

THE POSSIBILITY OF SUPPLEMENTING THE LISTENING COMPREHENSION COMPONENTS OF FOREIGN LANGUAGE EXAMINATIONS WITH AUDIO-VISUAL TASKS

doi.org/10.61425/wplp.2020.15sp.70.69

Ármin Kövér

Language Pedagogy PhD Programme
Eötvös Loránd University, Budapest,
armin.kover@gmail.com

Abstract: Due to the rapid nature of technological development, the array of language learning opportunities and tools has widened. To keep up with the changing trends in language learning, language testing also has to constantly improve its practices. In real-life language use the consumption of audio-visual material is growing. Consequently, language testing professionals might also want to consider including audio-visual materials in the listening comprehension component of foreign language examinations. The topic of using audio-visual material in listening tests has already been researched before (e.g., Kellerman, 1990; Ockey, 2007; Sueyoshi & Hardison, 2005), however, these studies arrived at contradictory results. For this reason, the present study investigates the audio-only and audio-visual text comprehension of 140 language learners at four language proficiency levels (i.e., A2-C1) in two languages (i.e., English and German). The participants of the study filled in both a paper-based and a computer-based test including both audio-only and audio-visual tasks. In addition, at the end of the tests, they were asked to fill in a questionnaire regarding the usefulness of the visual input in solving the tasks. The results suggest that the inclusion of audio-visual tasks into the listening comprehension component of language tests is possible, but it requires a sound methodological rationale.

Keywords: listening comprehension, audio-visual comprehension, language testing, computer-based language testing

1 Introduction

With the emergence of the new technological innovations of the 21st century, the field of education has experienced substantial changes. Being exposed to audio-visual material has become part of people's everyday lives, and using audio-visual material has also become a prevalent practice in the field of education (Woolfitt, 2015). This practice has also become widespread in foreign-language education (Suvorov, 2009), and presumably it has especially influenced the ways in which listening comprehension is taught and practiced. Since language testing aims to assess a skill in an artificial situation which successfully emulates the intended real-life situation, these aforementioned changes in the practice of teaching listening comprehension should necessitate the revision of the ways listening comprehension is assessed in foreign language tests.

Furthermore, the methods currently used in testing listening comprehension are deeply rooted in the approach of Communicative Language Teaching (and testing), which emerged in the 1980s. Comparing the present context to that of the 1980s, today's life is

substantially different regarding the way information is spread and used, or consumed. While the use of audio-only input in the form of telephone conversations and radio broadcasts was dominant at the emergence of the Communicative Language Teaching approach, it has not kept pace with the dramatic expansion of audio-visual sources of information. Today's examples would include a whole range from YouTube videos, online conference presentations, Skype calls, and online video courses, webinars, to name but a few. This change of the social setting should also initiate the re-evaluation of the methods currently used in testing listening comprehension.

Even though the using of audio-visual materials in the testing of listening comprehension has already been researched, researchers did not arrive at clear results. Bejar et al. (2000) found that content-related and context-related visuals can have a positive effect on the comprehension of the aural input. However, the findings of Ockey (2007) and Londe (2009) seem to suggest that the presence of visual input does not have any effect on the candidates' performance. The existence of contradictory findings and the fact that the field of testing listening comprehension with audio-visual materials is under-researched, especially in the Hungarian context, indicate that the issue needs further investigation. In order to start remedying this deficiency, the aim of the present paper is to analyse whether including audio-visual tasks in the listening comprehension component of language examinations is necessary and desirable.

2 Theoretical background

2.1 Listening comprehension

Listening comprehension has an essential role in people's everyday lives both in the form of face-to-face, telephone, and online conversations and in the form of watching and listening to pre-recorded materials on TV or the Internet. According to Wagner (2014), it is estimated that at least 50% of communication is spent with listening. In addition, the understanding of speech is also one of the main focuses of foreign language education since having good listening comprehension skills facilitates, as a precondition, the learners' successful communication in the foreign language.

In the past century of investigating listening comprehension, the term *listening comprehension* has been defined in various different ways. One of the more commonly used definitions is Rost's (1990) definition:

Understanding spoken language is essentially an inferential process based on a perception of cues rather than a straightforward matching of sound to meaning. The listener must find relevant links between what is heard (and seen) and those aspects of context that might motivate the speaker to make a particular utterance at a particular time. (p. 33)

Another important definition which must be taken into consideration when discussing language testing is the definition provided by the CEFR (Council of Europe, 2001). According to the CEFR, listening comprehension is defined as a listener receiving and processing "spoken input produced by one or more speakers" (Council of Europe, 2001, p. 65). As part of the process, the listeners' knowledge of the world and knowledge

of schematic structures are activated as well as their ability to decode the message on a phonological, syntactic and word level (Council of Europe, 2001).

