Investigating Iranian Test-takers’ Performance Over Taking Different Modalities of Listening Comprehension Test

Sajad Faramarzi
Young Researchers and Elite Club, Karaj Branch,
Islamic Azad University, Karaj, Iran

Atefeh Elekaei, Reza Biria
Islamic Azad University, Khorasgan (Isfahan) Branch,
Isfahan, Iran

Visual and pictorial materials are extensively used in teaching listening comprehension textbooks. Although the use of still image is currently implemented in some distinguished testing systems, video test is not a popular modality due to the fact that this kind of test may result in test-takers’ confusion during the test. In the present study, a sample of 80 Iranian EFL (English as a Foreign Language) students was selected to compare their performance on three different modalities of listening test: audio-only test, pictorial (listening with still image test), and visual (video) listening test. The analysis of one-way repeated measure ANOVA (Analysis of Variance) was used to compare the students’ scores. The results of Bonferroni Post Hoc Test confirm that there were significant differences in scores among audio-only, pictorial, and visual tests. The results of this study highlight the greater performance of students in pictorial listening test which is statistically significant.

Keywords: visual listening test, pictorial listening test, test-taker performance, listening assessment, still image listening test, video listening test

Introduction

Statement of the Problem

As evident for many teachers and instructors, listening is an indispensable skill in second and foreign language learning. Nunan (2000) believed that listening has been underestimated by paying more attention to other skills such as speaking. For a long time, listening was not the primary source of attention by second language teachers. For many years discrete items were used by test administrators which included audio-only passages, e.g., different ® Cambridge ESOL examinations; FCE (First Certificate in English), CPE (Certificate of Proficiency in English), and CAE (Certificate in Advanced English). Similarly, listening items in TOEFL PBT (Test of English as a Foreign Language, Paper-based Test) has been organized in a discrete way by ® ETS (Educational Testing System). Since 2003, TOEFL test has undergone a metamorphosis of incorporating still images into integrated listening items. The result was the advent of the new generation of integrated listening test
INVESTIGATING IRANIAN TEST-TAKERS’ PERFORMANCE

which is TOEFL iBT (Test of English as a Foreign Language, internet-based Test). Since then, the use of some visual texts has been introduced to different conversational courses like Top Notch test and American English File proficiency test. The practice of visual texts has never been implemented officially by any recognized and reputable testing system yet. Perhaps it was because of a belief that testing listening with incorporating visual texts is equal to testing something more than the listening comprehension ability (Buck, 2001; Kellerman, 1990; Ockey, 2007; Rost, 2002). Another belief concerns the difficulty in administering and organizing visual listening test (Wagner, 2010b).

There is a general consensus that utilizing graphic representations in the form of still images, e.g., the kind of still-image representation for listening and speaking tests in TOEFL iBT, can increase listening comprehension (Baltova, 1994; Gruba, 1997; Kellerman, 1990; Kellerman, 1992; Progosh, 1996; Shin, 1998; Sueyoshi & Hardison, 2005). However, another assumption by some researchers is focused on the idea that listening with extra resources may result in listeners’ confusion and distraction. While testing, test-takers get busy with test papers and overlook watching the monitors, and spend more time on viewing the test books rather than watching the episodes on the monitor (Bejar, Douglas, Jamieson, Nissan, & Turner, 2000; Brett, 1997; Coniam, 2001; Gruba, 1997; MacWilliam, 1986; Ockey, 2007; Wagner, 2010a). Likewise, incorporating visual materials and still images to listening tests resulted in improving students’ motivation and consequently improving students’ sense of satisfaction with the use of visual materials in listening tests (Progosh, 1996; Sueyoshi & Hardison, 2005; Wagner, 2002; Wagner, 2007).

The use of still-image in listening tests had conflicting results, too. Bejar, Douglas, Jamieson, Nissan, and Turner (2000) highlighted the significance of still-images in TOEFL test as a way to predict the upcoming events and to analyse the situation in which the test took place. However, Ockey (2007) revealed that participants in his research project demonstrated little interaction with the provided still-images.

