



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

Roman Olegovich Lesnov<sup>a \*</sup>

<sup>a</sup> Northern Arizona University, United States

Received 2 January 2017 | Received in revised form 14 February 2017 | Accepted 19 February 2017

---

## Abstract

The question as to whether an assessment construct of second language (L2) listening comprehension should include the decoding of visual information remains unanswered (see Buck, 2001; Ockey, 2007). This study aimed to fill this gap by investigating 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 students' performance on listening achievement tests developed by the researcher. The participants were 60 low- and high-proficiency ESL students enrolled in an American intensive English program. The participants' scores on the achievement tests were used to compare the difficulty of items 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 videos containing mostly content-related visuals were significantly easier than their audio-only counterparts were. On the contrary, the inclusion of videos with mostly context visuals 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.

© 2017 EJAL & the Authors. Published by *Eurasian Journal of Applied Linguistics (EJAL)*. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Audio; second language; listening test; video; content visuals

---

## 1. Introduction

Using videos in second language (L2) listening tests is an area that has been explored since the 1970s (e.g., Stallings, 1972). Some researchers focused on comparing test-takers' results on audio-only and video-mediated listening tests of L2 other than English (e.g., Baltova, 1994; Stallings, 1972). Others focused on how the integration of video in English as a second language (ESL) listening tests would affect participants' performance (e.g., Cubilo & Winke, 2013; Wagner, 2013). However, the latter studies have not been numerous or in line with one another. Many fewer studies investigated the relationship between ESL proficiency level and test-takers' performance on listening tests of both formats (Batty, 2015; Sueyoshi & Hardison,

---

\* Roman Olegovich Lesnov. Tel.: +1-929-225-9330  
E-mail address: [roman.lesnov@nau.edu](mailto:roman.lesnov@nau.edu)  
<http://dx.doi.org/>.....

2005) only to yield another portion of conflicting results. In addition, video type was not taken into consideration by researchers with notably rare exceptions (Suvorov, 2015a). In these respects, there is little unanimous understanding. This study sought to fill these gaps by comparing difficulty of listening testlets by format (video-enhanced and audio-only) with special attention paid to the interaction between format and video type for different levels of test-takers' ESL proficiency. The findings could inform ESL teachers and test developers as to whether video complements to ESL listening achievement tests increase students' comprehension. In addition, more light could be shed on the inclusion of visuals in the construct of academic listening comprehension.

## **2. Background**

### *2.1. Visuals and listening comprehension construct*

The construct of listening comprehension has been defined in different ways. Whereas it is widely agreed that listening is defined as hearing and understanding incoming speech (e.g., Mendelsohn, 1994; Underwood, 1989), researchers differed in whether the construct definition should include critical listening cognitive skills of analysis and synthesis, or nonverbal information such as gestures, facial expression, and movements (Witkin & Trochim, 1997). The latter category takes place in face-to-face conversations or via visual input (e.g., videos).

From a theoretical perspective, as briefly reviewed by Suvorov (2015a), one trend emerges – the majority of existing theoretical models fail to view processing non-verbal input as part of listening comprehension. Germane to this majority are the views found in some scholarly groundwork. In his seminal book on language testing, Lado (1961) argued that listening comprehension should include the processing of audio input solely. According to Buck (2001), the ability to comprehend nonverbal information may be irrelevant to the listening assessment construct.

Few models seem to allow for including visual information as a potential source of construct-relevant variance. They are Wolvin and Coakley's (1996) model and a connectionist cognitive processing model (Kintsch, 1998). The former directly involves visual input in the process of listening comprehension while the latter accounts for stimuli other than audio, giving some room for the inclusion of visuals.

Empirical research findings provide less agreement about the place of visuals in the listening construct. Some studies (e.g., Batty, 2015; Gruba, 1993) suggested that the inclusion of nonverbal information could misrepresent the construct of listening comprehension. Gruba (1993) doubted that “the visual presentation of information” can “claim to have listening trait validity” (p. 87). Similarly, Batty (2015) was cautious about the inclusion of videos into the listening construct because its influence on listening comprehension remains largely inconclusive.

In contrast, a growing number of researchers (e.g., Ockey, 2007; Suvorov, 2015a; Wagner, 2007) advocated for the enhancement of the listening construct by including

nonverbal input, specifically video-enhanced stimuli, in listening tests. For example, Ockey (2007) implied that incorporating videos in listening tests would comply with a connectionist cognitive processing model of listening comprehension and increase the authenticity of a test.

Both Wagner (2007) and Suvorov (2015a) pointed out that the majority of target language use (TLU) situations involve nonverbal listening. Still, it is irrelevant for certain cases (e.g., phone conversations), which suggests that visual complements may either maintain or diminish listening construct depending on a particular listening setting. Building on this, both researchers supported the benefit of having separate definitions of the listening comprehension construct that would reflect numerous social contexts in which listening occurs. This pertains to the domain of academic listening as well. In many authentic settings (e.g., lectures, study groups, seminars), academic listening is frequently accompanied by visual information containing content clues (e.g., power point presentations, graphs, pictures related to the subject). For these settings, it is probable that the academic listening construct would benefit from including visual information.

In sum, the nature of the listening comprehension construct has been not definitive. Yet, as some theories and empirical evidence suggested, there is a rationale for further investigation of the role of visuals in the listening construct. Such an investigation can be accomplished by more closely looking at how adding a video to a listening input affects comprehension difficulty. For this, it is relevant to review existing research into the effects of using videos on ESL and English as a foreign language (EFL) listening tests difficulty as well as the dependency of these effects on language proficiency and video type.

## *2.2. Videos and ESL/EFL listening test difficulty*

The main focus of previous research was on comparing test-takers' performance on audio-only versus video-enhanced tests (Suvorov, 2015a). As a rule, test-takers' scores were intended to be indicative of the difficulty of the tests.

### *2.2.1. Effects of videos on listening comprehension*

As is the case with the listening construct, researchers are in disagreement about the effects of videos on test-takers' performance. There are three major sets of results from this line of research.

Firstly, there have been a number of studies documenting that test-takers' performance was positively influenced by the use of video-enhanced listening passages (Latifi, Tavakoli, & A'lipour, 2013; Shin, 1998; Sueyoshi & Hardison, 2005; Wagner, 2010b). Shin found that students who took audiotape formatted listening tests scored much lower than those who took videotape formatted listening tests. Sueyoshi & Hardison (2005) and Latifi et al. (2013) also found that video-enhanced tests were easier for students than their audio-only counterparts. More specifically, Sueyoshi & Hardison reported a positive effect of videos regardless of proficiency level

(low-intermediate and advanced) while Latifi et al. found that intermediate students benefited from the presence of video more than the advanced students. Wagner (2010b) found that the video group significantly outperformed the audio-only group.