These definitions suggest that the process of speech comprehension involves the decoding of utterances. As part of decoding the utterances the acoustic characteristics of sounds, such as their length and loudness, aid the listener in identifying speech signals in the stream of sounds (Marslen-Wilson & Tyler, 1980). In addition, the time required for processing (Brazil, 1983; Chafe, 1980, 1982; Kreckel, 1981) and the identification of phonemic units (Chomsky & Halle, 1968) are also important components of speech comprehension. Nevertheless, real-time listening comprehension is an even more complex process where besides the identification of the physical characteristics of sounds and the derivation of abstract phonemes into their variations, the listeners also have to engage their pragmatic knowledge and their short-term memory in order to understand the meaning of the message (Berg, 1987; Bregman, 1978; Buck, 2001).

When discussing testing listening comprehension, the validity of the test should also be taken into consideration. One of the aspects of the validity of the test is concerned with the relationship between the test takers' performance in the testing situation compared to their performance in a similar related-life situation. This relationship is called *cognitive validity* (Glasser, 1991). As the aim of the test is to assess — through simulated tasks — how a candidate would perform in a real-life situation the test should represent the real-life context as much as possible. Since the everyday life listening habits of people these days heavily feature the input (consumption) of audio-visual materials (Brynjolfsson & McAfee, 2014), the cognitive validity of the currently used listening tests could be questioned and including audio-visual materials into the listening comprehension component of language examinations might enhance the tasks' reflection of the real-life situation.

2.2 Audio-visual comprehension

As a result of constant technological development, consuming audio-visual material has become an essential aspect of every person's life. According to the CEFR, audio-visual comprehension is defined as “the user simultaneously receiv[ing] an auditory and a visual input” (Council of Europe, 2001, p. 71), or “the user watch[ing] TV, video, or a film and us[ing] multimedia, with or without subtitles and voiceovers” (Council of Europe, 2018, p. 54). Based on this definition, the difference between audio-only and audio-visual comprehension is that in the case of audio-visual comprehension the listener has to comprehend both audio and visual input.

Despite the fact that using audio-visual material in foreign language teaching is becoming more and more popular, a reluctance can be observed on the part of the test developers towards including audio-visual materials into language tests. This reluctance is reflected by the CEFR as well since even its revised edition (Council of Europe, 2018) describes audio-visual comprehension only in a single scale. The audio-visual reception scale of CEFR (Council of Europe, 2018) focuses on three main concepts: the ability to understand and follow the main ideas, the ability to comprehend details and implied meaning, and the ability to understand different types of language use. In comparison with the other competencies the CEFR (Council of Europe, 2018) describes, audio-visual reception seems to be underrepresented.

There could be several explanations proposed as to why test developers are hesitant to include audio-visual materials into language tests. Even though some researchers encourage the inclusion of audio-visual tasks into foreign language tests, based on the fact that in real-life communication verbal and non-verbal elements are equally important (Progosh, 1996; Wagner, 2007), their potential to measure something different from listening comprehension is an often-raised criticism against making audio-visual tasks a part of language tests (Buck, 2001). Despite the fact that the possibility of construct-irrelevant variance can be a relevant concern, it must be taken into consideration that audio-visual tasks can probably recreate real-life language use, namely, target language use (Bachman, 1990), more successfully than audio-only tasks.

The influence of using different types of audio-visual tasks on L2 listening performance has already been investigated by several researchers, however, their findings are contradictory (Kellerman, 1990; Ockey, 2007; Raffler-Engel, 1980; Sueyoshi & Hardison, 2005). The visual input which can be used in listening comprehension tasks is usually divided into four categories: context-related images (e.g., a still image depicting two people talking to each other on the street), content-related images (e.g., the photo of a figure or a table accompanying a presentation about it), context-related videos (e.g., a video recording of two people talking to each other in a classroom), and content-related videos (e.g., the video recording of a set of presentation slides) (Suvorov, 2011). The different types of visuals are presumed to have a different influence on the test takers' performance. The most noteworthy studies in this topic are Bejar et al. (2000), Ginther (2002), and Ockey (2007). When investigating the performance of TOEFL test-takers, Bejar et al. (2000) found that the inclusion of context-related images had a positive effect on the candidates' test performance. Similar results were concluded by Ginther (2002) who found that content-related visual inputs positively influenced the test-takers' performance, nevertheless, another finding of her study was that context-related visuals negatively impacted the understanding of short talks, positively influenced the understanding of lectures, and had no notable influence on understanding conversations. In contrast, Ockey's (2007) study asserts that many of the test-takers did not use the visual components of the tasks at all and that the presence of the visual material did not have any effect on the participants' performance. Londe's (2009) study produced similar results where the video recording accompanying a 10-minute lecture seemed to have no effect on participants' performance.