In both pictorial and visual listening tests, some researchers (e.g., Shohamy & Inbar, 1991) underscored the importance of choosing the most suitable speaking samples in the listening test which represented certain aspects of target language unit domain. Accordingly, different speaking situations like conversations in forms of dialogue and monologue, academic lectures and lecturettes should be tested in both pictorial and visual listening tests.

In the current study, three different types of listening tests: an audio-only test, still image test, and a visual test were objectively compared with each other. These tests were presented to a group of individual students who studied conversation courses at an intermediate level.

Significance of the Study and Rational for the Research

It is believed that organizing non-verbal interaction between listener and speaker(s) has been made possible by incorporating visual materials into the listening text which accordingly, increases the comprehension level of the test (Baltova, 1994; Gruba, 1997; Kellerman, 1990; Kellerman, 1992; Progosh, 1996; Shin, 1998).

Some researchers (e.g., Sueyoshi & Hardison, 2005) argued that so called “Kinesic behaviors”, i.e., the use of facial expressions and body gestures can be beneficial in comprehension level of listening texts. Raffler-Engel (1980) emphasized the “redundancy” aspect of visual materials and debated that understanding the text would be easier by watching visual materials because many parts were assumed to be understood or redundant.
Little research has been carried out on the appropriateness of the use of still image and visual texts in an EFL context and also there has been no research in comparing these two modalities. In previous studies such as Wagner (2010b) the emphasis was on using video input for lecturettes (short lectures) and dialogues among test-takers who studied English as a second language. However, taking small lectures has little compatibility with real life communication, i.e., it is hardly a representative of target language domain. Accordingly, more research is needed in order to investigate the other areas such as conversations among different groups of people or direct face to face communication.

As far as the affective factors of test-takers were concerned, participants seemed to be more comfortable with dealing with visual rather than audio-only materials for listening comprehension tests and students reacted positively to take the visual test of listening (Baltova, 1994; Dunkel, 1991; Progosh, 1996; Sueyoshi & Hardison, 2005; Wagner, 2002).

According to Wagner (2010b), there are some reasons why test-developers are not willing to implement visual materials in their listening materials. First of all, video listening test is difficult to administer and hard to execute in comparison with audio-only listening test. Secondly, there is an assumption that when the test-taker can see the visual cues of the passage, it is hardly acceptable that the listening had tested the listening ability by itself. Therefore, in order to test the listening ability alone, no inclusion of video materials are allowed. However, comprehension of listening texts in real life does involve understanding the body gestures and facial features which are the tools in order to facilitate listening comprehension.

Although video technology has been incorporated to many foreign and second language textbooks, no official test has been devised which measures the students listening abilities with visual texts so far. Students’ previous familiarity with using visual resources can have a great impact on their scores in visual listening test.

Study Design and Research Questions

The current study involved collecting and analyzing data from EFL participants. The participants took three different modalities of listening comprehension test which consisted of an audio-only test, a pictorial test, and a visual test. After administering these three tests, scores were analyzed to find the significant statistical differences. In this study, a pre-test and a post-test design were used to investigate the usefulness of implementing still-image and visual materials in foreign language listening comprehension. The pre-test was conducted to ensure the homogeneity of test-takers. A group of Iranian students was selected in this experiment. Students’ scores over three different tests were compared to examine any statistical difference among the three test-types.

