# Using Videos in ESL Listening Achievement Tests: The Effects on Difficulty

Roman Lesnov

Northern Arizona University

Abstract

Even though modern video-technology has been used in a variety of educational contexts, second language (L2) listening comprehension testing remains one of the few areas that have made little use of video support. This study investigated the impact that video-enhanced listening passages had on English as a second language (ESL) students' test performance. The research utilized the achievement tests scores of 86 ESL students that were enrolled in an American intensive English program (EIP) in fall 2014. Students' results on video-enhanced and audio-only academic listening subtests were compared to examine the difference between the two formats in terms of their effect on test difficulty and whether this difference related to the students' proficiency level. The findings showed that the use of videos did not affect the difficulty of listening tests in general and had no interactions with proficiency level in particular. This partly supported the argument for inclusion of videos into ESL achievement tests as innocuous but helpful accompaniments.

*Keywords*: video, test, listening comprehension, proficiency

# Using Videos in ESL Listening Achievement Tests: The Effects on Difficulty

## Background

The research into the effects of using video-mediated listening passages as part of ESL listening tests on test-takers' performance has been inconclusive. A number of studies in this particular domain yielded different results, according to which they can be divided into two main groups – studies that did not show evidence of positive effects of videos on ESL students' listening comprehension and studies that did.

The studies from the first cohort (Gruba, 1993; Ockey, 2007; Cubilo &Winke, 2013; Suvorov, 2013; Batty, 2015) showed no statistically significant differences between the performance of audio-only and video groups of participants. Moreover, some researchers found that videos had a negative impact on participants' test results (Suvorov, 2009; Wagner, 2010a). The other cohort has a number of studies (Shin, 1998; Hernandez, 2004; Sueyoushi & Hardison, 2005; Wagner, 2010b; Wagner, 2013) whose findings are indicative of lesser difficulty of video-based listening comprehension tests for ESL students. The two of the studies – Batty's and Sueyoushi & Hardison's ones – bridge the two cohorts by providing evidence that there is no interaction between delivery format and test-takers' proficiency level. Overall, it is evident that no consensus on the effects of videos as part of ESL listening tests has been found.

**Research Questions**

1. Is there a difference between two types of delivery format – namely video-enhanced and audio-only – in ESL academic listening achievement tests in terms of their effect on test difficulty?

2. Is there a difference between two types of delivery format – namely video-enhanced and audio-only – in ESL academic listening achievement tests in terms of their effect on test difficulty for higher- and lower-level students?

## Methods

Scores on four ESL achievement tests of 86 students in the Program in Intensive English (PIE) in Northern Arizona University (NAU) in fall 2014 were used as a primary source of data for analysis. These students were enrolled in groups of different proficiency levels on the basis of their PIE placement test results or their previous academic work in PIE. The program had 6 levels of proficiency, level 1 being the lowest and level 6 the highest. Twenty-three of the 86 students were taking classes in Level 3, and the other 63 were in Level 5. Each group took two achievement tests during the semester, which were independent pre-developed group-unique tests.

Four PIE achievement tests – two administered to level 3 students in week 7 and week 14 respectively and the other two administered to level 5 students in week 7 and week 14 of the fall 2014 semester – were the measurement instruments used in the study. The tests were criterion-referenced assessment tools that had been designed to measure students' progress in a Listening and Speaking course as part of the curriculum and make informed decisions on students' gaps in required skills and knowledge.

There were two sections in each of the tests: listening and speaking. The week 7 level 3 listening section contained 5 testlets in total, two of which were lectures, one monologue, one dialogue, and one listening vocabulary definitions testlet. One of the listening passages was delivered in the audio-video (video-enhanced) format while the others used audio only. The topics of the passages corresponded to the instructional content covered in listening classes. Each

of the listening passages, except for the vocabulary definition part, was followed by 6 three-option multiple-choice questions of 3 types: identifying main idea, finding details, and making inferences. The performance on the listening section was scored dichotomously while the speaking production was rated according to a pre-developed rubric. For the sake of this study, only students' raw scores on the first four parts of the listening section were used.

Level 3 week 14 test, as well as two level 5 tests had analogous structure. The summary of the tests in terms of the number and type of listening passages are presented in Table 1.

Table 1

*Achievement Tests Design Summary*

| | Level 3 | | | | | | | | Level 5 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Week 7 | | | | Week 14 | | | | Week 7 | | | | | Week 14 | | | |
| Passage | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 |
| Type | M | M | M | M | M | D | M | M | M | M | M | M | M | D | M | M | M |
| Format | AV | AO | AO | AO | AV | AO | AV | AO | AV | AO | AV | AV | AV | AO | AO | AO | AO |
| Points | 6 | 6 | 6 | 6 | 7 | 8 | 8 | 9 | 6 | 6 | 5 | 5 | 6 | 7 | 7 | 8 | 8 |
| α | 0.53 | | | | 0.56 | | | | 0.52 | | | | | 0.65 | | | |

Note: M – monologue, D – dialogue; AO – audio-only, AV – audio-video; α – Cronbach's alpha internal consistency coefficient

The groups that took the achievement tests were a mix of Arabic, Chinese, and Portuguese students who intended to study one or more semesters at PIE to increase their English proficiency level.

## Results

The characteristics of the distribution of item difficulty values for each of the delivery modes were examined. Table 2 shows the distribution independently of proficiency level. It indicates that item difficulty values associated with both audio-only and audio-video modes produced negatively skewed distributions.