Candidates' attitudes towards the use of audio-visual material in language tests have also been examined in several research studies, and the results of these studies are also contradictory (Dunkel, 1991; MacWilliam, 1986; Ockey, 2007; Sueyoshi & Hardison, 2005; Wagner, 2002). On the one hand, Dunkel (1991), Sueyoshi and Hardison (2005), and Wagner (2002) found that their participants had a preference for audio-visual tasks over audio-only tasks because they felt that the visual material had a positive influence on their test performance. On the other hand, the participants of the studies conducted by MacWilliam (1986) and Ockey (2007) claimed that they found the video material distracting so they did not watch it at all. Despite the inconsistent results, these pieces of research are all designed appropriately. Therefore, one explanation for the contradictory results could be that the participants might not have been able to appropriately judge the efficiency of the methods used in the studies so the participants' claims about their preferences and attitudes regarding the use of audio-visual tasks should be taken with caution (Kirschner & van Merriënboer, 2013).

One of the most frequently mentioned concerns regarding the use of audio-visual tasks in testing is its potential for being a distractor rather than a facilitator. Nonetheless, it must also be taken into account that in the case of audio-visual listening tasks the test-taker is also able to rely on the kinesic behaviour (e.g., body language, facial expressions, gestures and visible stress patterns) of the speaker, which they would also do in their everyday language use (Raffler-Engel, 1980). According to Kellerman (1990) and Raffler-Engel (1980), the kinesic behaviour of the speaker is an essential part of verbal interaction and in instances of potential misunderstanding, the kinesic behaviour of the speaker increases. Furthermore, according to Burgoon (1994), in situations where the information deduced from the kinesic input contradicts the information deduced from the linguistic input listeners are more likely to accept the information deduced from the kinesic input as correct. The data seem to support the inclusion of audio-visual tasks into language tests as they are more capable of capturing all the nuances of the real-life communications. Since real-life language use appears to support the necessity of including audio-visual material into language tests, and the research conducted in the topic has ambiguous results, the aim of the present study was to analyse whether including audio-visual tasks into the listening comprehension component of language examinations is necessary and desirable. To investigate this issue, the following research questions were proposed:

1. Does the performance of the test-takers on the audio-visual-to-audio-only tasks differ from their performance on the audio-visual tasks?
2. Do the participants perceive the inclusion of audio-visual tasks as useful?

3 Research methods

3.1 Data collection

The present study was conducted in three phases from the beginning of September 2017 to the end of August 2018. Most of the data collection was conducted as part of a language examination development project whose aim was to develop a computer-based language examination for four language proficiency levels (i.e., A2, B1, B2, and C1). The project was conducted by a major Hungarian language school.

The aim of the first phase was to develop 16 sets of listening comprehension tasks (i.e., eight English and eight German), one paper-based and one computer-based set for each language proficiency level (i.e., A2, B1, B2, and C1). The paper-based sets of tasks were developed for the purposes of a larger study also conducted within the framework of the language examination development project. In order to be able to examine the potential effect of audio-visual tasks on the performance of the participants and the participants' preferences, the last task of each set was created based on audio-visual material; however, in the case of the paper-based test, the visual input was removed from these tasks during the test-administration. These tasks are called in the present study *audio-visual-to-audio-only* (henceforward *ATAO*) tasks.

In the second data collection phase, two questionnaires were developed in order to be able to assess the participants' opinions about the usefulness of including audio-visual material into the listening test. These questionnaires were developed independently of the language examination development project and they were piloted with the help of four

English learners who were asked to do both versions of the sets of tasks appropriate for their language proficiency levels and then perform a think-aloud protocol on filling-in a questionnaire.

In the third data collection phase, 140 participants completed both the paper-based and the computer-based sets of tasks matching their language proficiency levels. After completing the tests, the participants filled in the questionnaire appropriate for the test version. The data collected was subjected to statistical analyses.

3.2 Participants

The four participants of the second phase were all English learners (i.e., one female and three males, their ages ranging from 21 to 36). They all spoke English at different language proficiency levels. All of these participants had been studying English for a while and they were all preparing for different language examinations. The data provided by them were only used for finalising the questionnaire; their answers were not considered when answering the research questions because the circumstances of the data collection were different from those of the third phase of the study.

The third data collection phase had 140 participants (i.e., 60 males and 80 females between the ages of 12 to 42). The participants came from several different contexts as the data collection was carried out at a major Hungarian university, in several groups of two major Hungarian language schools, in six high schools from three Hungarian counties, and two elementary schools from two Hungarian counties. The large variety of the data collection sites was justified by the intention to avoid potentially skewed results caused by collecting data from one type of language learning context only. The participants' language proficiency levels were indicated by their language teachers based on the course books they used in class. Out of the 140 students, 73 completed the English tests and 67 participants the German tests. For the number of tests completed for each language proficiency level in each language, see Table 1. and Table 2.