Literature Review

The Use of Video and Visual Resources in Listening Comprehension Test

Numerous researchers (e.g., Baltova, 1994; Burgoon, 1994; Coniam, 2001; Gruba, 1997; Kellerman, 1990, 1992; Progosh, 1996; Shin, 1998; Sueyoshi & Hardison, 2005; Wagner, 2010a; Wagner, 2010b) have investigated the role of visual materials on ESL and EFL test-takers’ listening comprehension performance. Baltova (1994) studied 53 students of French by organizing randomly-assigned video and audio-only tests. The test types were Multiple Choice tests. It was concluded that students scored higher in the integrated video and audio questions almost twice as high as the audio-only group. The limitation for this research was that there were no pre-test and accordingly, students must have been in different levels. Baltva (1994) replicated the same
experience by organizing another multiple choice test to a different group but no statistically significant result was concluded. Similarly, Parry and Meredith (1984) organized video and audio-only multiple choice test types by randomly assigning it to 178 American students studying Spanish. Test-takers from video group scored higher but with no statistical significance.

Shin (1998) conducted a study on international students studying ESL in the United States of America. Test-takers took open-ended tests for the post test. The video group scored higher than the audio-only group and the results were statistically significant.

Sueyoshi and Hardison (2005) carried out a study on ESL intermediate and lower-intermediate students in United States. The students who were divided into two different groups of control and experimental, received the listening test in Audio-only and visual formats respectively. The type of test modality in this experiment was visual which demonstrated speakers’ physical and facial gestures. Listening comprehension tests included open-ended essay type tests which required short answers. It was concluded that the test-takers in the experimental group were scored higher than the audio-only group. However, Gruba (1997) studied 91 ESL students in United States by providing two types of tests: audio-only and the video listening tests. The test formats were based on True/False items. No outstanding difference between the two groups was observed.

The use of video listening test was examined alongside the audio-only and Multimedia listening test. The Multimedia listening test is a type of video test in which it is administered on the computer by receiving direct and immediate feedback. Brett (1997) examined three test types: audio-only, visual, and the multimedia tests on 49 European ESL students studying in UK. True/false and matching items were implemented to format the test types. It was demonstrated that among the three test types, Multimedia listening test scored higher among others and the observed difference was statistically significant. However, test-takers were scored higher in four of the six test types in video listening test and two of the six test types in the audio-only tests.

In another study, Coniam (2001) examined postgraduate teachers in Hong Kong by taking two tests: audio-only and video listening test. Essay type questions have been chosen for this test and no significant difference was observed between the two groups. Wagner (2010b) examined the effect of video materials on listening performance of ESL students. The study investigated the role of video materials by dividing students in two groups. The control group received the listening materials without any visual input and the experimental group received the same test with video materials. It was concluded that those students who took the test with video materials scored higher than those without the video materials and the differences were statistically significant.

The Implementation of Still-images Into Listening Comprehension Test

The use of image in order to facilitate the listening process has been a subject of controversy among different researchers. Bejar, Douglas, Jamieson, Nissan, and Turner (2000) identified still-image as a way to predict the situation in which pictures available in listening and speaking tests were related to TOEFL iBT test. It was concluded that there were a number of useful hints which helped the test-takers in taking the listening test. For example, if the listening section includes a still-image portraying two students talking around the campus, the listener can predict many features like the settings, the agents, the purposes, their tones, their status, their stances, and their reactions.

Test-Takers’ Engagement With Visual Materials
Ockey (2007) carried out a study in order to measure the students’ engagement with the visual test by providing pictures and videos. He found that students had little engagement with these resources while taking the test.

On the other hand, Wagner (2007) investigated the students’ viewing rate in taking the video listening test. After analyzing students’ behavior during the test, it was concluded that students had 69% eye contact with video materials comparing to total amount of time of viewing the video materials.

In another study however, Wagner (2010a) studied 56 ESL students from five different levels and examined the amount of interaction between test-takers and the video listening test by recording the test-takers during the test. It is believed that because of time which was spent to review the questions, test-takers tended to ignore watching the video materials in order to avoid losing the track of the questions. By recording test-takers during the test and analyzing the correlation between the viewing rate and the test-takers’ performance on the test, it was concluded that test-takers looked at the visual materials “less than half of the time”.