Table 2

*Descriptive Statistics for Item Difficulty values by Delivery Mode*

| | Audio-only group | Audio-video group |
|---|---|---|
| Sample size (N) | 43 | 71 |
| Mean (out of 1) | 0.76 | 0.77 |
| SD | 0.20 | 0.19 |
| Median | 0.79 | 0.83 |
| Skewness | -0.87 | -0.93 |
| Kurtosis | 0.34 | 0.13 |

Looking at Table 3, we can see that low level distributions displayed negative skewness for both formats. A similar pattern can be seen for high level distributions. It is notable that skewness values for low proficiency mode-dependent groups were more than two times higher than those for the high proficiency groups. This may suggest higher degree of easiness of the tests for lower level students.

Table 3

*Descriptive Statistics for Item Difficulty values by Delivery Mode and Proficiency Level*

| | Low | | High | |
|---|---|---|---|---|
| | Audio-only group | Audio-video group | Audio-only group | Audio-video group |
| Sample size (N) | 35 | 21 | 36 | 22 |
| Mean (out of 1) | 0.81 | 0.84 | 0.70 | 0.70 |
| SD | 0.19 | 0.17 | 0.19 | 0.19 |
| Median | 0.91 | 0.86 | 0.76 | 0.69 |
| Skewness | -1.37 | -2.20 | -0.65 | -0.30 |
| Kurtosis | 2.27 | 6.67 | -0.22 | -1.04 |

The first research question asked if there was a difference between video-enhanced and audio-only formats in terms of their effect on ID indices. The Mann-Whitney U test produced the *z* value of -0.30, which is less that the critical 1.96. This result yielded the conclusion that ID

indices for audio-only and video-enhanced testlets across the two proficiency groups were not statistically significantly different.

The second research question sought to find the difference between item difficulty indices for audio-only and video-enhanced listening subtests within each of the proficiency levels and to see if lower level indices were more mode-dependent than the higher ones. Z values in appeared to be much lower than the critical value, which indicated that the ID indices for audio-only and video-enhanced categories did not show statistically significant difference in either of the proficiency levels.

## Relevance to PIE

Despite the failure to find positive effects the videos in the achievement tests could have had on the tests difficulty, the findings do not leave the reader without the discussion of practical importance of the obtained results. There may be several implications to bear in mind. One possible suggestion would be for PIE teachers to not refrain from using videos in the achievement tests. Even though this study did not support a facilitating influence of videos on test difficulty, it did not undermine it either. According to some researchers, including videos can better represent the construct of listening comprehension (Ockey, 2007), as well as enhance the face validity of a measurement instrument (Bejar, Douglas, Jamieson, Nissan, and Turner, 2000). Therefore, the inclusion of videos in listening stimuli can add to the overall usefulness of a listening test. Another meaningful aspect to remember is that videos did not show bias towards level of students' proficiency. Basically, this finding may urge ESL teachers at PIE to keep using videos in achievement tests at all instructional levels without potential detrimental effects on students' comprehension. Moreover, as many studies suggested (Cubilo & Winke, 2013; Suvorov, 2009; Wagner, 2010a) students tend to have positive attitudes to having video

accompaniments as parts of listening tests. This, in turn, may positively affect test-takers' motivation and reduce anxiety.

## References

Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review, 50,* 507–531.

Batty, A. O. (2015). A comparison of video- and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing*, *32*, 3-20.

Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). TOEFL 2000 listening framework: A working paper (TOEFL Monograph Series Report No. 19). Princeton, NJ: Educational Testing Service.

Cubilo, J., & Winke, P. (2013). Redefining the L2 listening construct within an integrated writing task: Considering the impacts of visual-cue interpretation and note-taking. *Language Assessment Quarterly, 10,* 371–397.

Gruba, P. (1993). A comparison study of audio and video in language testing. *JALT Journal, 15,* 85–88.

Hernandez, S. S. (2004). The effects of video and captioned text and the influence of verbal and spatial abilities on second language listening comprehension in a multimedia learning environment (Unpublished doctoral dissertation). New York University, New York.

Latifi, M., Tavakoli, M., & A'lipour, J. (2013). Investigating the effect of video materials on testing foreign language learners' listening performance. *Middle-East Journal of Scientific Research, 13,* 1197-1201.

Miller M. D., Linn, R. L., & Gronlund, N. E. (2009). Measurement and assessment in teaching (10th ed.). Englewood Cliffs, NJ: Merrill.

Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing, 24,* 517–537.

Parry, T. S., & Meredith, R. A. (1984). Videotape vs. audiotape for listening comprehension tests: An experiment. *OMLTA Journal*. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED254107

Shin, D. (1998). Using videotaped lectures for testing academic listening proficiency. *International Journal of Listening, 12, 57–80.* Retrieved from http://shadow.cios.org:7979/JOURNALS/LIS/012/1/057/index.djvu

Stallings, W. M. (1972). A Comparison of Television and Audio Presentations of the MLA French Listening Examination. *The Journal of Educational Research, 65,* 472-474.

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55,* 661-699.

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.

Suvorov, R. (2013). Interacting with visuals in L2 listening tests: An eye-tracking study (Doctoral thesis). Iowa State University, Ames, IA.

Wagner, E. (2010a). Test-takers' interaction with an L2 video listening test. *System, 38,* 280-291.

Wagner, E. (2010b). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing, 27,* 493-513.

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,* 178-195.