ORIGINAL

Using Video in Web-Based Listening Tests

Cristina Pardo-Ballester* (iD)

Department of World Languages and Cultures, Iowa State University, U.S.A. {cpardo@iastate.edu}

ABSTRACT

With sophisticated multimedia technology, there is a renewed interest in the relationship between visual and auditory channels in assessing listening comprehension (LC). Research on the use of visuals in assessing listening has emerged with inconclusive results. Some learners perform better on tests which include visual input (Wagner, 2007) while others have found no difference in the performance of participants on the two test formats (Batty, 2015). These mixed results make it necessary to examine the role of using audio and video in LC as measured by L2 listening tests. The current study examined the effects of two different types of listening support on L2 learners' comprehension: (a) visual aid in a video with input modified with redundancy and (b) no visuals (audio-only input) with input modified with redundancy. The participants of this study included 246 Spanish students enrolled in two different intermediate Spanish courses at a large Midwestern university who participated in four listening tasks either with video or with audio. Findings of whether the video serves as a listening support device and whether the course formats differ on intermediate-level Spanish learners' comprehension will be shared as well as participants' preferences with respect to listening support.

KEYWORDS: WRITING, DESIGN-BASED RESEARCH, ASYNCRONOUS PEER-REVIEW, INTERACTIVE RUBRICS

1   INTRODUCTION

The project for intermediate Spanish learners began in 2010, when Marta Lence shared the results of her listening research thesis with me. She was thrilled by the positive impact that the device of redundancy of an elaborated text had on intermediate Spanish learners when inferring information. I started to work on this video project based on Lence's thesis (2010) results because of a) my passion for working with technology; b) the scarcity of research studies on assessing listening comprehension (LC); and c) the fact that many of the studies have mixed results when assessing LC. For instance, there were research studies that had measured the language proficiency of L2 learners enrolled in two different formats: offline or classroom-based students compared to blended-learning students (Chenoweth & Murday, 2003; Chenoweth et al., 2006). None of these studies assessed the effects that the video listening test (VLT) itself might have had on understanding LC. These studies were designed to measure students' language performance comparing

---

*To whom correspondence should be addressed:
Iowa State University, World Languages and Cultures
3102 Pearson Hall
Ames, Iowa, 50011, USA

conventional and blended courses, and their listening final tests included only audio input. Moreover, there were studies that had investigated how the use of video compared to how only audio input affected students' listening performance (Coniam, 2001; Suvurov, 2008), yet no research studies were found assessing listening comprehension (LC) with the use of online video tests delivered in hybrid courses.

For this paper, the term *online-hybrid* is used to refer to hybrid or blended courses in which students meet two days in the classroom and one day online. In addition, students are required to do extra work online to compensate for the time they are not in the regular classroom. The term *face2face-blended* is used to refer to what we used to call "traditional" or "face-to-face" courses where they meet weekly face-to-face three days in the classroom and once a week face-to-face in the computer lab. The term *redundancy* is used to refer to repetition and /or paraphrasing of information. The inference term refers to the task of deducing meaning from the aural text when not specific information is not said. The study presented here aims to add to the literature on the subject of video versus audio listening assessment. Despite the fact that studies have investigated how the use of video vs. audio texts has had an impact on the performance of L2 learners on listening tests, there is a dearth of research investigating web-based listening tests for enrolled intermediate learners Spanish in *online-hybrid* courses compared to *face2face-blended* courses. Specifically, the main goals of this article are: (a) to examine the effect of eight listening tests using video and audio when looking at inference (draw implication from the text) items, (b) to find out the effect of different instruction formats (*online-hybrid* or *face2face-blended*) on learners' ability to infer content, and (c) to investigate learners' attitudes towards the use of video on listening tests.

2   LITERATURE REVIEW

Videos are often used in the L2 classroom to teach language and culture (Pardo-Ballester, 2012; Wagner, 2010a), but they seem to be used much less often in assessing LC (Coniam, 2001; Wagner, 2007, 2008, & 2010). If videos are an excellent teaching tool, why are they not used more frequently in the classroom for assessing listening? Logic dictates that if videos are used in the classroom for teaching purposes, we teachers should also use them to assess the performance of our students. Lee and Van Patten (2003) remind us that "tests should not be divorced from how one learns something" (p. 183). However, Wagner (2008) reported that L2 test developers are not passionate about the use of video because of issues related to technology and practicality, and the uncertainty of measuring listening ability. These are valid reasons for not attempting the use of video in high-stakes standardized language tests [i.e., the