Language proficiency level	Number of participants
A2	11
B1	19
B2	26
C1	17

Table 1. The number of participants taking the English tasks in the third phase

Language proficiency level	Number of participants
A2	11
B1	19
B2	24
C1	13

Table 2. The number of participants taking the German tasks in the third phase

3.3 Instruments

The main data collection instruments of the present study were the paper-based and computer-based sets of tasks. Altogether eight audio-visual tasks, eight ATA0 tasks, and 60 audio-only tasks were developed for the third data collection phase. The tasks were organised into eight sets of tasks intended for the paper-based tests (i.e., four in English and four in German), and eight sets of tasks intended for a computer-based tests (i.e., four in English and four in German). Both the English and the German paper-based tests contained three audio-only tasks and one ATA0 task on the A2 language proficiency level, and four audio-only tasks and one ATA0 task on the B1-C1 levels. In the case of the computer-based tests, both the English and the German tests contained three audio-only tasks and one audio-visual task on the A2 level, and four audio-only tasks and one audio-visual task on the B1-C1 levels. The terms *paper-based test* and *computer-based test* indicate the method of test delivery. In the paper-based test the candidates solved the test in a paper-and-pen format; while in the computer-based test they solved the task on computers (in a lab improvised for the pretests).

Another important instrument of the study was the questionnaire used to investigate the participants' attitudes towards the use of audio-visual tasks in the language test. This questionnaire had two different versions, one tailored for the paper-based test and another one tailored for the computer-based test. Both versions of the questionnaire had 18 items investigating four constructs: *disturbing features*, *structure of the test*, *perceived difficulty*, and *necessity of the video*. The necessity of the video construct was the only one which contained different items in the two versions of the questionnaire: in the paper-based test its items referred to the degree to which the participants think some videos could have successfully aided them in solving the tasks; whereas in the computer-based version the items of the necessity of the video construct investigated the degree to which the participants found the video material useful for solving the last task. As the questionnaire was developed for the purpose of a larger project, the only questionnaire construct relevant for the present study is the *necessity of the video*. The relevant questionnaire items are presented in the table below (Table 3.). The participants needed to answer the questions on a five-point Likert scale (i.e., participants could choose from the following options: strongly disagree, disagree, neither agree nor disagree, agree, or strongly agree).

Paper-based version	Computer-based version
<ul style="list-style-type: none"> • Jobb lett volna, ha vannak videók is a szövegekhez. // It could have been better if there had been a video to some of the recordings. 	<ul style="list-style-type: none"> • A videó segítette az utolsó szöveg jobb megértését. // The video helped the better understanding of the last recording.
<ul style="list-style-type: none"> • Egy-egy videó segíthette volna a hangzó szövegek jobb megértését. // A video could have helped the better understanding of some of the audio recordings. 	<ul style="list-style-type: none"> • A videó megkönnyítette az utolsó szöveg gondolatmenetének követését. // The video made it easier to follow the lines of thoughts in the last recording.
<ul style="list-style-type: none"> • Egy-egy videó segíthette volna a hangzó szövegek gondolatmenetének követését. // A video could have helped following the lines of thoughts in some of the recordings. 	<ul style="list-style-type: none"> • A videóban szereplő képi információk segítették az utolsó feladat megválaszolását. // The visual information in the video helped to better answer the questions of the last task.
<ul style="list-style-type: none"> • Egy-egy videó segíthette volna a feladatok megválaszolását. // A video could have helped answer the questions in some of the tasks. 	<ul style="list-style-type: none"> • A videót hasznosnak találtam az utolsó feladat megoldásához. // The video was useful in answering the questions of the last task.

Table 3. The relevant questionnaire items for the paper-based and the computer-based versions of the tests (The original, Hungarian version, and the English translation)

3.4 Data analysis

In order to be able to answer the first research question the results of the participants achieved on the ATA0 and the audio-visual task were entered and compared in a Microsoft Excel spreadsheet. Furthermore, the data collected with the help of the questionnaires were entered in SPSS 22.0 and descriptive statistics were calculated for the *necessity of the video* construct in order to be able to answer the second research question.

4 Results and discussion

In order to answer the first research question, the participants' results on the ATA0 and the audio-visual tasks were compared. This comparison could indicate the degree to which the participants' performance was potentially influenced by the presence or the absence of the visual material. Before this analysis, the Cronbach's alpha values of the tests were calculated and based on these values it was concluded that the English B1 and C1 and German B2 level tests did not have a satisfactory Cronbach's alpha value. These values were below 0.7 based on Kline's (2000) categorisation of Cronbach's alpha values. Therefore, the present analysis focuses only on the comparison of the remaining proficiency levels. For the participants' results on the ATA0 and audio-visual tasks see Tables 4. – 8.

