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 how audio-only and video-enhanced delivery formats of listening passages compared in terms of difficulty for English as a second language (ESL) students. It utilized listening achievement tests scores of 60 ESL students that were enrolled in an American intensive English program (EIP) in the Fall 2015 semester. Students' results were used to compare the difficulty of items between testlets of different formats and determine whether this difficulty related to video type (context versus content) and students' proficiency level. The findings suggested that, at least for higher-level students, listening testlets enhanced with content videos might be significantly easier than their audio-only counterparts. On the contrary, the inclusion of context videos did not affect the difficulty of testlets in any proficiency category. The findings are discussed in terms of their practical significance for ESL teachers as well as theoretical implications for the field of ESL listening assessment.

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

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

Background

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 (e.g., Ockey, 2007; 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 (e.g., 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 – also provided evidence that there is no interaction between delivery format and test-takers' proficiency.

Few studies (e.g., Suvorov, 2014) considered different types of videos in their instruments, namely context and content videos. Context videos contain information about the context associated with the verbal interaction such as the setting and the speakers. In contrast, content pictures provide important information about the actual content of the audio stimulus (e.g., a content-related drawing or scheme, or presentation slides with content textual information). Since videos are "complex multimodal texts" (Suvorov, 2014, p. 18) they often, to a larger or lesser degree, contain elements of both context and content.

Only one study (Suvorov, 2014) controlled for the effects of video type on ESL listening comprehension by including videos that were more clearly related to either content

or context type. Suvorov found that the time test-takers attended to videos of both types did not significantly correlate with their comprehension scores.

Overall, it is evident that no consensus on the effects of videos as part of ESL listening tests has been found. What is more, the dependence of these effects on video type and test-takers' proficiency was not sufficiently explored.

Research Questions

1. Does delivery format (audio-only vs. video-enhanced) affect listening achievement testlet difficulty for lower-level students? If yes, how does it depend on video type (context vs. content)?
2. Does delivery format (audio-only vs. video-enhanced) affect listening achievement testlet difficulty for higher-level students? If yes, how does it depend on video type (context vs. content)?

Methods

Scores on two ESL achievement tests of 60 students in the Program in Intensive English (PIE) in Northern Arizona University (NAU) in Fall 2015 were used as a primary source of data. These students were enrolled in groups of different proficiency levels on the basis of their PIE placement test results. Sixteen out of 60 were Level 3 students. Their overall proficiency can be described as lower intermediate (equivalent to the TOEFL score of 32-44). There were two classes of students at Level 3 – 3A and 3B. The remaining 44 students were studying at Level 5 (intermediate to higher intermediate, TOEFL score in the range of 57-69). Level 5 consisted of four classes of students – 5A, 5B, 5D, and 5C.

Two listening tests (LT) – Level 3 LT and Level 5 LT – were developed to monitor students' progress in Listening and Speaking courses at the corresponding proficiency level. The tests fully matched the instructional objectives of the courses and were parts of the curricula. For Level 3 LT, four video passages were found on the Internet. As indicated in

Table 1, all the passages were monologic and mostly context (i.e., containing less than 50% of content clues).

Table 1

Features of Level 3 LT Video Passages

Passage	Main topic	Non-verbal cues	Content clues
Passage BPlan: Writing a **B**usiness **P**lan	An interview with an economic professor about the parts a good business plan should have. Mostly monologic.	Upper half of the body: face, gestures	0% of the time: No content clues
Passage HugeCh: A **H**uge **C**hallenge	Three short monologues by three different experts in the area of water shortages. The main topic – the challenges the lack of clean water brings and the ways to deal with them	Upper half of the body: face, gestures	0% of the time: No content clues
Passage NoWater: A Country with **N**o **W**ater	A TED-talk presentation (monologue) about the history of water challenge in Qatar and the solution to it.	Entire body: face, gestures, movements	30% of the time: pictures, numbers, text
Passage Social: A **S**ocial **E**ntrepreneur	A TED-talk presentation (monologue) about the essence of being a social entrepreneur exemplified by the story of Mohummad Yunus.	Entire body: face, gestures, movements	11% of the time: pictures, text

Then five to six multiple-choice questions of different types (i.e., main idea, detail, and inference questions) were developed for each passage. Finally, the test contained 4 listening testlets. Each testlet contained either audio-only or video-enhanced passages; audio-only passages being a product of excluding video channel from the original video passages. 3A and 3B versions of the test differed in relation to the order of format and passages as shown in Table 2.

Table 2

Features of Level 3 LT Versions

Passage	Length	Speed	Hearings	Items	Answer time, per item	Order		Format	
						3A	3B	3A	3B
NoWater	02:54	slow to medium	2	6	27 sec	1	3	V	A
Social	03:07	slow to medium	2	6	27 sec	2	4	A	V
BPlan	04:50	medium to fast	2	6	27 sec	3	1	V	A
HugeCH	03:15	medium to fast	2	6	27 sec	4	2	A	V

Note: A = audio-only; V = video-enhanced

Level 5 LT and its versions (i.e, 5A, 5B, 5C, and 5D) were created analogously with a set of another four passages. One of the video passages used in Level 5 LT was mostly content (Ethics, 63%) while the others were more of the context type, as can be concluded from Table 3.