The second set of results did not find a difference between the scores of audio-only and video-enhanced groups (Batty, 2015; Cubilo & Wilke, 2013; Gruba, 1993; Londe, 2009). Cubilo & Winke concluded that the use of videos did not make a difference in terms of participants' performance; moreover, it distracted test-takers from note-taking process. In addition to the general conclusion of no effect of format on listening comprehension, Batty also found no interactions between format and test-takers' proficiency levels (four tiers of proficiency).

Lastly, two studies reported a negative effect of video-enhanced listening texts on test performance. Suvorov (2009) found that video-mediated lectures were harder for students than their audio-only counterparts. Wagner (2010a) reported a negative relationship between the amounts of time test-takers attended to the video and their performance, showing that watching videos may have distracted students and led to lower comprehension scores.

### 2.2.2. *Video type*

One of the plausible explanations for the conflicting findings may be the failure to control for video type. Very few studies (i.e., Suvorov, 2015a; 2015b) considered different types of videos in their instruments, namely context and content videos. This distinction largely corresponds to the general classification of visuals discussed in Bejar, Douglas, Jamieson, Nissan and Turner (2000) and examined in relation to pictures in Ginther (2002). According to Ginther, context images contain information associated with verbal interaction unrelated to listening content (i.e., visuals showing the setting and the speakers). In contrast, content images provide important information on the actual subject of the audio stimulus (e.g., a related drawing or scheme, or presentation slides with essential information).

The distinction between content and context videos in ESL/EFL tests has rarely been the object of investigation. Few studies (Suvorov, 2015a, 2015b) controlled for the effects of video type on ESL/EFL listening comprehension by including videos that were related to either content or context type. In Suvorov (2015a), test-takers spent statistically significantly more time watching content than context videos but the difference in watching time did not affect test-takers' comprehension scores. Suvorov (2015b) found no impact of video type on test-takers' performance.

The failure to find the effect of video type may have been due to the overlap between the context and content categories. As Suvorov (2015a) concludes, these categories may not be applicable to videos with the same success since videos are "complex multimodal texts" (p. 18) that often contain elements of both context and content. Therefore, he suggests considering other dimensions of videos such as the degree of semantic congruity between audio and video inputs, rhetorical structure, and discourse type. Another, perhaps less sophisticated yet more practical, way to

approach the blurriness of the context versus content distinction would be to account for the amount of content information, or clues, in a video. For example, it may be worth exploring if a greater amount of content clues would make a difference for the difficulty of listening comprehension for test-takers.

Failing to appropriately define and take into account video types, recent research may have suffered threats to internal validity. Different studies have used different video types attempting to answer the same question. Thus, an intervening variable of video type has been underexplored, which can partly explain the conflicting research findings.

### *2.2.3. Item type*

Among other reasons for mixed findings may be the failure to control for the relationship of comprehension questions to content video input. All previous studies used comprehension items that could be answered from an audio input alone (i.e., video-independent items). Even though answering these items could be facilitated by content-related information (e.g., charts, graphs, tables, photos) in videos, watching was not necessary. Though some studies discussed to what degree content clues from videos could lead test-takers' to the correct choices (Gruba, 2004; Ockey, 2007), none attempted to use this degree as a quantifiable variable. One way to account for this would have been to measure the proportion of items within a testlet that relied on the content information from a video. This could have made comparisons of testlet difficulty by format more informative and moved the field forward in terms of investigating the role of individual items in video-enhanced listening test difficulty. In the light of Batty et al.'s (2015) finding that even context videos can affect performance on originally video-independent questions, considering the reliance of individual items on video-based content information may be of much promise.

### *2.3. Research gaps*

The present review has highlighted several issues in the literature about using videos in listening comprehension tests. First, the construct of listening comprehension may be underspecified in relation to the role of visual information in it. More research is needed to pave the way for a logical argument that the inclusion of videos will have the construct represent the TLU domain more fully. Second, no consensus has been reached on the effect of videos as part of ESL/EFL listening tests. The results of the reviewed studies give no clear picture as to whether video complements to listening test passages made their comprehension easier with respect to any proficiency level. Another conundrum, which is aggravated by the scarcity of research, is how context and content videos differ in their effects on test-takers' performance given that some comprehension questions can be answered from video-based content-related information. Furthermore, the way context and content videos have been defined may require reconsideration. It seems unreasonable to label a video a certain type if the video contains the elements of the other, which is often the case in videos. A new way to account for the mixture of context and content elements in a

video should be proposed and investigated. Finally, the degree to which items tie in with content video information and how this degree affects performance on the items remains unexplored.

#### 2.4. Research questions

The general aim of this study was to assess the effect of using videos on the difficulty of ESL academic listening achievement tests. Primarily, the study investigated whether test-takers' performance differed on segments of achievement listening tests (i.e., testlets) delivered in different formats (video-enhanced and audio-only). Secondly, the study explored the relationship between delivery format, video type, and test-takers' ESL proficiency level to see if particular video format or type favored students of a certain level.

To guide the above-set goals, the study sought to answer the following research questions:

1. Does delivery format (audio-only vs. video-enhanced) affect listening achievement testlet difficulty for lower-level students? If so, how does this effect depend on the amount of video-mediated content clues?
2. Does delivery format (audio-only vs. video-enhanced) affect listening achievement testlet difficulty for higher-level students? If so, how does this effect depend on the amount of video-mediated content clues?

### 3. Method

#### 3.1. Research design

The study investigated the effects of delivery format on the difficulty of ESL listening achievement testlets for two groups of different proficiency levels. For this, an *ex post facto* design with between-groups comparisons was employed. The data for the study were collected using secondary data from intact ESL listening and speaking (LS) classes. Students in these classes were given listening achievement tests that contained testlets in audio-only and video-enhanced formats. Item difficulty (ID) indices for each item were the object of measurement in the study.

#### 3.2. Participants

The data were obtained from 60 international students who studied in an intensive English program (IEP) at an American university in Fall 2015. These students were enrolled in classes of different proficiency levels on the basis of placement test results and, if any, their academic progress in the IEP over previous semesters. The program had six levels of proficiency, Level 1 being the lowest and Level 6 the highest. Sixteen of the 60 students were taking classes in Level 3, while the other 44 students were enrolled in Level 5. Level 3 had two classes of students (3A and 3B) and Level 5 consisted of four classes (5A, 5B, 5C, and 5D). Level 3 and Level 5 students were

approximately in the ranges of 32-44 and 57-69 on the Test of English as a Foreign Language (TOEFL iBT) scale respectively.