Teaching of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS) or the "*Diplomas de Español como Lengua Extranjera*" - Diplomas of Spanish as a Foreign Language (DELE)], but as Wagner (2008) stated, low-stakes video language tests exist. Moreover, despite the existence of low-stakes language tests using video as a component to measure LC, there is an intriguing mismatch between the L2 instructional listening material that use audivisuals (e.g. Grgurović & Hegelheimer, 2007; Montero, et al., 2013; Pardo-Ballester, 2012; Winke, Gass, & Sydorenko, 2010) and the assessment of LC with no visuals. Even though today computer-based listening exams including visual are becoming the norm, given that in most situations target language use requires that L2 learner make use of visual stimuli (Ockey, 2007), most listening tests are administered using only audio and paper/ pencil (e.g. Batty, 2015; Lence, 2010). Lence examined the role of different input modifications, namely those of redundancy, transparency and signaling, on L2 intermediate Spanish learners' comprehension of listening texts. Redundancy, or the repetition or paraphrasing of information, was found to help participants extrapolate information. According to Lence's research, students were more likely to respond correctly to inference items when listening to redundancy-enhanced oral texts. In Wagner's (2010a) review of the impact of using video texts on test performance, he points out that Rubin's (1995) description of LC provides important insights into the concept of what the comprehension process is: "an active process in which listeners select and interpret information which comes from auditory and visual cues in order to define what is going on and what the speakers are trying to express." Wagner presents a highly informative treatment of testing listening and understanding the listening process. He emphasizes the importance of including visuals in a listening test.

With respect to L2 listening assessment, Sydorenko (2010) examined the effect of input modality (video, audio, and captions) with 26 first-year students of the Russian language. She found that the group using the video-only format (i.e., no captions, just video and audio) had better scores than the group using video with captions on recognizing aural words. However, when recognizing written words, the groups using captions outperformed the group using the video-only format. Another study by Suvorov (2008) examined: (a) whether there was a difference among different types of visual input (single photograph and a video) that affected ESL test-takers' performance, and (b) whether the use of different visuals facilitated this performance. He also looked at the test-takers' preferences. He wanted to know if test-takers' preferences corresponded to their actual scores on different listening tests. Results of his study indicated that ESL participants' performance on audio-only and still-image listening passages were significantly higher than on the video passages. In a more recent Suvorov's research (2013), he found no difference on students' performance between video and audio listening tests.

Within the same ESL context, more research studies on the use of visuals assessing listening have emerged but results were inconclusive. Ginther (2002) reported some participants performed better on tests which include visual input while Coniam (2001) had found that the use of video or audio test did not make a difference on students' performance. Ockey (2007) has reported on studies (Coniam, 2001; Ginther, 2002) and stated that a possible explanation for this disparity in findings may be the different types of visuals frequently used in listening passages. Where context-only visuals include visual input about the speaker, and the setting only meant to set the scene for communication, content-only visuals are meant to supplement the speaker's discourse by providing additional information to illustrate meaning. Ockey examined how and to what degree learners engage with context-only visuals, specifically those of still images versus video stimuli in order to determine the appropriateness of a listening test construct. Pertaining to how visuals were useful and when, participants revealed that the still images were helpful at the beginning of the text to provide a situational context, but not helpful and even distracting thereafter. As for the video, there were mixed results: three participants expressed that when they saw the speaker's lip movements they comprehend the information better; two participants used gestures to alert them of topic changes; and four used facial gestures to understand the speaker's opinion about a topic. Suvorov (2014) also investigated the video-based L2 listening assessment focusing on comparing the effect of context videos and content videos by using eye-tracking technology. No significant difference was found between context and content videos. More recently, Batty (2015) applies a many-facet Rasch model to compare the video and audio formats of a L2 listening test. The results of his study were in line with those of Coniam (2001) since he found no significant differences on the responses of the test between the two formats affected.

Other studies that also have examined the impact of the use of video on listening performance have reported other results. Wagner (2007) reported that video is not a distraction and participants performed better on tests with video input. Furthermore, participants tend to report positive attitudes toward the use of video texts (Coniam, 2001; Ockey, 2007; Suvorov, 2008; Wagner 2010a). These findings, comparing video and audio texts and students' perceptions, revealed mixed results in the literature on this topic.

2.1 Study design and research questions

According to Wagner (2010a), most of the studies examining how the use of a VLT affects L2 listening test-taker performance have used a quasi-experimental design in which one group of participants takes a VLT, and another group takes the same test but with audio-only input. Instead of doing a quasi-experimental design, a cross-over design was implemented in the current study with the purpose of having a balance between groups and also because some of these participants were enrolled in the *online-hybrid* courses instead of the *face2face-blended* courses. Participants were also grouped by proficiency level based on the classes they were enrolled in. The study design evolved by collecting test results and perceptional data from participants taking four web-based Spanish listening tasks (i.e., 2 listening tests with audio format and 2 tests with video format) during each course. Perceptional data involved a questionnaire to find out the participants' preferences with respect to the different listening assessment types. Data was collected in three different sequences (see Tables 1, 2, and 3). The use of a cross-over design could shed more light on the difficulty of web-based listening tests with audio-only compared to the same test that incorporates video. This study will also compare the students' performance on the web-based tests (i.e., with audio or video input) with participants enrolled in the same intermediate course, but delivered with two different formats, *online-hybrid* and *face2face-blended.*