The Influence of Utilizing Visual Texts on Students’ Attitude

A number of studies have conducted to confirm the positive and negative effects of using video texts for listening test. In some studies students gave positive responses regarding the use of video tests (Progosh, 1996; Wagner, 2002; Sueyoshi & Hardison, 2005).

Kellerman (1992) introduced a new term “kinesic behavior” as a broad category that included the body movements, which included body gestures and different positions of the face. However, they referred to kinesic behavior as “co-verbal” with spoken language, which was an indispensable part of human communication as “human communicative interaction, whether the interactants are conscious of this or not” (Kellerman, 1992, p. 240).

Shohamy and Inbar (1991) proposed a range of different genres in using video listening test which are matched with target language domain. They recommended that interpersonal communication involves more non-verbal behavior in testing students’ comprehension in lectures and presentations.

In another study, Wagner (2010a) argued that the students’ strong reaction in favor of the inclusion of video listening test may indicate that this modality of listening test reduced the test-takers’ stress and anxiety. Participants reacted positively toward receiving the visual materials for and the test had been made interesting by adding video materials. Video listening materials, accordingly, made test-takers to be more “attentive” toward the test and less distractive while taking it.

The controversial results mentioned here show that more research is needed in order to investigate test-takers’ mixed feelings, different attitudes, and perceptions toward the inclusion of visual and pictorial materials (Feak & Salehzadeh, 2001).

There might be some reasons for not using video listening test and because of these explanations, test developers remained reluctant and hesitant to integrate video materials to some distinguished and reputable international tests like ® TOEFL and ® IELTS or different ® Cambridge ESOL examinations. First, the lack of resources and facilities and the difficulty in administering the test are the main shortcomings of this methodology. Second, some researchers (Kellerman, 1992) believed that testing listening with video texts, although relevant to real life communication, was equal to test something beyond listening as there were many instances in everyday
communication such as talking on the telephone which did not require any face to face correspondence.

**Video Listening Test as a Distractor Rather Than a Facilitator**

Some researchers argue that the video listening test, though look more authentic and valid, may distract students’ attention and as a result inhibit their comprehension (Alderson, Clapham, & Wall, 1995; Bejar, Douglas, Jamieson, Nissan, & Turner, 2000; MacWilliam, 1986).

Some studies examined the distractive side of video listening test. Brett (1997) showed that while taking the test, students became so busy with their test books that they ignored paying enough attention to the video materials. It showed that the inclusion of video materials had no effect on attracting test-takers’ attention. Distraction is another reason that some researchers (e.g., Coniam, 2001; Ockey, 2007) confirmed to be the result of incorporating video materials with listening test. Buck (2001) argued against the influence of video materials to listening comprehension test by stating that language ability was different from the ability to understand “subtle visual information” (p. 172). Similarly, Buck stipulated that because test-takers were different regarding their abilities to engage with the video materials, it was better to highlight the auditory aspect of the listening test rather than the visual one.

All in all, majority of researchers are in favour of using video texts with listening tests although there has been some mixed results in which the researchers remain doubtful about. Bachman (1990) and Palmer (1996) suggest that it is necessary to include “non-verbal” information to listening comprehension test in order to make it more authentic as listening in real-life situation involves non-verbal communication.

The present study sought to investigate the following questions.

Are there any significant differences among different modalities of L2 (Second Language) listening comprehension test?

Are there any significant differences between success of EFL students in pictorial and visual listening tests?

**Method**

**Participants and Setting**

In the present study, a sample of 85 Iranian EFL students, both male and female, studying English at Andisheh Nou Foreign Language Institute in Qazvin, Iran was selected.

In this foreign Language Institute, many languages like English, French, and German are taught to the students, studying English as a foreign language. In order to be certain about the homogeneity of students in, the Michigan Test of English Language Proficiency was administrated and the results were taken into account. After the pre-test, the number of participants was reduced to 80. Moreover, the students’ ages ranged from 16 to 20. Persian was reported to be the student’s first language and English was regarded as a Foreign Language course by focusing on ® American English File textbooks as the main source. Based on ACTFL (American Council on Teaching Foreign Languages) proficiency guidelines and the results of the listening pre-test, participants were ranked as intermediate level students.