The participants were a mix of Arabic and Chinese students who studied for one or more semesters at the IEP before beginning mainstream undergraduate or graduate studies at the university. The majority of the students were 18-to-25-year-olds who had been in the USA for less than six months. Even though most of the students were newcomers to the program, some were continuing their studies from previous semesters. The overall group of test-takers was believed to be typical of students' population in IEPs across the USA.

### *3.3. Materials and instruments*

Two listening tests (LT) – Level 3 LT and Level 5 LT – were developed to monitor students' progress in Listening and Speaking courses at each proficiency level. The tests fully matched the instructional objectives of the courses and were parts of the curricula. Principally, Level 3 LT and Level 5 LT were achievement tests as they traced students' progress and diagnosed deficiencies to further work on. Formally, they were precursors of the course achievement tests and were referred to as “skills assessments”. They served the purpose of informing teachers and students on what listening areas might need special attention before an actual achievement test came. The format of the LTs mirrored the format of the actual achievement tests in regards to both design and content.

#### *3.3.1. Level 3 LT*

For Level 3 LT, four videos were found on the Internet that matched the course in relation to content and difficulty. They formed the bank of four video-enhanced listening passages. After that, these four passages were first duplicated and then deprived of video input in Adobe Premiere software (Version PRO CC) to form a parallel set of the same four passages in the audio-only format. Finally, passages and formats were used in different sequences for each class of students within the level.

The quality of video and audio input was controlled by the experienced judgment of the ESL instructors who were responsible for teaching and supervising corresponding LS courses. Open access video materials from YouTube and TED Talks were used. The use of the videos was in compliance with Paragraph 107 “Fair Use” of the US Copyright Law (“Copyright Law of the US”, 2011), which excluded possible infringement of copyright.

The four listening passages were on the topics covered in Level 3 LS classes – two about the problem of access to clean water in the world and the remaining two on business start-ups (see Table 1). As described above, these four passages were originally video-enhanced. Table 1 presents three pieces of information about each video passage – main topic, the amount of non-verbal communication (i.e., to what extent a speaker is seen in each video), and types and amount of content cues (i.e., any visual aids that explained or illustrated the content of listening). All the passages

were mostly monologic and showed speakers' entire bodies or halves of their heights, which means that gestures and facial expression were not a point of difference between the testlets.

What the passages differed in was the amount of content clues displayed (e.g., graphs, tables, pictures, organizers, figures, text). In the absence of a clear distinction between context and content videos in the literature, this study resorted to view this categorization as a continuum. In other words, videos were labeled by the amount of content clues in them. This amount was calculated as the ratio of time content clues were displayed in a video to the total length of the video. In this respect, passages "NoWater" and "Social" made use of content clues more extensively – around 30% and 11% of the entire passage length respectively, compared to 0% in the other two passages.

Table 1. Features of Level 3 LT video passages.

Passage	Main topic	Non-verbal cues	Content clues
Passage <b>NoWater</b> : A Country with No Water	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 <b>Social</b> : A Social Entrepreneur	A TED-talk presentation (monologue) about the meaning of being a social entrepreneur exemplified by the story of Mohammad Yunus.	Entire body: face, gestures, movements	11% of the time: pictures, text
Passage <b>BPLan</b> : Writing a Business Plan	An interview with an economics 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 <b>HugeCh</b> : A Huge Challenge	Three short monologues by three experts in the area of water shortages. The main topic – challenges the lack of clean water brings and ways to deal with them	Upper half of the body: face, gestures	0% of the time: No content clues

As indicated in Table 2 below, passages were mostly two to three minutes long, delivered twice, and had a medium speech rate as judged by the instructors of the LS course. To form four testlets, six three-option multiple-choice questions followed each of the passages and tested students' ability to identify main ideas, details, and make inferences. Each testlet contained one main idea question, two to three detail questions, and two to three inference questions. The test was scored dichotomously with 24 as a maximum possible number of points.

In testlets enhanced by videos with content clues, video input could help to answer certain questions. Text, pictures, or schemes might contain or allude to the right answer, as judged by the researcher. Thus, answers to some items of testlets "NoWater" (3 out of 6) and "Social" (1 out of 6) could be stimulated by video-based content clues (see Table 2).



Two versions of Level 3 LT were created in accord with the number of intact Level 3 student groups. The order and format of testlets received by Group 3A and 3B students were counter-balanced to exclude corresponding intervening effects. Table 2 summarized the features of length, speed, number of hearings, number of items answerable from a video input, answer time, sequences of passage order and format for the 3A and 3B versions. The final form of the Level 3 LT is given in Appendix along with the YouTube video source links.

Table 2. Features of Level 3 LT testlets and versions.

Testlet / Passage	Length	Speed	Number of Hearings	Items answerable from a video input, if present	Answer time, per item	Order and format of testlets in versions	
						3A	3B
NoWater	02:54	slow to medium	2	3 out of 6	27 sec	1V	3A
Social	03:07	slow to medium	2	1 out of 6	27 sec	2A	4V
BPlan	04:50	medium to fast	2	0 out of 6	27 sec	3V	1A
HugeCh	03:15	medium to fast	2	0 out of 6	27 sec	4A	2V

Since Cronbach's *alpha* reliability is not an appropriate consistency measure for classroom assessments, its criterion-referenced counterpart – dependability – was used (Fulcher, 2010). With the cut score set at 75%, the dependability index for the Level 3 LT was 0.75, showing substantial consistency.

### 3.3.2. Level 5 LT

Level 5 LT test was created following the same steps with a set of different four videos in accordance with the topics covered in the Level 5 LS course – namely business ethics and media use. Looking at Table 3, we can see that, similar to Level 3 LT, all the Level 5 LT passages were monologic and displayed facial expressions and gestures of a speaker. Two video passages – “Ethics” and “MTask” – were closer to the content end of the “context-content” continuum (content-related clues took up 63% and 29% of video time accordingly) while the remaining two listenings (i.e., “Fraud” and “Zombies”) were purely context videos (i.e., with no content clues).

Table 3. Features of Level 5 LT video passages.