Table 1. First sequence of four listening test instruments

Group 1 Face2face-blended (F2FB) N=68	Group 2 Online-hybrid (OH) N=40
Test 1 Audio Redundancy (T1AR)	Test 1 Video Redundancy (T1VR)
Test 2 Video Redundancy (T2VR)	Test 2 Audio Redundancy (T2AR)
Test 3 Audio Redundancy (T3AR)	Test 3 Video Redundancy (T3VR)
Test 4 Video Redundancy (T4VR)	Test 4 Audio Redundancy (T4AR)

Table 2. Second sequence of four listening test instruments

Group 1 Face2face-blended (F2FB) N=53	Group 2 Online-hybrid (OH) N= 32
Test 5 Video Redundancy (T5VR)	Test 5 Audio Redundancy (T5AR)
Test 6 Audio Redundancy (T6AR)	Test 6 Video Redundancy (T6VR)
Test 7 Video Redundancy (T7VR)	Test 7 Audio Redundancy (T7AR)
Test 8 Audio Redundancy (T8AR)	Test 8 Video Redundancy (T8VR)

Table 3. Third sequence of four listening test instruments

Group 1 Face2face-blended (F2FB) N= 36	Group 2 Online-hybrid (OH) N= 17
Test 1 Audio Redundancy (T1AR)	Test 1 Video Redundancy (T1VR)
Test 2 Video Redundancy (T2VR)	Test 2 Audio Redundancy (T2AR)
Test 3 Audio Redundancy (T3AR)	Test 3 Video Redundancy (T3VR)
Test 4 Video Redundancy (T4VR)	Test 4 Audio Redundancy (T4AR)

Following Lence's research (2010), it was assumed that Spanish learners at the intermediate level infer information better when redundancy is added to aural texts. The research questions in the current study addressed the effects of listening support on L2 learners' comprehension: (a) visual aid in a video-text enhanced with redundancy, (b) no visuals (audio-only input) enhanced with redundancy. Therefore, the study sought to explore three research questions:

Is there a difference between video and audio formats, in terms of their effect on intermediate level language learners' ability to infer information while listening to texts on web-based tests?

Is there a difference in the same circumstances between students enrolled in an online-hybrid course and students in a face2face-blended course while responding to the same inference items?

Do students' preferences of audio versus video format in listening tests correspond to their performance when inferring information?

3 METHOD

3.1 Participants and setting

246 Spanish learners (181 participants were female and 65 were male) at a large Midwestern university participated. Participants came from two different intermediate classes in the Spanish program, SPAN 201 and SPAN 202, which correspond to the third and fourth semesters of Spanish courses. At the time data was collected, SPAN 201 was offered only in the fall and SPAN 202 was offered in the spring. Both courses were offered in two formats: *online-hybrid* and *face2face-blended*. The participants ranged in age from 18 to 33 (a mean of 19.4 years), and the majority reported English as their native language, except for one native speaker of Arabic, four of Chinese, one of German, and one of Russian.

3.2 Materials

3.2.1 Listening test instruments

The instruments were created based on an interaction between the listening ability and the main topics that are normally taught in the second year of the Spanish language curriculum. To measure the listening ability of Spanish, three components of Buck's (2001, p. 104) framework were included in the tasks: grammatical, discourse, and sociolinguistic knowledge. Text difficulty took into account the rate of delivery of 160 words per minute for intermediate learners (Long, 1990). The instruments used consisted of eight listening tests with monologues in a Spanish target language use domain, and a total of five items of multiple-choice format for each assessment. The first multiple-choice item inquired about the main idea of the text; three items focused on vocabulary; and the last item targeted inferred meaning or information based on clear evidence from the text. In this study, the inference items are the only ones analyzed.

Each test was delivered on a web-based computer. The video and audio were embedded into a Flash file and then saved as a shockwave file to be added to the test in WebCT. The audios were saved as mp3 files and the videos as AVI files. The size for the videos was scaled down from the original 640 X 480 to 320 X 240 to ensure a well stream when playing online. Video and audio inputs were embedded in the listening test within the WebCT platform. All listening tests included a play button and it could be played only twice. See Figure 1 for the VLT (T8VR) titled *En casa*, "At home." The audio listening test (T8AR) for this topic (i.e., home) was the same, except for the inclusion of the video.