**Instruments and Materials**

The pre-test and post-tests were organized to assess the ability to understand the implicit information. The post-test included three types of listening comprehension tests which were extracted from American English
Files official placement tests. The listening tests, consisted of multiple choice items, were divided into three modalities. In the first section, the students took an audio-only listening test without any complementary information. In the second test type, the students took a listening test with a still image in which a picture related to the topic was presented (see Figure 1).

![Still image listening test.](image1)

Figure 1. Still image listening test.

The purpose of the still image, according to Bejar, Douglas, Jamieson, Nissan, and Turner (2000), was to familiarize students with the settings and the situation in which the test had been taken place. The third test type was administered by providing a video track before answering the questions (see Figure 2).

![Visual (video) listening test.](image2)

Figure 2. Visual (video) listening test.

The speakers in all three test modalities were native speakers of English speaking with the American accent and the tracks have been chosen from AEF test maximizer. Each test type consisted of a conversation in a form of a dialogue and a set of multiple choice questions. The speakers in the selected dialogues read a conversation, talking interactively about real-life situations. The topic of the audio-only listening test was a dialogue taking place at the hotel. The conversation which was exchanged by the speakers concerned with asking about personal information at the reception desk and asking about the address of a particular place. For the listening test with a
still-image, the conversation took place at the shopping centre. The kind of discussion taking place included asking about the price and choosing a particular item. The topic for the visual listening test was chosen from a dialogue at the restaurant. Speakers in the third section discussed issues regarding how to order a meal. The full description of the three tests is provided in table 1.

There is no general agreement on whether students are allowed to view the questions before taking the test. As Chang and Read (2006) argued, previewing the questions before starting the listening test might increase test-takers’ performance. Test-takers were allowed to have one minute in order to read the questions before taking the test. All different test conditions and the instructions on how to do the test were given to the test-takers before starting any test type. Test materials were explicitly designed for these participants and it was revised several times.

Table 1

<table>
<thead>
<tr>
<th>Description of Three Modalities of Listening Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test modality</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Audio-only listening test</td>
</tr>
<tr>
<td>Pictorial Listening test (with still image)</td>
</tr>
<tr>
<td>Visual listening test (with video)</td>
</tr>
</tbody>
</table>

Procedures

In the current study, the Michigan Test of English Language Proficiency was administered for the pre-test to make sure that the students were in the same level. The test consisted of 50 listening items, 40 grammar items and 50 reading items. Students with (z) scores of 1 standard deviation below and above the mean were selected. After narrowing down the number of students to 80, test-takers received a test booklet including three different modalities of the listening comprehension test with an answer sheet for the post-test. Each 20-item multiple choice test, was scored as one point being allocated for each correct response and zero point for an incorrect response. The instruction on how to do different test types was given to the students by the examiner at the start of each section.

Students listened to an audio-only listening conversation for the first section. Afterwards they had to answer the multiple choice comprehension questions. The second test type was administered in a way that students were presented with a large still image on their test booklets as well as showing the same picture on a large LCD in front of them. The sound was projected by two large high quality speakers. There is no general consensus to play the track once or repeat the test for the second time. Some of the test administrators like IELTS and TOEFL do not allow the second time repetition perhaps due to the belief that naturally listening in real situations does not involve any second repetition. However, some researchers (e.g., Berne, 1995; Chang & Read, 2006) believed that repeating the listening test over different times may result in improving the comprehension level. The researchers in this study decided to avoid the second repetition and play the listening tracks only once.

Data Analysis

In order to compute the statistical analysis of this study, SPSS software version 18.0 was used. After taking
the post-test, papers were corrected meticulously and the descriptive statistics for test-takers’ scores were calculated and analyzed for each test type. Moreover, by using Cronbach’s alpha, the internal consistency of post-test was also measured. Items with high item difficulty were eliminated from the test analysis. To examine the differences among three different modality tests, one-way repeated measure ANOVA was conducted.