Passage	Main topic	Non-verbal cues	Content clues
Passage <b>Ethics:</b> “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 <b>Fraud:</b> “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 <b>Zombies:</b> “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 <b>MTask:</b> “Multitasking and Switchtasking”	A presentation (monologue) about the ineffectiveness of switchtasking and the value of managing one’s time properly.	Entire body: face, gestures	29% of the time: pictures, text, schemes

Each passage was characterized by three to five minutes of length, one hearing, and medium to fast speech rate. Paralleling Level 3 LT, all the passages were followed by three-option multiple-choice questions of the same three types (i.e., main idea, detail, and inference questions). Testlets “Ethics” and “MTask” contained items that could be fully or in part answerable from video content clues – five out of five and two out of six in number respectively. The number of questions in a testlet was five to six, adding up to the total of 22. The final blueprint of the Level 5 LT can be found in Appendix along with the corresponding YouTube video source links. With the cut score set at 75%, the dependability index for the Level 5 LT was 0.74, showing substantial consistency. Since there were four available intact classes of Level 5 students (unlike only two in Level 3), the resulting number of the Level 5 test was four. The four versions had different format and order combinations, as reflected in Table 4.

Table 4. Features of Level 5 LT testlets and versions.

Passage	Length	Speed	Number of Hearings	Items answerable from a video input	Answer time, per item	Order and format of testlets in versions			
						5A	5B	5C	5D
Ethics	04:31	medium	1	5 out of 5	25 sec	1A	4A	3V	2V
Fraud	04:52	medium to fast	1	0 out of 6	25 sec	2V	3V	4A	1A
Zombies	04:13	medium	1	0 out of 5	25 sec	3A	2A	1V	4V
MTask	03:05	medium to fast	1	2 out of 6	25 sec	4V	1V	2A	3A

### 3.4. Data collection

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 the time. The total number of administrations was six – two for Level 3LT and four for Level 5 LT. Before each administration, students were given test papers and scantrons. Then they were instructed to put their last and first names only on test papers leaving personal information sections on scantrons blank.

The administration took place in familiar classrooms with a teacher as a sole supervisor. The tests were given in a paper-and-pencil, closed book, timed format. The students were allowed to spend about two minutes reading the questions before listening to each passage. While listening, the students were free to take notes, if needed. After having listened to a passage, there was about two minutes for test-takers to choose their answers. The students were asked to mark their answers both on test papers by circling right options and on their scantrons by filling out chosen bubbles.

The role of the teacher during administrations was limited to launching pre-recorded computer video files. For each version of the tests, there was a video master file that automatically ran directions and listening passages, as well as controlled the allocation of time for preparation, answer, and revision. The directions were recorded by the researcher and then included in the master video files.

It was ensured that every test-taker could comfortably see and hear the content of the test. The master videos were played on classroom computers and projected to wide white screens situated at the front of a classroom. The audio was amplified and delivered via classroom speakers. Before each administration, the master video files were downloaded on the desktops of classroom computers and then tested to ensure the absence of lags and skips. All the teachers were given oral and written instructions on where to find and how to launch the master files.

As the researcher worked with unnamed scantrons after the administrations, the scores were considered secondary data void of any trace of personal information from the students, which precluded approval from the Institutional Review Board. Still, the permission from the IEP to use the students' scores for research purposes was required and successfully received.

### *3.5. Data analysis*

#### *3.5.1. Order effects*

Before running the analyses for the research questions, it was necessary to control for possible effects of order in which passages and formats appeared on the tests. It was achieved by comparing test-takers' performance on different versions of the tests in SPSS software (Version 22).

Raw scores for 3A and 3B versions of Level 3 LT were compared using the Independent Samples *t*-test procedure as the data obtained from both versions met

the required assumptions (including independence, normality, and equal variance). The results of the test showed no statistically significant difference in the scores for 3A version ( $M = 13.43$ ;  $SD = 2.77$ ) and 3B version ( $M = 14.67$ ;  $SD = 3.04$ );  $t(14) = -0.87$ ,  $p = 0.39$ . Henceforth, it was assumed that the two versions were unaffected by the order of formats and passages, and, thus, were equivalent in terms of difficulty.

To compare raw scores among the four versions of Level 5 LT, a one-way ANOVA was run. It was ensured that all the assumptions required for ANOVA were met including independence of observations, normality of score distributions for each cell, and homogeneity of variance. No significant difference between mean scores was found,  $F(3, 40) = 1.18$ ;  $p = 0.33$ , which showed that order of passages and formats did not affect performance. In turn, it confirmed the equivalence of the four versions.

### 3.5.2. Effects of videos on testlet difficulty

Unlike the examination of order effects on the basis of raw scores, the analyses for the actual research questions in the present study compared item difficulty indices, or ID-values, which were calculated for every single item of each version of the tests in MS Excel. These data were then transferred to IBM SPSS software to run statistical tests according to the specifications of the research questions.

To answer the first research question (i.e., Does delivery format affect listening achievement testlet difficulty for lower-level students?), a two-way 2x4 ANOVA was run with test difficulty (operationalized as ID-values calculated for each item) set as the dependent variable and format and testlet being the two fixed factors. Format had two values – audio-only and video-enhanced. The four values of testlet were “NoWater,” “Social,” “BPlan,” and “HugeCh”. Before running the test, all the necessary assumptions were met and the outlier check was done.

The analysis for RQ 2 (i.e., Does delivery format affect listening achievement testlet difficulty for higher-level students?) was similar to RQ 1 in many respects. Three analogous variables were made use of – test difficulty (dependent), format (independent), and testlet (independent). The operationalization of test difficulty entailed the ID-values calculated for each item of the Level 5 LT versions (5A, 5B, 5C, and 5D). Testlet was operationalized with four values – “Ethics,” “Fraud,” “Zombies,” and “MTask”. The assumptions were checked and a two-way ANOVA was run to estimate the interaction effect between format and testlet and the main effect of format. From there, *post hoc* analysis was performed. All the statistical procedures were run using IBM SPSS software (version 22).

## 4. Results

### 4.1. Difficulty of format for lower-level students

Before running a two-way ANOVA test on the Level 3 LT dataset, it was required to check a number of assumptions. First, the number of observations in each cell, levels of measurement of the variables and the independence of observations attested

to have met the requirements for the test. Second, the analysis of boxplots yielded the conclusion of no significant outliers. Third, the normality of data distributions for each combination of the variables was examined on the basis of  $z$  scores for skewness and kurtosis as well as the Shapiro-Wilk test. Table 5 shows these values for data distributions in each cell as well as descriptive statistics for item difficulties averaged by testlet.

Table 5. Descriptive statistics and distribution parameters for Level 3 LT testlets.