Test version	Candidate No.	Item No.						Σ	%	Test version	Candidate No.	Item No.						Σ	%
		19	20	21	22	23	24					20	21	22	23	24	25		
paper-based	1a	0	1	1	1	1	1	5	83	computer-based	1b	0	1	1	1	1	1	5	83
	2a	1	0	1	1	1	1	5	83		2b	0	1	1	1	1	1	5	83
	3a	0	1	1	1	1	1	5	83		3b	0	1	1	0	1	1	4	67
	4a	1	1	1	1	1	1	6	100		4b	0	0	1	1	1	1	4	67
	5a	1	1	1	0	1	1	5	83		5b	0	1	1	0	1	1	4	67
	6a	0	1	1	1	1	1	5	83		6b	0	1	1	1	1	1	5	83
	7a	1	1	1	1	1	1	6	100		7b	1	0	1	1	1	1	5	83
	8a	0	0	1	1	1	1	4	67		8b	0	0	0	0	0	0	0	0
	9a	1	0	1	0	1	0	3	50		9b	1	1	1	1	1	1	6	100
	10a	0	0	1	1	1	1	4	67		10b	1	1	0	0	1	0	3	50
	11a	0	1	0	1	1	0	3	50		11b	0	1	1	1	1	1	5	83
		Average %						77			Average %						70		

Table 4. Comparison of the Participants' Results on the last tasks in the English A2 paper-based and computer-based tests
Note. Light grey shading indicates higher performance. Dark grey shading indicates average percentages.

Test version	Candidate No.	Item No.					Σ	%	Test version	Candidate No.	Item No.					Σ	%
		25	26	27	28	29					23	24	25	26	27		
paper-based	31a	1	1	1	1	1	5	100	computer-based	31b	0	0	1	1	1	3	60
	32a	1	1	0	0	1	3	60		32b	0	1	1	1	1	4	80
	33a	1	1	1	1	1	5	100		33b	0	1	1	1	1	4	80
	34a	1	1	1	1	1	5	100		34b	1	1	0	1	1	4	80
	35a	1	1	1	1	0	4	80		35b	1	1	1	1	0	4	80
	36a	1	1	1	0	1	4	80		36b	1	0	0	1	0	2	40
	37a	1	1	1	1	1	5	100		37b	0	0	0	0	1	1	20
	38a	1	1	1	1	0	4	80		38b	1	1	1	1	1	5	100
	39a	0	1	1	0	1	3	60		39b	1	1	0	0	1	3	60
	40a	1	1	0	1	1	4	80		40b	1	1	1	1	0	4	80
	41a	1	1	1	1	1	5	100		41b	0	1	0	1	0	2	40
	42a	1	1	1	1	0	4	80		42b	0	1	0	0	1	2	40
	43a	1	1	1	1	0	4	80		43b	0	0	0	1	1	2	40
	44a	1	1	0	1	0	3	60		44b	1	1	0	0	1	3	60
	45a	1	1	1	1	1	5	100		45b	0	0	0	0	1	1	20
	46a	0	0	0	0	0	0	0		46b	0	1	0	1	1	3	60
	47a	0	1	0	0	1	2	40		47b	1	1	0	1	1	4	80
	48a	1	0	0	1	0	2	40		48b	1	1	0	1	0	3	60
	49a	1	1	1	1	1	5	100		49b	0	0	0	0	0	0	0
	50a	1	1	1	1	0	4	80		50b	1	1	0	0	1	3	60
	51a	0	0	1	0	1	2	40		51b	0	1	1	0	1	3	60
	52a	0	0	0	1	1	2	40		52b	0	1	1	0	0	2	40
	53a	0	0	0	0	0	0	0		53b	0	1	0	0	1	2	40
	54a	0	0	0	0	0	0	0		54b	1	1	1	0	0	3	60
	55a	1	1	1	1	1	5	100		55b	1	1	0	1	1	4	80
	56a	1	1	0	0	1	3	60		56b	1	1	1	1	1	5	100
			Average %					68				Average %					58

Table 5. Comparison of the participants' results on the last tasks in the English B2 paper-based and computer-based tests
Note. Light grey shading indicates higher performance. Dark grey shading indicates average percentages.

Test version	Candidate No.	Item No.						Σ	%	Test version	Candidate No.	Item No.					Σ	%
		19	20	21	22	23	24					20	22	23	24	25		
paper-based	74a	1	1	1	0	1	0	4	67	computer-based	74b	1	0	1	0	0	2	40
	75a	0	1	1	0	1	1	4	67		75b	1	1	1	1	1	5	100
	76a	1	1	1	0	1	0	4	67		76b	1	1	1	1	1	5	100
	77a	1	1	1	0	1	1	5	83		77b	1	0	1	1	1	4	80
	78a	0	1	1	1	1	1	5	83		78b	1	1	1	1	1	5	100
	79a	1	1	1	1	1	1	6	100		79b	1	1	1	1	1	5	100
	80a	1	1	1	0	1	1	5	83		80b	1	1	1	1	1	5	100
	81a	0	1	1	0	0	0	2	33		81b	1	1	1	1	1	5	100
	82a	0	1	1	0	1	1	4	67		82b	1	1	1	1	1	5	100
	83a	0	1	1	1	1	1	5	83		83b	1	1	1	1	1	5	100
84a	1	1	1	0	1	1	5	83	84b	1	1	1	1	1	5	100		
		Average %						74			Average %					93		

Table 6. Comparison of the participants' results on the last tasks in the German A2 paper-based and computer-based tests
Note. Light grey shading indicates higher performance. Dark grey shading indicates average percentages.