Table 3

Features of Level 5 LT Video Passages

Passage	Main topic	Non-verbal cues	Content clues
Passage **Ethics:** "Business Ethics"	An introductory lecture (monologue) on business ethics delivered by a professor of economics.	Upper half of the body: face, gestures	63% of the time: organized text
Passage **Fraud:** "A Fraud Triangle"	An interview with an expert about the causes of fraud in a workplace and the ways to deal with it. Mostly monologic.	Upper half of the body: face, gestures	0% of the time: No content clues
Passage **Zombies:** "Economics and Zombies"	A presentation (monologue) on the topic of the relationship between economics and ethics.	Upper half of the body: face, gestures	0% of the time: No content clues
Passage **MTask:** "Multitasking and Switchtasking"	A presentation (monologue) about the ineffectiveness of switchtasking and managing one's time properly.	Entire body: face, gestures	29% of the time: pictures, text, schemes

It should also be mentioned that content information in Ethics as well as in MTask testlets could be used by test-takers to answer comprehension questions. The sequences of passages and formats as well as other features of the four versions of Level 5 LT are presented in Table 4.

Table 4

Features of Level 5 LT versions

Passage	Length	Speed	Hearings	Items	Answer time, per item	Order and Format of versions			
						5A	5B	5C	5D
Ethics	04:31	medium	1	5	25 sec	1A	4A	3V	2V
Fraud	04:52	medium to fast	1	6	25 sec	2V	3V	4A	1A
Zombies	04:49	medium	1	5	25 sec	3A	2A	1V	4V
MTask	03:05	fast	1	6	25 sec	4V	1V	2A	3A

Note: A = audio-only; V = video-enhanced

Level 3 LT and Level 5 LT were logical parts of the corresponding curricula and, thus, were administered in a routine fashion by regular teachers of Level 3 LS and Level 5 LS courses during week 12 and 13 of the Fall 2015 semester respectively. Classes of students within each level were given the corresponding versions of the tests at different occasions in keeping with schedules in operation at that time.

Results

To answer RQ1, to examine the effect of format on testlet difficulty, a two-way AVONA was run with item difficulty set as the dependent variable and format and testlet being the two fixed factors. The analysis yielded no main effect for format, F (1, 40) = 0.25, p = 0.62 < 0.05, such that the average item difficulty for audio-only testlets (M = 0.57, SD = 0.26) and video-enhanced testlets (M = 0.60, SD = 0.18) were not significantly different. Similarly, the interaction effect was not significant, F (3, 40) = 0.38, p = 0.77 < 0.05, indicating that format made no difference for any testlet condition. The magnitude of the interaction effect was miniscule (*Eta Squared* = 0.03) suggesting that only 3% of variation of the dependent variable between testlets could be explained by format.

As all the testlets in Level 3 LT were mostly of the context type, the results yielded the following interpretation. The difficulty of listening achievement testlets for lower-level students was unaffected by the presence of context video stimuli. In other words, the inclusion of context videos in listening testlets makes testlets or the whole test neither less nor more difficult.

The two-way ANOVA for RQ 2 produced the following results. The main effect of format was non-significant, F (1; 74) = 0.005, p = 0.94 < 0.05. This indicates that item difficulty averaged for all audio-only items (M = 0.71, SD = 0.18) did not show contrast with that for items from video-enhanced testlets (M = 0.71, SD = 0.21). Meanwhile, the analysis yielded significant results for the interaction between format and testlet, F (1; 74) = 3.67, p =

$0.02 < 0.05$, which called for further pairwise comparisons to locate the differences. Subsequent *post hoc* Bonferroni comparisons of each testlet by format revealed that there was a statistically significant effect of format only for Ethics testlet, F (1; 74) = 6.37, $p = 0.01 < 0.05$. This indicated a significant difference in item difficulty between the audio-only version of Ethics (M = 0.55, SD = 0.15) and Ethics in the video-enhanced format (M = 0.76, SD = 0.23).

Since only Ethics testlet was enhanced by a video of mostly content type, it can be said that the inclusion of content videos can make achievement listening testlets easier for higher-level students. In contrast, testlets' enhancement by context videos seems to bring on effect on testlets' difficulty.

Relevance to PIE and Second Language Studies

Based on the findings of this study, there may be several implications to bear in mind. One possible suggestion would be for PIE teachers to not refrain from using context videos in the achievement tests. Even though this study did not support a facilitating influence of context 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 context videos in listening stimuli can add to the overall usefulness of a listening test. Another meaningful aspect to remember is that context videos did not show bias towards level of students' proficiency. This finding may urge ESL teachers at PIE to keep using context videos in achievement tests at all instructional levels without potential detrimental effects on students' comprehension.

In case the majority of information in a video is related to the actual content of a listening passage, the exclusion of such a video from a testlet should be considered carefully.

It seems unfair to turn a passage that was originally enhanced with a content video into the audio-only format as this may increase the difficulty of the testlet for test-takers.

On a larger scale, the findings of this study may also support the argument for considering context videos a part of the academic listening comprehension construct. As the presence of content-related visual information is an essential part of L2 use in authentic academic contexts and displayed facilitative effects on listening comprehension, it may be advisable to reconsider the role of context videos in academic listening tests.

References

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.

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

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.