Testlet	NoWater		Social		BPlan		HugeCh	
Format	Audio	Video	Audio	Video	Audio	Video	Audio	Video
k	6	6	6	6	6	6	6	6
M	0.65	0.59	0.55	0.69	0.56	0.57	0.53	0.56
SD	0.19	0.14	0.23	0.19	0.34	0.20	0.30	0.19
$Z_{skew}$	-0.87	-0.54	-1.63	0.73	-0.70	-0.06	0.06	-0.49
$Z_{kurtosis}$	0.53	-1.37	0.78	0.46	0.66	-0.14	-0.97	-1.03
SW $p$	0.80	0.04*	0.07	0.86	0.79	0.96	0.50	0.24

Note: \* = significant at the  $\alpha$  level of 0.05; SW = Shapiro-Wilk test; Each testlets was taken by N = 16 students.

None of the  $z$  scores were significant (greater than the critical value of 1.96). The Shapiro-Wilk test for the “NoWater” data was slightly significant ( $p = 0.04 < 0.05$ ) whereas the others did not approach significance. However, skewness and kurtosis values for this passage did not display serious deviations from normality. Taken all the evidence together, it was assumed that the assumption of normality was not violated. Finally, Levene’s test indicated equal variances across all the eight distributions,  $F(7, 40) = 0.65$ ;  $p = 0.72$ .

Table 6. Results of the Two-way ANOVA for the Level 3 LT data.

Source	$df$	$F$	$p$	$R^2$	Effect size ( $\eta^2$ )
Corrected Model	7	0.35	0.93	0.06	0.06
Intercept	1	311.59	0.00		0.89
Format	1	0.25	0.62		0.01
Testlet	3	0.35	0.79		0.03
Format * Testlet	3	0.38	0.77		0.03
Error	40				
Total	48				
Corrected Total	47				

Note: The object of the ANOVA analysis was item difficulties ( $k = 24$  for each format, totaling to 48 SPSS data entries), not test-takers’ scores

#### 4.2. Difficulty of format for higher-level students

In the process of checking assumption of the Level 5 LT data distributions, three outliers were detected using boxplots. They were the items with ID-values of 0.00 and 0.27 in the audio-only “Fraud” passage and the ID-value of 0.33 in the audio-only Zombie passage. These values were deleted from the dataset. As the very same items were used in corresponding parallel passages under the video-enhanced condition in other versions of Level 5 LT, their ID-values also had to be removed from that

analysis. This was done to avoid unbalance between the numbers of items under both conditions. Next, the normality check was run, the results of which are shown in Table 7 along with descriptive statistics for each cell. According to the  $z$  scores and Shapiro-Wilk test results, the data in each cell did not significantly deviated from normality. Lastly, Levene's test result was non-significant,  $F(7, 74) = 1.00, p = 0.43 > 0.05$ , which suggested equality of variances across groups. In sum, it was concluded that all the assumptions were met.

Table 7. Descriptive statistics and normality parameters for the Level 5 LT testlets.

Testlet	Ethics		Fraud		Zombies		MTask	
Format	Audio	Video	Audio	Video	Audio	Video	Audio	Video
k	10	10	10	10	9	9	12	12
M	0.55	0.76	0.78	0.63	0.78	0.69	0.74	0.76
SD	0.15	0.23	0.18	0.20	0.11	0.22	0.19	0.18
$Z_{skew}$	0.03	-1.23	-0.57	-0.002	-0.80	0.70	-0.24	-0.77
$Z_{kurtosis}$	-1.29	0.42	-0.39	-1.25	0.08	-0.92	-0.67	-0.83
SW $p$	0.25	0.30	0.59	0.40	0.62	0.23	0.63	0.25

Note: SW = Shapiro-Wilk test; Each testlets was taken by N = 44 students.

The two-way ANOVA (see Table 8) 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 (see Table 9) revealed that there was a statistically significant effect of format only for "Ethics",  $F(1; 74) = 6.37, p = 0.01 < 0.05$ . 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 ( $M = 0.76, SD = 0.23$ ).

Table 8. Results of the two-way ANOVA for the Level 5 LT data.

Source	$df$	$F$	$p$	$R^2$	Effect size ( $\eta^2$ )
Corrected Model	7	2.03	0.06	0.16	0.16
Intercept	1	1198.78	< 0.01		0.94
Format	1	0.005	0.94		< 0.01
Testlet	3	1.06	0.37		0.04
Format * Testlet	3	3.67	0.02*		0.13
Error	74				
Total	82				
Corrected Total	81				

Note: The object of the ANOVA analysis was item difficulties ( $k = 22$ ; each testlet occurred 2 times under the audio-only format and 2 times under the video-enhanced format, totaling the number of data entries to  $22 \times 4 = 88$  before removing outliers), not test-takers' scores;

\* = significant at the  $\alpha$  level of 0.05;

Table 9. Post hoc (multiple comparisons) analysis for the Level 5 LT testlets.

Testlet	Mean difference (Video – Audio)	<i>df</i>	<i>F</i>	<i>p</i> <sup>a</sup>	Effect size ( <i>Eta</i> <sup>2</sup> )
Ethics	0.21	(1, 74)	6.37	0.01*	0.08
Fraud	-0.15	(1, 74)	3.33	0.07	0.04
Zombies	-0.10	(1, 74)	1.20	0.28	< 0.01
Mtask	0.03	(1, 74)	0.12	0.73	0.02

Note: \*. statistically significant at the *alpha* level of 0.05; *a*. Adjustment for multiple comparisons: Bonferroni

## 5. Discussion and conclusions

The results of the statistical tests indicated that, for either proficiency level, the presence of videos did not affect the collective difficulty of testlets. In other words, test-takers' cumulative performance on all the audio-only testlets was not different from that on all video-enhanced testlets for each proficiency category. However, it is unreasonable to draw conclusions from this finding since different testlets were enhanced with videos containing varying amounts of content clues. The impacts of different videos may have been contrasting and, thus, cancelled out in the collective estimation of the video effect. Therefore, it is more meaningful to look at how the difficulty of every single testlet was affected by the presence of video and how this effect may be linked to the amount of content clues in videos.

### 5.1. Effect of video-mediated content clues

The difficulties of seven out of the eight testlets used in this study were not affected by the presence of videos. All the four videos used in the lower-level test contained content clues in the range of 0 to 30%. None of them made any difference for difficulty when included in testlets. A similar conclusion can be made in regard to the three videos in the higher-level test (i.e., "Fraud," "Zombies," and "MTask"). The videos used for these testlets contained from 0 to 29% content clues and did not make a difference for the testlets' difficulties.