ANOVA was used as a procedure in testing to make sure whether the differences among the means of the three sets of scores were significant. This technique allowed statistical analysis to measure the effect of a single independent variable (main effect) and a combination of different factors (interactions) on the dependent variable. As an example, we have a single independent factor X and three groups of subjects. Therefore, the goal of this design is to determine whether the dependent variable varied as a function of the factor X in the ANOVA procedure. If the null hypothesis is true, both measures will estimate the same population variance. However, if the null hypothesis is rejected, the two or more measures will not estimate the same population variance since the first measurement will be influenced by any real effect of the factor while the other ones will not (Hatch & Farhady, 1982; Jaccard & Becker, 1997; Kirk, 1968; Shavelson, 1996; Winer, 1971).

In this study, the first null hypothesis is that there are no significant differences among different modalities of L2 listening comprehension test. The second null hypothesis, similarly, is that there are no significant differences between success of EFL students in pictorial and visual listening tests.

Results

The first research question sought to investigate whether there were any significant differences among different modalities of listening test by testing them in an EFL context. For this reason, participants were examined in an audio-only listening test, still image test, and a visual test. To examine the effect of these different tests on students’ scores, one-way repeated measure ANOVA was conducted. Tables 2a and 2b showed descriptive statistics and multivariate tests.

Table 2a
Descriptive Statistics for Modality Test

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio-only Test</td>
<td>7.1875</td>
<td>1.52692</td>
<td>80</td>
</tr>
<tr>
<td>Listening test with still image</td>
<td>8.8750</td>
<td>1.24651</td>
<td>80</td>
</tr>
<tr>
<td>Visual listening Test</td>
<td>6.8500</td>
<td>1.76535</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 2b
Multivariate Test for Modality Test

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modality Test</td>
<td>Wilks’ Lambda</td>
<td>.238</td>
<td>125.134</td>
<td>2.000</td>
<td>78.000</td>
</tr>
</tbody>
</table>

According to Tables 2a and 2b, the mean for the audio-only test was 7.18 and the standard deviation for this test was 1.52. The mean for the Listening test with still image was 8.87 and the standard deviation was 1.24. However, for the visual listening test, the mean was equal to 6.85 and the standard deviation was 1.76. The overall internal consistency reliability or the Cronbach alpha coefficient for the test was 0.82.

According to Table 2a, the means of the three tests differed significantly and also based on the results of Table 2b, F-value was significant (Sig = .000). F-value showed the effect of a test on the students’ scores.
Moreover, there were interactions among all the three tests, meaning that the kinds of tests implemented by the students had a significant effect on their performances. Then, Mauchly’s test was run in order to assess sphericity. Table 3 presented the Mauchly’s Test of Sphericity.

Table 3
Mauchly’s Test of Sphericity for Modality Test

<table>
<thead>
<tr>
<th>Within Subjects Effect</th>
<th>Mauchly's W</th>
<th>Chi-Square</th>
<th>df</th>
<th>Sig.</th>
<th>Epsilon Greenhouse-Geisser</th>
<th>Huynh-Feldt</th>
<th>Lower-bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modality_Test</td>
<td>.916</td>
<td>6.802</td>
<td>2</td>
<td>.033</td>
<td>.923</td>
<td>.944</td>
<td>.500</td>
</tr>
</tbody>
</table>

The significance value was .033 which was less than .05, so, the variances of the differences among tests were significantly different. In other words, the assumption of sphericity has been violated \((p = .033)\). Therefore, ANOVA for within-subjects variable was run and degrees of freedom were corrected using Huynh-Feldt estimate of sphericity \((\epsilon = .094)\). Table 4 represented the significance of F which was .000. So, it can be concluded that the students’ scores were significantly different in the three tests.