Test version	Candidate No.	Item No.					Σ	%	Test version	Candidate No.	Item No.					Σ	%
		26	27	28	29	24					25	26	27	28			
paper-based	85a	0	0	1	1	2	50	computer-based	85b	1	0	0	1	0	2	40	
	86a	0	0	1	1	2	50		86b	1	1	1	1	1	5	100	
	87a	0	0	1	0	1	25		87b	1	0	0	1	0	2	40	
	88a	0	0	1	1	2	50		88b	0	0	0	0	0	0	0	
	89a	0	1	1	0	2	50		89b	1	0	1	0	0	2	40	
	90a	0	1	0	1	2	50		90b	1	0	0	0	0	1	20	
	91a	1	1	0	1	3	75		91b	1	1	0	1	1	4	80	
	92a	1	1	0	0	2	50		92b	1	0	0	1	0	2	40	
	93a	1	1	1	1	4	100		93b	1	0	1	0	1	3	60	
	94a	0	1	0	0	1	25		94b	1	1	1	1	0	4	80	
	95a	0	0	1	0	1	25		95b	1	1	1	1	0	4	80	
	96a	1	0	1	0	2	50		96b	1	0	1	0	0	2	40	
	97a	1	0	0	1	2	50		97b	1	0	1	0	1	3	60	
	98a	0	1	0	0	1	25		98b	1	0	0	1	0	2	40	
	99a	0	1	0	0	1	25		99b	1	0	0	0	1	2	40	
100a	0	1	1	0	2	50	100b	1	1	1	0	0	3	60			
101a	0	1	0	0	1	25	101b	1	1	0	1	0	3	60			
102a	0	1	0	0	1	25	102b	1	0	1	1	0	3	60			
103a	1	1	1	1	4	100	103b	1	1	1	0	0	3	60			
		Average %					47			Average %					53		

Table 7. Comparison of the participants' results on the last tasks in the German B1 paper-based and computer-based tests

Note. Light grey shading indicates higher performance. Dark grey shading indicates average percentages.

Test version	Candidate No.	Item No.					Σ	%	Test version	Candidate No.	Item No.						Σ	%
		25	26	27	28	29					27	28	29	30	31	32		
paper-based	128a	1	1	1	0	1	4	80	computer-based	128b	1	1	1	1	1	0	5	83
	129a	0	0	1	1	0	2	40		129b	1	1	1	1	1	1	6	100
	130a	1	0	1	0	0	2	40		130b	1	0	0	1	1	0	3	50
	131a	0	1	1	0	1	3	60		131b	1	1	1	1	1	0	5	83
	132a	1	0	1	1	0	3	60		132b	1	0	1	0	1	0	3	50
	133a	1	1	1	0	1	4	80		133b	0	0	0	0	1	0	1	17
	134a	0	1	0	0	1	2	40		134b	1	1	1	1	1	1	6	100
	135a	1	0	1	0	1	3	60		135b	1	1	1	1	1	1	6	100
	136a	1	0	1	0	1	3	60		136b	1	1	1	0	1	1	5	83
	137a	0	0	1	0	1	2	40		137b	1	1	1	0	1	1	5	83
	138a	1	1	1	0	1	4	80		138b	0	1	1	0	0	1	3	50
	139a	1	0	1	1	1	4	80		139b	1	1	1	0	1	1	5	83
	140a	0	1	1	1	0	3	60		140b	0	0	1	1	1	0	3	50
			Average %					60				Average %						72

Table 8. Comparison of the participants' Results on the last tasks in the German C1 paper-based and computer-based tests
Note. Light grey shading indicates higher performance. Dark grey shading indicates average percentages.

In the analysis the average percentages of the total number of correct answers provided by the participants on the ATA0 and the audio-visual task were compared and the results appear to suggest that the participants of the A2 and B2 English tests had a slightly better performance on the ATA0 task; whereas the participants of the A2, B1, and C1 German tests achieved higher results on the audio-visual task. The most notable difference could be observed in the case of the A2 level German test participants who had a 19% higher rate of correct answers on the audio-visual task compared to the ATA0 task. Regarding the rest of the language proficiency levels, the difference does not seem to be significant so it can be concluded that the participants generally did not have a lower performance on the audio-visual tasks than on the ATA0 tasks.