The seven videos bearing no effect for testlet difficulty contained content-unrelated, situational information for more than a half of the video length, content clues being in the minority or not present. Thus, it is concluded that using videos that contain mostly context information exerted no effect on testlet difficulty for any proficiency level. This finding is in accord with other studies (Batty, 2015; Cubilo & Winke, 2013; Gruba, 1993; Suvorov, 2015a) that found no interaction between delivery mode and performance on a listening test, and, in particular, between delivery mode and proficiency level (Batty, 2015) or video type (Suvorov, 2015b).

Unfortunately, only one video in the study had the prevalence of content-related information. It was the lecture on business ethics ("Ethics") as part of the higher-level test. The video employed content clues in the form of organized text – bulleted lists of major points in the lecture – 63% of the time. The results showed that the presence of this video facilitated test-takers' performance on the "Ethics" testlet. In other words,

the video reduced the difficulty of the testlet. This echoes with Gruba's (2004) finding that written text as well as other content-related scenes in a video aid in L2 learners' listening comprehension and Suvorov's (2015b) finding that test-takers perceive content videos to be more helpful for their comprehension than context videos.

The conclusions above may show a trend in regard to the relationship between video-mediated content clues and testlet difficulty. In general, the presence of content clues did not display any detrimental effects on testlet difficulty for any proficiency level. On the contrary, it seems that the more content information there is in videos, the more probable a facilitative effect on test difficulty is, given that comprehension questions latch onto videos. More specifically, the decrease in difficulty may be predicated on the prevalence of content over context information. Still, this inference may be applicable only to higher-level L2 learners as this study failed to include videos dominated by content clues into the investigation of lower-level test difficulty.

The number of comprehension items that can be answered with the help of content-related video information is another factor to consider. "Ethics" testlet had comprehension items that all were answerable from the video input. In contrast, video-dependent items in the other testlets were much less numerous or absent. Thus, it seems that the effect of videos on difficulty may depend not only on the amount of content clues, but also on the degree to which testlet items are tied in to these clues. It may be argued that a large amount of content clues in a video coupled with the preponderance of video-dependent items work together to lower testlet difficulty.

### *5.2. Construct implications*

The finding that content-related clues may reduce L2 listening test difficulty may trigger the discussion about their place in the listening construct. At the core of such a discussion lays the fact that some authentic listening settings do include content-related visuals in one form or another (e.g., visual aids during lectures). In light of this, it would be unfair to deprive listeners of information they would normally encounter in a corresponding authentic situation. Another unfair consequence of not including content-related visuals may be the increase of listening message difficulty. To avoid these biases, it seems reasonable to define L2 listening as a visual-inclusive skill for every visual-enriched TLU domain of interest (Suvorov, 2015a). It would allow to have separate listening constructs that would reflect the characteristics of corresponding TLU domains, particularly in regard to types visuals used in each domain.

Another subject of debate is whether non-verbal visual cues (e.g., facial expression, gestures) and other context visual clues (e.g., setting) should be a part of the listening construct. The findings of this study add little to bridging this gap of knowledge. However, since the use of videos with the majority of context clues did not make much difference for L2 listeners, it may be harmless to include context clues into the construct. This idea finds support from empirical evidence suggesting potential benefits of including non-verbal cues on L2 listening comprehension (e.g., Sueyoshi &



Hardison, 2005). In a similar vein, non-verbal cues may constitute desirable accompaniments to better represent certain TLU domains and increase authenticity.

### *5.3. Testing implications*

The results of this study may bear several implications in regard to the use of videos in ESL/EFL classroom tests.

One suggestion would be for ESL teachers to not refrain from using context videos in achievement tests. Even though this study did not support a facilitating influence of context videos on listening comprehension, it did not undermine it either. In addition, context videos did not show bias toward level of test-taker's proficiency. This finding may urge English teachers to keep using context 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 towards having video accompaniments as parts of listening tests. This, in turn, may positively affect test-takers' motivation and reduce anxiety.

Another suggestion is to avoid removing video from originally video-enhanced passages. Such practice may be common for IEP teachers in different countries due to possible technological or practical challenges associated with using videos in classroom tests. In case the majority of information in the video is related to the actual content of the listening passage, the exclusion of such a video from a testlet should be taken carefully. Turning a passage that was originally enhanced by a content-related video into the audio-only format may increase the difficulty of the testlet for test-takers. As well, it will surely reduce the authenticity of the listening.

### *5.4. Limitations and future directions*

One of the biggest limitations of this study was the imbalance between the numbers of video types. There was only one video that displayed content clues more than half of the time. This limited the comparison of the effect of content-enriched videos on testlet difficulty to the higher-level proficiency group solely. Future studies could overcome this limitation by making the number of videos with the preponderance of content clues comparable to the number of mostly context videos within a proficiency category. Such studies are needed to confirm that the prevalence of content-related information can be a catalyst of the facilitative effect videos have on listening comprehension.

Investigating interactions between quantified amounts of video-mediated content clues and test-takers' performance on individual items can also present an interesting avenue for further investigation. More sophisticated statistical procedures based on item response theory, such as Rasch analysis used by Batty (2015), could bring additional insights into the degree to which individual items of certain types (i.e.,

main idea, detail, inference) are dependent on the amount of content clues in video passages.

It should be mentioned that this study used subjective human judgment to specify how answerable questions were from a content-related video input. An item was deemed answerable if the correct response for this item was either clearly seen in or vaguely implied by the input, as judged by the researcher. This method may need refinement or even reform, which calls for new research ideas in this area. It may be worth developing a rubric for estimating the degree of relationship between individual items and visual clues and employing a second judge to ensure inter-rater reliability.

Finally, it should be mentioned that the generalizability of the study is restricted to this sample because the participants were not randomly selected and assigned to the stratified groups. This problem is the most difficult to address due to the rare accessibility of true randomization in applied linguistics contexts. However, if future studies are able to produce similar results, we may be able to accrue a body of evidence that would be more easily generalizable to the target population.