The visual input for the video format includes context and content visuals. The context visual is the title of the test projected on the screen as a caption as well as the first visual they see from the video. According to Ginther (2002), this helps to set the scene for the spoken input. Participants could see the title of the video and the visual input related to a house (See Figure 1). The content visual includes photos and videos and tends to be equivalent to the aural content. If participants hear '*olla*' ("pot"), they also see an image of that spoken word.

3.2.2 Spoken texts

The listening passages themselves were written to facilitate L2 communication beyond the classroom. The first step was to select a theme and some low-frequency utterances for the intermediate level. Second, the researcher created a recording

without preparation by using the selected utterances (e.g., 'olla'). Then, the recording was transcribed with fillers (e.g., umm, eh…) or other audible non-word fillers. The transcribed text was elaborated with redundancy to aid language comprehension (e.g., *…olla, es decir un recipiente para cocinar, "…*pot, that is a recipient to cook"). The revised script was given to the research assistant who was a Spanish native speaker from Argentina. She was asked to record the text by reading it as naturally as possible. After reading the text several times, she recorded the final spoken text. The final version was slightly changed; there were some words not present in the text and more fillers. The reason being that the speaker spoke as she normally would, and she was trying not to read the text. The same process was used for all spoken texts, except for the first text because the speaker was a different person and the video had already been developed for a different project.[i]
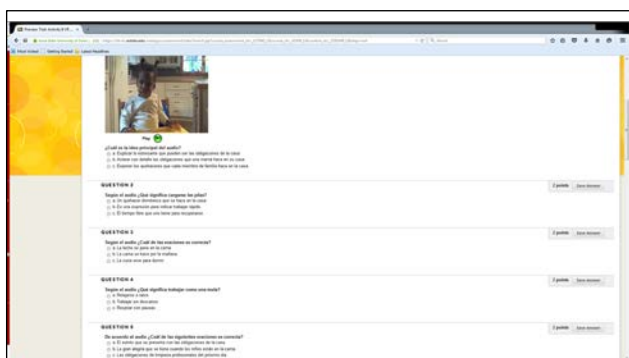

Figure 1. Test 8 video redundancy (T8VR) titled "At home"

3.2.3 Post-test questionnaire

Participants completed a questionnaire about their background and preferences in working with authentic listening multimedia exercises in WebCT. The questionnaire was created in WebCT with multiple-choice, yes or no, and open-ended questions.

3.3 Procedures

Participants were informed about the purpose of the study and all agreed to be part of the project. A consent form was signed prior to data collection. This study was conducted as a review for the listening section before each exam and took place over 15 weeks for each semester. The study included a total of three semesters from spring 2010 to spring 2011. Participants completed a questionnaire at the end of each semester to learn about their background and their perceptions of working with listening using multimedia. A problem with technology impeded the collection of perceptions data from participants in spring 2011.

The assessments were administered during normal class sessions. Each semester of the data collection, *face2face-blended* participants took four listening assessments (two using video and the other two only-audio input) in the computer language lab. *Online-hybrid* participants took the same assessments in the classroom with the aid of a laptop and headphones which were distributed by the researcher. Instructions were given to access WebCT and locate the specific assessment in a space called "audio project." A task with a pilot video was developed in WebCT and administrated before the actual data collection. Participants were told to follow the same procedures as when taking a real test in their regular classroom. The only difference was that the test was taken using WebCT. They were told that

they could listen twice whenever they were ready, but they could not stop or rewind the audio. They had access to the questions before listening to the oral input. They were told to read the questions off the screen in order to reduce their cognitive load by listening for only the pertinent information in the text. Figure 1 shows the listening review with video for Test 8 taken by Spanish 201 participants. Test 8 for audio was exactly the same, but without video.

3.4 Analysis

Odds Ratio was used to describe the association between audio enhanced with redundancy and video enhanced with redundancy to describe whether students are more or less likely to answer an item correctly. Only seven inference items were analyzed. I used a two by two table to compute the odds ratio which is a special case of logistic regression. Whereas X_{11} and X_{12} are participants who responded to the same item correctly, X_{21} and X_{22} are participants who responded to the same item incorrectly. See Table 4. "Odds" are the probability of an event occurring divided by the probability of that event not occurring. "Odds ratio" is the ratio of the odds of an event occurring in one group compared to another.