Table 4
ANOVA Test of Within-Subjects Effects for Modality Test

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modality_Test</td>
<td>Sphericity Assumed</td>
<td>2</td>
<td>94.163</td>
<td>103.551</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>1.846</td>
<td>102.026</td>
<td>103.551</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>1.888</td>
<td>99.748</td>
<td>103.551</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>1.000</td>
<td>188.325</td>
<td>103.551</td>
<td>.000</td>
</tr>
</tbody>
</table>

In addition, according to Table 4, the significance of the value of F was changed. It meant that the effect of tests became even more significant after correcting for sphericity. Therefore, the first null hypothesis was rejected.

In order to answer the second research question which sought to investigate differences between the audio-only test and the listening test with still image, as well as the differences between the listening test with still image and the visual listening test, Bonferroni Post Hoc Test was run. Table 5 showed the result of the Bonferroni Post Hoc test and Pair-wise comparisons.

Table 5
Bonferroni Post Hoc Test and Pair-wise Comparisons for Modality Test

<table>
<thead>
<tr>
<th>(I) Modality_Test</th>
<th>(J) Modality_Test</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig. (a)</th>
<th>95% Confidence Interval for Difference (a)</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5% Confidence Interval for Difference (a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-1.688*</td>
<td>.128</td>
<td>.000</td>
<td>-2.000 -1.375</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.338</td>
<td>.165</td>
<td>.132</td>
<td>-0.66 -0.741</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.688*</td>
<td>.128</td>
<td>.000</td>
<td>1.375 2.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.025*</td>
<td>.157</td>
<td>.000</td>
<td>1.641 2.409</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-.338</td>
<td>.165</td>
<td>.132</td>
<td>-.741 .066</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-2.025*</td>
<td>.157</td>
<td>.000</td>
<td>-2.409 -1.641</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on estimated marginal means

* The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.
By looking at the significance values in Table 4, it can be seen that the mean differences between the audio-only listening test and the still image listening test (mean = 1.688) and also between the still image test with the video test (mean = 2.025) are significantly different compared to the mean differences between the audio-only test and the video listening test (mean = .338). In other words, F-value of the audio-only test and the listening test with still image; as well as listening test with still image and the video listening test are statistically significant (Sig = .000) in comparison to the F-value of the audio-only test and the video listening test (Sig = .132). Therefore, the students performed significantly different in these two tests (i.e., the audio-only test and the listening test with the still image, as well as the listening test with still image test and the visual listening test). Therefore, the second null hypothesis is rejected, too. Figure 3 shows the differences among the three tests more clearly.

As Figure 3 shows, there are slight differences between the scores of the audio-only test and the visual listening test compared to the scores of the audio-only test and the listening test with still image, as well as between the scores of the test with still image test and the visual listening test.

**Discussion**

**Research Question 1**

Are there any significant differences among different modalities of L2 listening comprehension test among EFL students?

The first research question sought to investigate whether there are any significant differences among different modalities of listening test. The significant value in the Mauchly’s Test of Sphericity for Modality Test which is equal to 0.33 corroborates the fact that there are significant differences between three types of listening test. Moreover, the amount of F value in Huynh-Feldt estimate of sphericity confirms that students’ scores were significantly different in the three tests. Therefore the first null hypothesis is rejected. There are significant differences among different modalities of L2 listening comprehension test. In order to analyse the extent of the differences between these three tests, the mean scores should be taken into account.

In this study, interestingly, in the listening test with still image, the mean is the highest among the three tests.
which is 8.87. Listening to audio-only test is scored the second amount of mean which is 7.18. The mean for the still image test is 12.35% higher than the mean for the audio-only test. Surprisingly, the mean for the visual listening test is equal to 6.85 which is the lowest score, even less than that of audio-only test. The mean for the still image test is 12.94% higher than the mean for the visual listening test which is statistically significant.