## 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. <http://doi.org/b4s9>
- Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). TOEFL 2000 listening framework: A working paper (TOEFL Monograph Series Report No. 19). Retrieved from <https://www.ets.org/Media/Research/pdf/RM-00-07.pdf>
- Buck, G. (2001). *Assessing listening*. Cambridge, UK: Cambridge University Press.
- Chan, C., Wang, L., & Xu, L. (2014). A study of video effect of English listening comprehension. *Studies in Literature and Language*, 8, 53-58. <http://doi.org/b4tb>
- Copyright Law of the United States and Related Laws Contained in title 17 of the United States Code. (2011). Retrieved from <http://www.copyright.gov/title17/circ92.pdf>
- 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. <http://doi.org/b4tc>
- Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, 19, 133–167. <http://doi.org/b9kst8>
- Gruba, P. (1993). A comparison study of audio and video in language testing. *JALT Journal*, 15, 85–88.
- Gruba, P. (2004). Understanding digitized second language videotext. *Computer Assisted Language Learning*, 17, 51-82. <http://doi.org/d8jx2b>
- Kintsch, W. (1998). *Comprehension*. Cambridge: Cambridge University Press.
- Lado, R. (1961). *Language Testing: The construction and use of foreign language tests*. London: Longman.

- 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. Retrieved from [https://www.idosi.org/mejsr/mejsr13\(9\)13/11.pdf](https://www.idosi.org/mejsr/mejsr13(9)13/11.pdf)
- Londe, Z. C. (2009). The effects of video media in English as a second language listening comprehension test. *Issues in Applied Linguistics*, 17, 41-50. Retrieved from <http://escholarship.org/uc/item/0c080191#page-1>
- Mendelsohn, D. J. (1994). *Learning to listen: A strategy-based approach for the second language learner*. San Diego: Dominie Press.
- Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, 24, 517-537. <http://doi.org/dk58d9>
- Shin, D. (1998). Using videotaped lectures for testing academic listening proficiency. *International Journal of Listening*, 12, 57-80. <http://doi.org/b4td>
- 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. <http://doi.org/fv5v5k>
- 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. (2015a). 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*, 21, 1-21. <http://doi.org/b4tf>
- Suvorov, R. (2015b). Interacting with visuals in L2 listening tests: An eye-tracking study. In V. Berry (Ed.), *ARAGs research reports online (Report #AR-A/2015/1)*. Retrieved from [http://www.britishcouncil.org/sites/default/files/interacting\\_with\\_visuals\\_in\\_l2\\_listening\\_tests\\_suvorov.pdf](http://www.britishcouncil.org/sites/default/files/interacting_with_visuals_in_l2_listening_tests_suvorov.pdf)
- Underwood, M. (1989). *Teaching listening*. New York: Longman.
- Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning & Technology*, 11, 67-86. Retrieved from <http://llt.msu.edu/vol11num1/pdf/wagner.pdf>
- Wagner, E. (2010a). Test-takers' interaction with an L2 video listening test. *System*, 38, 280-291. <http://doi.org/d6s5kd>
- Wagner, E. (2010b). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing*, 27, 493-513. <http://doi.org/cjg7cg>
- 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. <http://doi.org/b4tg>
- Witkin, B. R., & Trochim, W. W. K. (1997). Toward a synthesis of listening constructs: A concept map analysis. *International Journal of Listening*, 11, 69-87. <http://doi.org/fvmrjx>
- Wolvin, A., & Coakley, C. G. (1996). *Listening* (5th ed.). Dubuque, IA: Brown & Benchmark.

**Appendix A. Level 3 Listening Achievement Test**

## Listening 1.1: A Country with No Water

**You will hear the passage twice. For questions 1 through 6, fill in the letter of the best answer on your answer sheet. You may take notes while you listen. You now have about 2 minutes to read the questions before you start listening.**

**Source:**

**[https://www.ted.com/talks/fahad\\_al\\_attiya\\_a\\_country\\_with\\_no\\_water](https://www.ted.com/talks/fahad_al_attiya_a_country_with_no_water)**

1. The passage mainly talks about how the Qataris \_\_\_\_\_.
  - A. changed their meals because of water
  - B. lived without fresh water in 1940
  - C. developed without access to water
2. There were no beautiful cities in Qatar because people were without \_\_\_\_\_.
  - A. abilities
  - B. knowledge
  - C. resources
3. We can infer that the Qataris will \_\_\_\_\_ in future.
  - A. drink less water
  - B. continue to have enough water
  - C. need better technology
4. Qatar will probably stop growing without \_\_\_\_\_.
  - A. rivers
  - B. energy
  - C. salt
5. According to the passage, in the 1940s the Qataris were mostly \_\_\_\_\_.
  - A. fishermen
  - B. teachers
  - C. oil workers
6. According to the passage, the Qataris \_\_\_\_\_ than other nations.
  - A. use more water
  - B. live longer
  - C. have more money

**Listening 1.2: A Social Entrepreneur**

**You will hear the passage twice. For questions 7 through 12, fill in the letter of the best answer on your answer sheet. You may take notes while you listen.**

**Source: <https://www.youtube.com/watch?v=N8LVa9pb-n8>**

7. The passage mainly talks about how to \_\_\_\_\_.
  - A. win the Nobel Peace prize
  - B. start a business in Bangladesh
  - C. use business to help people
8. We can infer that the speaker's belief in doing good things comes from \_\_\_\_\_.
  - A. Mohummad Yunus
  - B. good parents
  - C. university teachers

9. A social entrepreneur is a person who solves \_\_\_\_\_.
  - A. social problems with business skills
  - B. business problems with social skills
  - C. women's problems around the world
10. The problem that Mohammad Yunus saw was that \_\_\_\_\_.
  - A. world economy was improving
  - B. many people still stayed poor
  - C. people had to pay their loans back
11. Yunus' idea of micro lending sounded strange because it \_\_\_\_\_.
  - A. took place in Bangladesh
  - B. was for poor people
  - C. started working in 30 years
12. Yunus' micro lending was probably \_\_\_\_\_ successful than American commercial loans.
  - A. more
  - B. less
  - C. equally

### Listening 1.3: Writing a Business Plan

You will hear the passage twice. For questions 13 through 18, fill in the letter of the best answer on your answer sheet. You may take notes while you listen.

Source: <https://www.youtube.com/watch?v=RoytcBiOpDI>

13. The talk is mainly about \_\_\_\_\_.
  - A. what a good business plan should have
  - B. how long a business plan should be
  - C. where to find women for your business
14. According to the passage, we can best learn by \_\_\_\_\_.
  - A. teaching
  - B. doing
  - C. listening
15. Without a business plan, it will probably be impossible to \_\_\_\_\_ a business.
  - A. start up
  - B. get money for
  - C. have an idea for
16. In writing their *business ideas* students need more success in \_\_\_\_\_.
  - A. deciding what to write about
  - B. using more words
  - C. describing the idea clearly
17. We can infer that *Customer Description* (Phase 2) should be more \_\_\_\_\_.
  - A. general
  - B. basic
  - C. specific
18. The right way to think through business is to \_\_\_\_\_.
  - A. go fast from the beginning to the end
  - B. keep returning to the starting point
  - C. follow the guidelines in business books

**Listening 1.4: A Huge Challenge**

You will hear the passage twice. For questions 19 through 24, fill in the letter of the best answer on your answer sheet. You may take notes while you listen.