Table 4. Computation of odds ratio for audio and video groups

Possible response	Audio	Video
1	X^{11}	X^{12}
0	X^{22}	X^{23}

The formula to compute odds ratio is as follows:

$$\frac{p_1 / (1-p_1)}{p_2 / (1-p_2)} = \frac{p_1 / q_1}{p_2 / q_2} = \frac{p_1 \, q_2}{p_2 \, q_1}$$

p_1 represents the proportion of participants who responded correctly to inference items using the audio format. $1-p_1$ is the proportion of participants who responded incorrectly to inference items using the audio format. The result of $1-p_1$ is q_1. In order to calculate the odds for participants grouped in audio enhanced with redundancy, we have $p_1 / (1-p_1)$ (i.e., p_1 divided by $1-p_1$). p_2 represents the proportion of participants who responded correctly to the inference items using the video format. $1-p_2$ is the proportion of participants who responded incorrectly to inference items using the video format. The result of $1-p_2$ is q_2. In order to calculate the odds for participants grouped in video enhanced with redundancy, we have $p_2 / (1-p_2)$ (i.e., p_2 divided by $1-p_2$). If odds are higher or larger than 0, they contribute to the right answer. That is, it is more possible to get the question right.

4 RESULTS

4.1 Inference items

Table 5. Logistic regression for first sequence when the predictor variable is audio

	B	S.E.	Odds Ratio	Sig.
Audio	-0,246	0,245	0,634	0,063
Constant	1,099	0,174	3	0

Model x2 (1)= 0.45, N=108; p<.005

Table 5 represents the first sequence (spring 2010) of 108 participants of Spanish 202 who answered three[ii] inference items

using audio and video formats. The B values or the log of the odds ratio is a negative number, meaning that participants are less likely to respond correctly to inference items when using audio enhanced with redundancy. The value is not significant at a .05 level, so the model indicates that it is less possible to answer correctly when using audio redundancy.

Table 6. Logistic regression for first sequence when the predictor variable is online-hybrid

	B	S.E.	Odds Ratio	Sig.
Online-hybrid	0,15	0,257	1,162	0,558
Constant	827	0,151	2,286	0

Model x2 (1) =3.46, N=108; p<.005

Table 6 indicates that participants enrolled in *online-hybrid* courses are more likely to respond correctly to inference items by a 16% (1.16-1=0.16%), but there is no significant difference. As in Table 5, only three items were analyzed.

Tables 7 and 8 represent the second sequence of this study which took place in fall 2010 with 85 participants of Spanish 201. Results for this sequence of data indicate quite the opposite of Tables 5 and 6. Table 7 shows a positive number for the B values or the log of the odds ratio indicated that the participants enrolled in Spanish 201were more likely to respond correctly to inference items when using audio with redundancy. The B value is significant at a .05 level, so the model indicates that it is more possible to answer correctly when taking audio redundancy.

Table 7. Logistic regression for second sequence when the predictor variable is audio

	B	S.E.	Odds Ratio	Sig.
Audio	0,707	0,242	2,028	0,003
Constant	0,505	0,158	1,656	0,001

Model x2 (1) =8.77, N=85; p<.005

By looking at the negative B value in Table 8, we are able to know that participants enrolled in online-hybrid courses during fall 2010 were less likely to respond correctly to inference items, but there is no significant difference. Table 9 represents the third sequence (spring 2011) of 85 participants of Spanish 202 who answered three inference items using audio and video formats. The B values or the log of the odds ratio is a negative number, meaning that participants are less likely to respond correctly to inference items when using audio enhanced with redundancy. The value is not significant at a .05 level, so the model indicates that it is less possible to answer correctly when taking audio redundancy.

Table 8. Logistic regression for second sequence when the predictor variable is online-hybrid

	B	S.E.	Odds Ratio	Sig.
Online-hybrid	-0,248	0,241	0,78	0,304
Constant	0,93	0,152	2,533	0

Model x2 (1) =1.051, N=85; p<.005

Table 9. Logistic regression for third sequence when the predictor variable is audio

	B	S.E.	Odds Ratio	Sig.
Audio	-0,858	0,299	0,737	0,337
Constant	1,124	0,226	3,07	0,001

Model x2 (1) =8.50, N=53; p<.005

By looking at the negative B value in Table 10, we are able to know that the participants enrolled in online-hybrid courses during spring 2011 were less likely to respond correctly to the three inference items, but there is no significant difference.

Table 10. Logistic regression for third sequence when the predictor variable is online-hybrid

	B	S.E.	Odds Ratio	Sig.
Online-hybrid	-0,305	0,318	0,737	0,337
Constant	0,875	0,226	4,4	0,001

Model x2 (1) =.937, N=53; p<.005

4.2   Results for questionnaire

At the end of the semester, participants were asked about the type of activity they preferred when working with listening multimedia activities. A total of 193 participants responded to this question. Table 11 indicates their preferences. Their first choice was working with video listening in WebCT. Their second choice was also working with a video projected on a large screen. Their third choice was working with only audio in WebCT. Their final choice was working with audio in the lab or in the classroom. These results indicate that participants would rather work with video than with audio even when the instructor is the one controlling the video. See Table 12 and Table 13 for details of their preferences by their level of Spanish.