Nevertheless, it must be taken into consideration that the present results can only be concluded for the present data set and as every participant only completed one ATA0 and one audio-visual task calibrated for their language proficiency level, thus, the results cannot be generalised. The usefulness of the video might only reflect the features of these particular tasks so it cannot be claimed that every task would benefit from the addition of audio-visual material. Furthermore, as the tables show, some of the participants achieved 0% altogether on the audio-visual task while they achieved 100% on the ATA0 task. This might be the outcome of the digital platform not recording their answers, and this potential data loss could lead to the distortion of the results.

Regarding the second research question, the data provided by the participants on the items related to the necessity of the video construct of the questionnaire were subjected to statistical analysis. The results are summarised in Table 9., and they appear to suggest that in the paper-based test versions the participants had a medium to low preference for including audio-visual material in the listening comprehension test. However, these results might be skewed by the data collection procedure, which required every participant to complete the paper-based test first. Therefore, the participants might not have had any idea what the inclusion of audio-visual material could mean in terms of their task-solving processes. This presupposition might also be supported by the relatively high standard deviation values indicating a high variance in the answers.

Proficiency level	Language	Test version	Mean	Std. deviation
A2	English	Paper-based	3.68	0.96
		Computer-based	3.43	1.28
	German	Paper-based	2.89	0.82
		Computer-based	4.02	0.86
B1	English	Paper-based	3.10	1.17
		Computer-based	3.89	1.19
	German	Paper-based	2.97	1.17
		Computer-based	3.95	0.92
B2	English	Paper-based	2.71	1.29
		Computer-based	3.23	1.15
	German	Paper-based	3.52	1.07
		Computer-based	3.52	1.02
C1	English	Paper-based	3.04	0.93
		Computer-based	3.78	1.01
	German	Paper-based	2.69	1.25
		Computer-based	3.12	0.90

Table 9. Questionnaire results: The necessity of the video

Regarding the usefulness of the video in the computer-based test, at most language proficiency levels the mean values are between 3.50 and 4.0, which indicates that the participants found the videos useful in solving the tasks. The results also appear to suggest that the lower language proficiency level participants found the presence of the video material more useful than their higher language proficiency level peers. However, it must be noted that the sample size for each language proficiency level was small so the investigation should be repeated with a larger sample in order to be able to arrive at generalizable results. Nevertheless, it appears that according to the participants of the present study, the inclusion of the video material is not disturbing and it is even useful in some cases.

5 Conclusions

In conclusion, the results suggest that in terms of the test scores the participants appeared to give a similar performance on the ATA0 tasks and the audio-visual tasks. Regarding the second research question, the results appeared to be in line with the findings of Bejar et al. (2000) and Ginther (2002) as the majority of the participants claimed to find the presence of the videos in the audio-visual tasks non-disturbing. Some of the lower proficiency level students appeared to even find the presence of the video content helpful and they claimed that it had a positive influence on their test-taking processes. These results and the fact that including the audio-visual material into the listening comprehension component of language examinations could lead to improved cognitive validity and a more authentic representation of the real-life language use indicate that the present approach to testing listening comprehension should be reconsidered.

As any research endeavour, the present study also has its limitations. First, during the data collection procedures, every participant had to first solve the paper-based set of tasks and only then the computer-based tasks so the participants' assessment of the usefulness of the video might be skewed as some of the participants might not have been able to imagine the ways audio-visual materials could have influenced their task-solving processes. A further limitation related to the data collection could have emerged from the fact that at that time the digital platform used for data collection was still in the development phase. Therefore, the possibility of potential technical difficulties concerning the recording of the participants' answers cannot be excluded. Lastly, as only one ATA0 and one audio-visual task was developed for each language proficiency level and each language, the results of the present study cannot be generalised. For this reason, further research would be needed on the topic using a larger number of tasks and a larger data collection sample.

6 Pedagogical implications

Including audio-visual content into the listening comprehension component of language examinations thus can provide several benefits for the field of language testing. Firstly, the listening component of the tests would be able to represent real-life language use more authentically, which could increase the cognitive validity of the test. Secondly, including audio-visual tasks into language examinations could prompt language test centres to gradually move their tests to a computer-based platform. A well-designed computer-based platform would be able to gather more information related to the test-solving processes of the candidates, thus, providing important metadata and insights for all stakeholders involved.

Besides its effects on the language testing community, the inclusion of audio-visual materials into foreign language examinations would also greatly influence the field of foreign language education in Hungary. On the one hand, it would produce a positive washback effect as it could increase the use of video material in foreign language classes. In a scenario where students had to practise for a language examination featuring audio-visual tasks, their tutors would be encouraged to incorporate more video-based tasks and audio-visual material into their lessons. On the other hand, such an innovation could also have a negative washback effect on foreign language education. For example, some teachers may feel the need to incorporate videos in their class material for the sake of being “fashionable” and if such an inclusion is not based on sound methodological reasons, it would not have the desired effect on the learning process. It has to be ensured that the video material used in the classes is designed to enhance the learning process of the students and it is based on justifiable methodological reasons.