Sources:

[https://www.youtube.com/watch?v=PYYiY7\\_yGh8](https://www.youtube.com/watch?v=PYYiY7_yGh8)

<https://www.youtube.com/watch?v=pf7nK8G3B3s>

1. The listening is mainly about \_\_\_\_\_.
  - A. the problem of making clean water more available
  - B. how much fresh water is there on Earth
  - C. the cost of making fresh water from salt water
2. According to the listening, we have the smallest amount of \_\_\_\_\_ water on Earth.
  - A. ocean
  - B. salt
  - C. fresh
3. We can infer that the right of having clean water is \_\_\_\_\_ people.
  - A. unimportant for some
  - B. mostly for educated
  - C. difficult to deliver to
4. According to the listening, one of the challenges for people is getting \_\_\_\_\_.
  - A. clean water to places
  - B. higher education
  - C. interested in finding water
5. We can infer that making clean water will be more \_\_\_\_\_ in future.
  - A. expensive
  - B. achievable
  - C. traditional
6. Traveling to the moon was similar to making clean water because it \_\_\_\_\_.
  - A. was cheap and interesting
  - B. took a lot of work
  - C. had to do with finding water

**THIS IS THE END OF THE TEST.**

## **Appendix B. Level 5 Listening Achievement Test**

### **Listening 1.1: Economics and Zombies**

**You will hear the passage once. For questions 1 through 5, fill in the letter of the best answer on your answer sheet. You may take notes while you listen. You now have about 2 minutes to read the questions before you start listening.**

**Source: <https://www.youtube.com/watch?v=HOwRd9kaJhk>**

1. The main goal of the presentation is to talk about the relationship between\_\_\_\_\_.
  - A. economics and ethics
  - B. advertisements and ethics
  - C. movies and economics
2. The speaker used the example of zombies and spirits to \_\_\_\_\_.
  - A. show a common trend in the American movie industry
  - B. illustrate what happens if economics and ethics are split
  - C. explain how zombies are different from economics
3. We can infer that the role of ethics is to \_\_\_\_\_ economics.
  - A. praise
  - B. blame
  - C. guide
4. When separate, ethics and economics \_\_\_\_\_.
  - A. function better
  - B. balance each other
  - C. work poorly
5. In the past, economics and ethics were \_\_\_\_\_.
  - A. separate
  - B. together
  - C. ridiculous

### **Listening 1.2: Multitasking and Switchtasking**

**You will hear the passage once. For questions 6 through 11, fill in the letter of the best answer on your answer sheet. You may take notes while you listen.**

**Source: <https://www.youtube.com/watch?v=787Q41eflYo>**

6. The passage mainly talks about \_\_\_\_\_.
  - A. the effects of multitasking
  - B. the time multitasking takes
  - C. how to multitask effectively
7. According to the speaker, switchtasking is \_\_\_\_\_.
  - A. doing several tasks simultaneously
  - B. going back and forth between tasks
  - C. typing and talking at the same time

8. To accomplish three tasks as soon as possible, we probably should \_\_\_\_\_.
  - A. work on them at the same time
  - B. do them one after another
  - C. take a couple of short breaks
9. Multitasking and switchtasking can lead to \_\_\_\_\_.
  - A. making a smaller number of mistakes
  - B. reducing levels of stress dramatically
  - C. poor understanding of instructions
10. We can infer that people in the past were multitasking \_\_\_\_\_ often.
  - A. less
  - B. more
  - C. equally
11. The main goal of this presentation is to show people the \_\_\_\_\_ of multitasking.
  - A. examples
  - B. ineffectiveness
  - C. popularity

### **Listening 1.3: Business Ethics**

**You will hear the passage once. For questions 12 through 16, fill in the letter of the best answer on your answer sheet. You may take notes while you listen.**

**Source:** <https://www.youtube.com/watch?v=vmVu66Fpd9U>

12. The main purpose of the lecture is to introduce business ethics \_\_\_\_\_.
  - A. on the whole
  - B. in corporations
  - C. for individuals
13. We can infer that business ethics comes from \_\_\_\_\_.
  - A. personal relationships in a workplace
  - B. general ethical principles
  - C. business technology and science
14. The main concern of ethics in general is \_\_\_\_\_ good behavior.
  - A. religious rules for
  - B. standards of
  - C. feelings about
15. We can infer that ethics is \_\_\_\_\_ law.
  - A. fully covered by
  - B. less effective than
  - C. broader than
16. One level of business ethics discussed in the lecture is \_\_\_\_\_.
  - A. community
  - B. society
  - C. institution



### **Listening 1.4: Fraud Triangle**

**You will hear the passage once. For questions 17 through 22, fill in the letter of the best answer on your answer sheet. You may take notes while you listen.**

**Source:** <https://www.youtube.com/watch?v=SPR6UssAYus>

17. The talk is mainly about \_\_\_\_\_ in an organization.
  - A. how weak internal control causes problems
  - B. different reasons why fraud takes place
  - C. why high integrity individuals commit fraud
18. For fraud to happen, there probably should be both financial pressure and \_\_\_\_\_.
  - A. a past history of dishonesty
  - B. high integrity
  - C. weak internal control
19. An example of rationalization the speaker gives is closest in meaning to \_\_\_\_\_.
  - A. "My boss pays me less than I deserve"
  - B. "I need money to buy my boss a cup of coffee"
  - C. "My boss has too much money"
20. High integrity individuals \_\_\_\_\_ to commit fraud.
  - A. are typically unable
  - B. need stronger reasons
  - C. are most expected
21. New employees can help prevent fraud by \_\_\_\_\_ in the company.
  - A. paying less attention to their co-workers
  - B. looking for the elements of fraud triangle
  - C. finding the golden key to deal with fraud
22. The speakers discuss the elements of fraud triangle to help employees \_\_\_\_\_.
  - A. see fraud risk and prevent it
  - B. prepare for a quiz on fraud structure
  - C. earn more money from companies

**THIS IS THE END OF THE TEST.**

---

#### **Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the Journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).