Table 11. Percentages of preference for first and second sequence in different formats

	The most preferred	Preferred	The less preferred	The least preferred
Part 1 Prefer working with video in WebCT	44%	32%	18%	7%
Part 2. Prefer working with audio in WebCT	20%	26%	33%	21%
Part 3. Prefer video projected on large screen	38%	32%	16%	13%
Part 4. Prefer audio in the classroom or lab	13%	35%	27%	24%

Note: N=193

Table 12. Percentages of preference for first sequence in different formats

	The most preferred	Preferred	The less preferred	The least preferred
Part 1 Prefer working with video in WebCT	44%	31%	20%	6%
Part 2. Prefer working with audio in WebCT	28%	28%	34%	11%
Part 3. Prefer video projected on large screen	34%	33%	15%	19%
Part 4. Prefer audio in the classroom or lab	18%	44%	25%	18%

Note: N=100

Table 13. Percentages of preference for second sequence in different formats

1. What type of activity did you prefer doing?	The most preferred	Preferred	The less preferred	The least preferred
Part 1 Prefer working with video in WebCT	44%	32%	16%	9%
Part 2. Prefer working with audio in WebCT	12%	24%	32%	32%
Part 3. Prefer video projected on large screen	43%	30%	16%	8%
Part 4. Prefer audio in the classroom or lab	9%	30%	30%	30%

Note: N=93

Participants were asked about their level of satisfaction when working with video and audio either in the classroom or in the WebCT for assessment. In general, they were satisfied or very satisfied working with listening with any format. They were the most satisfied working in the multimedia lab with video in WebCT (90.66% including answers from satisfied and very satisfied). Their second choice was using video that was projected on the large screen (88.08% including answers from satisfied and very satisfied). Their third choice was working with audio in WebCT (81.34% including answers from satisfied and very satisfied) and their last choice was listening the audio in the lab (80.30% including answers from satisfied and very satisfied). See Table 14.

Table 14. Percentages of satisfaction for first and second sequences in different formats

Q.2 How satisfied were you…	Very dissatisfied	Dissatisfied	Satisfied	Very Satisfied
…watching videos with large screen	2%	Very dissatisfied	72%	17%
…listening to audio clips in the lab?	3%	Very dissatisfied	68%	12%
…working with audio in WebCT?	4%	Very dissatisfied	68%	13%
…working with video in WebCT?	2%	Very dissatisfied	68%	23%

Note: N=193

5 DISCUSSION

5.1 Effects of video vs. audio texts

The first research question inquired about the difference between intermediate Spanish participants' performance of inferring information in web-based video tests versus web-based audio tests. Tables 5 and 9 show the degree of likelihood between participants' scores from the listening tests comparing the audio redundancy measures (i.e., three inferences items from three different tests) to the video redundancy measures (same items measured in the audio tests). Table 5 indicates the results of B values and odds ratio (B= -.246; OR=.634) for Spanish 202 in spring 2010 indicating that these participants were more likely to respond correctly to inference items in web-based video tests than in web-based audio tests. Results from Table 9 also from the same Spanish level, Spanish 202 in spring 2011 (B= -.858; OR=.737), indicated again that these participants were more likely to respond correctly to inference items in web-based video tests. Although significant differences were not found in Tables 5 and 9, considering the results from the two semesters and the large number of participants (n=161), these findings imply that the participants in this study performed better with the web-based video tests. The results of this study were in line with those of Ginther (2002) and Wagner (2007; 2010b) since they demonstrated that the use of video texts on L2 listening tests leads to a better performance in comparison to audio-only texts. It should be noted, however, in the above mentioned studies, the participants were from an ESL context. Also, in Wagner' studies, the participants did not take a computer-based listening test. A statistically significant difference at <.005 was found in the results from Table 7 (B= .707; OR=2.028), however, the study revealed that participants from Spanish 201 level were more likely to respond correctly in four inference items when taking the web-based audio-tests. Perhaps these participants were more likely to respond correctly with audio because they are less proficient compared to their counterparts (i.e., Spanish 202 participants) and therefore, they did not spend as much time listening than participants enrolled in Spanish 202. Moreover, for those with less Spanish proficiency, the video could have required more concentration while paying attention to the visuals and trying to listen in order to answer items. It should be noted that the spoken texts were elaborated with redundancy because in Lence's (2010) research, the redundancy device made a difference inferring information with the same type of students

(i.e., Spanish 201). In this study, participants came from two different levels, Spanish 201 and 202. Given that there was a difference between Spanish levels when tested with video and audio redundancy formats, these results imply that participants with the higher level of Spanish are better at inferring information with visuals and redundancy.

