any significant difference in test-takers’ scores; however, the use of video stimulus had a negative impact on students’ performance. As this study involved only single photographs, further research is needed to evaluate the effect of multiple photographs in a listening test on test-takers’ performance.

In addition, different text types also appeared to have an effect on test-takers’ results on the listening test. While in dialogues the use of photographs and video did not affect test-takers’ performance, in lectures the use of video appeared to be detrimental. The use of photographs did not seem to make any difference. Therefore, more research is needed to determine whether the use of video with different text types affects students’ performance on L2 listening tests differently.

Furthermore, except for those who preferred the audio-only format, participants did not score statistically significantly higher on the part of the test with their preferred type of visual input than on the other parts of the listening test. In other words, the participants’ preferences of photographs and video did not correspond with their performance on these types of visuals. If future studies corroborate this finding, further research would be necessary to determine the reasons why the students who prefer a photograph or a video do not perform better on the listening tests that use these visuals. These reasons might include but not be limited to individual differences, such as the role of cognitive load within the test-takers’ visual and acoustic information processing systems (Mayer, 1997), learning styles of the test-takers, and their L2 proficiency levels.

Finally, as only context visuals were used in this study, a comparative research study on the interaction of content and context photographs and video with different text types is required to examine their effect on test-takers’ performance on L2 listening tests.

Additionally, an interesting approach would be to use “media inclusion,” i.e., the use of graphics, video, and audio within an item or set of items in an L2 listening test (Zenisky & Sireci, 2002, p. 348). Such multimedia can be employed for better illustration of a particular context, visualization of a problem, or evaluation of a specified construct. The findings of a research study that uses visuals not only in listening passages but also in test items would greatly contribute to the understanding of the roles visuals play in listening comprehension and testing.

## REFERENCES

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review*, 50(3), 507-531.
- Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL 2000 listening framework: A working paper*. Princeton, NJ: Educational Testing Service.
- Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. *System*, 25(1), 39-53.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Chung, U. K. (1994). *The effect of audio, a single picture, multiple pictures, or video on second-language listening comprehension*. Unpublished PhD dissertation, University of Illinois at Urbana-Champaign.
- Coakley, C. G., & Wolvin, A. D. (1986). Listening in the native language. In B. H. Wing (Ed.), *Listening, reading, and writing: Analysis and application. Northeast Conference Reports: Northeast Conference on the Teaching of Foreign Languages* (pp. 11-42). Middlebury, VT: The Conference.
- 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, 1-14.
- Ginther, A. (2001). *Effects of the presence and absence of visuals on performance on TOEFL CBT listening-comprehension stimuli* (Report 66). Princeton, New Jersey: Educational Testing Service.

- Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, 19(2), 133-167.
- Gruba, P. (1993). A comparison study of audio and video in language testing. *JALT Journal*, 15(1), 85-88.
- Gruba, P. (1997). The role of video media in listening assessment. *System*, 25(3), 335-345.
- Gruba, P. (1999). *The role of digital video media in second language listening comprehension*. Unpublished PhD dissertation, Department of Linguistics and Applied Linguistics, University of Melbourne. Retrieved May 5, 2008, from <http://eprints.unimelb.edu.au/archive/00000244/>
- Hearst, M. A. (2000). The debate on automated essay grading. *IEEE Intelligent Systems*, 15(5), 22-37.
- Jones, L. C. (2003). Supporting listening comprehension and vocabulary acquisition with multimedia annotations: The students' voice. *CALICO Journal*, 21(1), 41-65.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman.
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1), 1-19.
- Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, 24(4), 517-537.
- Progosh, D. (1996). Using video for listening assessment: Opinions of test-takers. *TESL Canada Journal*, 14(1), 34-44.
- Rubin, J. (1995). The contribution of video to the development of competence in listening. In D. Mendelsohn, & J. Rubin (Eds.), *A guide for the teaching of second language listening* (pp. 151-165). San Diego, CA: Dominic Press.
- Schraver, K. A. (1997). *Dynamics in document design: Creating text for readers*. New York, NY: John Wiley & Sons, Inc.
- Thompson, I. (1995). Assessment of second/foreign language listening comprehension. In D. Mendelsohn, & J. Rubin (Eds.), *A guide for the teaching of second language listening* (pp. 31-58). San Diego, CA: Dominic.
- Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning and Technology*, 11(1), 67-86.

Zenisky, A. L., & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, 15(4), 337-362.