The findings of this research show that students perform weaker in the visual listening test due to the fact that video test might have been confusing and distracting for them. The results of this study confirm the result of the previous studies (e.g., Brett, 1997; Buck, 2001; Coniam, 2001; Ockey, 2007) that the use of the video with listening comprehension test is more of a distractor rather than a facilitator. Students are undecided to follow the visual materials or to pay attention to the questions. Perhaps, some students prefer to follow the questions exclusively instead of following the video track. For those other students who follow the visual materials enthusiastically, it is estimated that it might have been difficult to associate the ideas after finishing the test.

The results of this study clearly support the study conducted by Bejar, Douglas, Jamieson, Nissan, and Turner (2000). This study confirms that the use of still-image in listening comprehension test will increase students’ understanding by predicting the upcoming events and by anticipating information such as the settings, types of characters, situation, tone, and speaker’s stance.

The findings of this research clearly indicate that there are statistically significant differences between three different modalities of the listening test which are not in accordance with some other studies (e.g., Baltva, 1994; Coniam, 2001; Gruba, 1997) in which demonstrate no significant difference between audio-only and visual listening tests.

In some previous studies (e.g., Parry & Meredith, 1984; Shin, 1998; Sueyoshi & Hardison, 2005; Wagner, 2010b), video listening test scored higher than audio-only listening test. However, in this study, the video listening test was scored less than both the audio-only and still-image listening tests.

Research Question 2

Are there any significant differences between success of EFL students in pictorial and visual listening tests?

The statistical analysis indicated that in order to find the differences between pictorial and visual modalities of listening test, the difference between the mean scores of each test should be investigated in comparison to the mean score of the audio-only test. The mean difference between audio-only test and the still image test is less than the mean difference between audio-only test and the visual listening test. It shows the greater consistency between pictorial and audio-only test that they introduce the same traits.

Conclusion

This study investigated test-takers’ performance in three different modalities of listening comprehension test: audio-only listening test, still image listening test, and visual listening test. The major finding of this study which has been conducted on a sample of 80 EFL students in Iran is that using still image in listening tests can significantly improve test-takers’ comprehension level. Although the findings of this research are not in agreement with the results of the previous studies (e.g., Parry & Meredith, 1984; Shin, 1998; Sueyoshi & Hardison, 2005; Wagner, 2010b), the statistically significant difference can validate the idea that visual materials in the form of video prompts in the listening test can confuse and distract students’ attention, and reduce their scores as a result. Having considered the fact that students received instruction and awareness on how to view the
visual materials and how to do the test afterwards, students are quite uncertain on how to allocate the test time to either watching the visual materials or paying attention to the test booklet.

**Implications of the Study in Testing**

Although the use of video is encouraged in teaching methodology in order to enhance the level of listening comprehension among ESL and EFL learners, the use of such techniques should be considered more cautiously in testing. This is a vast area and there are immense variables which should be considered in detail. Although non-verbal information which are conveyed in the form of body gestures and facial expressions are the essential characteristics of the target language use domain in listening, the use of visual materials in listening comprehension may result negatively in many different situations.

**Limitations of the Study and Recommendations for Further Research**

There are some limitations in this study that are needed to be investigated for further research. This study is particularly focused on testing students in intermediate level and in an EFL environment. Further research can be conducted on testing students in advanced levels especially those who like to take integrated tests, e.g., TOEFL iBT. The types of questions in all three tests were multiple choice tests. However, testing other different test types, e.g., open-ended test or essay types, short answer items, matching and completion items can also be good areas of investigation. However, other researchers can test students’ abilities on taking pictorial and visual listening tests by devising True/False items, matching items, and selection items. In this study, dialogue was the text genre in all three tests. Future studies investigating other different genres of texts are needed. Test-takers’ performance can be measured in monologues in a form of public speech or academic lectures. As for academic assessment, a useful study which measures students’ performances in three modalities of listening test in long run lecture test can reveal useful information. More research is needed in observing students’ note-taking abilities in long run lectures during visual listening test.

**References**