5.2   Effects of different delivery formats

While there was no significant difference between the effect of different delivery formats on the ability to infer information, there was a positive B value (.150) plus the odds ratio (1.162) which indicated that participants enrolled in online-hybrid courses were more likely to respond correctly to inference items. Even though the results of the odds ratio 16% (1.16-1=0.16%), is a fairly low percentage, it indicated that in spring 2010 those participants enrolled in online-hybrid courses were more likely to respond correctly to inference items than participants who were enrolled in face2face-blended courses. However, results of B values and odds ratio for fall 2010 (B= -.248; OR=.780) and spring 2011 (B= -.305; OR=.737) indicated that participants who were enrolled in online-hybrid courses were not more likely to respond correctly to inference items. The only non-significant difference found in the spring 2010 semester presents evidence that online-hybrid test-takers' performance on inference items outperformed face2face-blended test-takers' performance on the exact inference items. Results of the odds ratio analysis (Table 6, 8 and 10) suggest that more empirical research is needed in assessing listening with a variety of learning formats, not only on a face2face setting, but also distance and hybrid settings. As mentioned in the introduction of this paper, to my knowledge there are no research studies assessing listening with the use of web-based listening tests which use visuals and online-hybrid courses. So the results of this study provide insights into the concept of measuring students' L2 listening using visuals and comparing students enrolled in *online-hybrid* and *face2face-blended* courses.

5.3   Students' preferences

Table 15. Summary of percentages for first and second sequences in different formats

	The most preferred & Preferred for Span 201	The most preferred & Preferred for Span 202
1st choice video in WebCT	76,33%	75%
2nd choice video on the large screen	73,11%	67%
3rd choice audio in the lab or in class	38,16%	62%
4rd choice audio in WebCT	35.47%	56%

The results of the questionnaire showed that most participants enrolled in Spanish 201 prefer the video format when doing a listening activity (see Table 15). However, their preference for visuals doesn't correspond to their performance on inference items since the analysis of odds ratio (2.028), although statistically significant, showed that participants are more likely to respond correctly to inference items with the audio redundancy format. Regarding Spanish 202 participants'

preference, the results on the questionnaire coincide with Spanish 201 participants' preference of working with video when taking a listening test. Their preferences correspond to their performance on inference items, since the results of the odds ratio (.636 and .737) which are not statistically significant, showed that students were less likely to correctly answer the inference items when taking the listening test using audio with redundancy. Even though students' preferences for working with LC using video or audio formats don't always coincide with their performance on listening tests, it appears imperative to develop computer-based tests using visual support. First of all, in this study the participants' first choice of preference was working with web-based language testing with video. Secondly, the majority of the students tested better with the video format: 65% (or 161 out of 246) performed higher with the use of video and 34.55% (or 85 out of 246) performed lower with the use of video.

5.4   Curricular implications

Participants of Spanish 202 are more likely to respond correctly using the video format in listening tests. As for participants enrolled in the first semester of intermediate Spanish (i.e., Spanish 201), their performance on inference items indicated that they are more likely to respond correctly to audio format in listening tests. The results of this study present evidence that different formats of listening tests might lead to different performances in inferring information. The fact that participants with lower Spanish levels performed better with audio-only texts could be because: a) they have been less exposed to Spanish, and b) they are not used to listening to the spoken text, watching the visuals, and paying attention to the written items and answers all at the same time.

As other researchers have shown, test takers have positive attitudes toward the use of video texts (Coniam, 2001; Ockey, 2007; Suvorov, 2008; Wagner 2010a). Moreover, Wagner's (2010a) research emphasized the importance of having visuals in listening tests to help students improve their test performance. Ginther (2002) also showed that including content visual information can help students' performance. Ockey's (2007) findings confirmed that a still image as a context visual helped students with LC. The fact that all videos used in my study were made with context and content visual support might have helped those participants who performed better with the video.

6   CONCLUDING REMARKS

This study has attempted to contribute to L2 listening assessment by focusing on the impact of visual support on learners' test performance and their perceptions of the use of visuals and audio. Findings revealed patterns: Learners with lower proficiency levels, who answered three inference items of three different listening tests, performed better with only audio. Where, learners with higher proficiency levels, responding to four inference items each coming from four different listening tests, performed better with video. With regards to the instructional format, inconclusive findings indicated that more research in this area is needed. More conclusive results will undoubtedly be yielded by increasing the number of inference items, as well as an equal number of participants in online-hybrid and face2face-blended formats. Redundancy was a characteristic of the listening texts, but it was not investigated in this study. I began with Lence's (2010) findings on redundancy

as a device that helped Spanish learners inferring information with listening tests. It will also be beneficial to investigate redundancy as a predictor variable to learn participants' performance on listening tests with visuals. Additionally, this research reveals that, when assessing listening, video is the first choice of preference for learners regardless of their test performance. As argued before, if we teach a foreign language with the use of videos logic dictates the use of video listening tests. In other words, if we teach listening or ask our students to practice listening with the use of videos the listening tests used in the classroom should reflect what we teach and how we teach it (Lee and Van Patten, 2003). That said, teachers are encouraged to start using visual support with their students when testing listening comprehension. As Rubin (1995) pointed out, the listening process in any individual requires not only auditory cues but also visual cues for interpreting the information heard. By using visuals with listening tests, we are close to what happens in the real world. We teach our students with visual support and equip them with a range of visuals to understand the foreign language. It is imperative for any individual listening process to be proactive by including visuals not only for our instruction, but also to assess listening in short listening quizzes.

REFERENCES

Batty, A. O. (2015). A comparison of video-and audio-mediated listening testes with many-facet Rasch modeling and differential distractor functioning. *Language Testing*, *32*(1), 3-20. doi:10.1177/0265532214531254

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511732959

Chenoweth, N. A., Ushida. E, & Murday, K. (2006). Students learning in hybrid French and Spanish courses: An overview of language online. *CALICO Journal*, *24*(1), 115-145.

Chenoweth, N. A., & Murday, K. (2003). Measuring student learning in an online French course. *CALICO Journal*, *20*(2), 284-314.

Conian (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. doi:10.1016/S0346-251X(00)00057-9

Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, *19*(2), 133-167. doi:10.1191/0265532202lt225oa

Grgurović, M., & Hegelheimer, V. (2007). Help options and multimedia listening: Students' use of subtitles and the transcript. *Language Learning & Technology*, *11*(1), 45-66. Retrieved from http://llt.su.edu/vol11num1/pdf/grgurovic.pdf

Lee, J.A, & Van Patten, B. (2003). *Making communicative language teaching happen*. New York: McGraw Hill.

Lence, M. (2010). *Assisting the intermediate language listener through the use of elaborated texts* (Master's thesis). Iowa State University, Ames, Iowa, USA.

Long, D. R. (1990). What you don't know can't help you: An exploratory study of background knowledge and second language listening comprehension. *Studies in Second Language Acquisition*, *12*(1), 65-80. doi:10.1017/S0272263100008743

Ockey, G. (2007). Construct implications of including still image on computer-based listening tests. *Language Testing*, *24*(4), 517-537. doi:10.1177/0265532207080771

Montero, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*, 720-739. doi:10.1016/j.system.2013.07.013

Pardo-Ballester, C. (2012). CALL Evaluation: Students' perceptions and use of LoMásTV. *CALICO Journal*, *29*(3), 535-547. doi:10.11139/cj.29.3.532-547

Rubin, J. (1995). The contribution of video to the development of competence in listening. In D. J. Mendelsonhn & J. Rubin (Eds.), *A guide for the teaching of second language listening* (pp. 151-165). San Diego, CA: Dominie Press.

Suvorov, Russlan. (2008). *Context visuals in listening tests: The effectiveness of photographs and video vs. audio-only format* (Master's thesis). Iowa State University, Ames, Iowa, USA.

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

Suvorov, R. (2014). The use of eye tracking in research on video-based second language (L2) listening assessment: a comparison of context videos and content videos. *Language Testing*, 1-21.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73. Retrieved from http://llt.msu.edu/vol14num2/sydorenko.pdf

Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning & Technology*, *11*(1), 67–86. Retrieved from http://llt.msu.edu/vol11num1/wagner/default.html

Wagner, E. (2008). Video listening tests: What are they measuring? *Language Assessment Quarterly*, *5*, 218-243. doi:10.1080/15434300802213015

Wagner, E. (2010a). Test-takers' interaction with an L2 video listening system. *System*, *38*, 280-291. doi:10.1016/j.system.2010.01.003

Wagner, E. (2010b). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing*, *27*(4), 493-513. doi:10.1177/0265532209355668

Wagner, E. (2013). An investigation of how the channel of input and access to test questions affect L2 listening test performance. *Language Assessment Quarterly*, *10*(2), 178-195. doi:10.1080/15434303.2013.769552

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 65-86. Retrieved from http://llt.msu.edu/vol14num1/winkegasssydorenko.pdf

NOTES

[i] We started this project by using a pilot activity to ensure that the technology worked well and to make sure participants knew what to do. Then we used a video that was already created by the research assistant using a different speaker. This activity was created to show how to teach culture, and the focus was food. The speaker was a female from Colombia. Unlike the rest of the other tests, Test 1 included content visuals in terms of facial and hand gestures. These gestures seemed to facilitate comprehension. The speaker´s body on the other tests (i.e., T2VR, T3VR, T4VR, T5VR, T6VR, T7VR, and T8VR) was not in the video.

[ii] The first listening test (T1AR) was too easy for our learners and the inference item for this specific test could not be included in the analysis because the algorithm did not converge. This test was the only one that included a speaker talking in front of the camera and showing Hispanic products to the camera. The visual content was equivalent with the aural content.