In conclusion, including audio-visual materials into the listening comprehension component of language tests can provide many benefits both for the foreign language testing and foreign language teaching community including the learners as well. Besides potentially enhancing the cognitive validity of the tests, computer-based language tests could also provide metadata about the test-taking strategies of the candidates. In addition, it could also help language education keep up with the technological advancements of the 21st century and the constantly changing context of language use.

Proofread for the use of English by: Amy N. Soto and Gergely A. Dávid, Department of English Language Pedagogy, Eötvös Loránd University, Budapest.

References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL 2000 listening framework: A working paper*. Educational Testing Service.
- Berg, T. (1987). The case against accommodation: Evidence from German speech error data. *Journal of Memory and Language*, 26(3), 277–299.
- Brazil, D. (1983). Intonation and discourse: Some principles and procedures. *Text*, 3(1), 39–70. <https://doi.org/10.1515/text.1.1983.3.1.39>
- Bregman, A. (1978). The formation of auditory streams. In J. Requin (Ed.), *Attention and performance* (pp. 63–75). Earlbaum.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age*. W.W. Norton & Company.
- Buck, G. (2001). *Assessing listening*. Cambridge University Press.
- Burgoon, J. (1994). Non-verbal signals. In M. Knapp & G. Miller (Eds.), *Handbook of interpersonal communication* (pp. 344–393). Routledge.
- Chafe, W. (1980). The deployment of consciousness in the production of a narrative. In W. Chafe (Ed.), *The pear stories* (pp. 9–50). Ablex.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper and Row.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.

- Council of Europe (2018). *The CEFR companion volume with new descriptors*.
<https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Dunkel, P. (1991). Computerized testing of nonparticipatory L2 listening comprehension proficiency: An ESL prototype development effort. *Modern Language Journal*, 75(1), 64–73. <https://doi.org/10.1111/j.1540-4781.1991.tb01084.x>
- Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, 19(2), 133–167.
- Glasser, R. (1991). Expertise and assessment. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 17–30). Prentice Hall.
- Kellerman, S. (1990). Lip service: The contribution of the visual modality to speech perception and its relevance to the teaching and testing of foreign language listening comprehension. *Applied Linguistics*, 11 (3), 272–280.
- Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist*, 48(3), 169–183.
<https://doi.org/10.1080/00461520.2013.804395>
- Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). Routledge.
- Kreckel, M. (1981). Tone units as message blocks in natural discourse: Segmentation of face-to-face interaction by naïve native speakers. *Journal of Pragmatics*, 5(5), 459–476.
- Londe, Z. C. (2009). The effects of video media in English as a second language listening comprehension tests. *Issues in Applied Linguistics*, 17(1), 41–50.
<https://doi.org/10.5070/14171005108>
- MacWilliam, I. (1986). Video and language comprehension. *ELT Journal*, 40(2), 131–135.
- Marslen-Wilson, W., & Tyler, L. (1980). The temporal structure of spoken language comprehension. *Cognition*, 8(1), 1–72. [https://doi.org/10.1016/0010-0277\(80\)90015-3](https://doi.org/10.1016/0010-0277(80)90015-3)
- Ockey, G. (2007). Construct implication of including still image or video in computer-based listening tests. *Language Testing*, 24(4), 517–537.
<https://doi.org/10.1177/0265532207080771>
- Progosh, D. (1996). Using video for listening assessment: Opinions of test-takers. *TESL Canada Journal*, 14(1), 34–44. <https://doi.org/10.18806/tesl.v14i1.676>
- Raffler-Engel, W. (1980). Kinesics and paralinguistics: A neglected factor in second language research and teaching. *Canadian Modern Language Review*, 36(2), 225–237. <https://doi.org/10.3138/cmlr.36.2.225>
- Rost, M. (1990). *Listening in language learning*. Longman.
- Sueyoshi, A., & Hardison, D. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(4), 661–699.
- Suvorov, R. (2009). Context visuals in L2 listening tests: The effects of photographs and video vs. audio-only format. In C. A. Chapelle, H. G. Jun, & I. Katz (Eds.), *Developing and evaluating language learning materials* (pp. 53–68). Iowa State University.
- Suvorov, R. (2011). The effects of context visuals on L2 listening comprehension. *University of Cambridge ESOL Examinations Research Notes*, 45, 2–8.
- Wagner, E. (2002). Video listening tests: A pilot study. *Working Papers in TESOL & Applied Linguistics, Teachers College, Columbia University*, 2(1), 1–39.
- Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning & Technology*, 11(1), 67–86.
- Wagner, E. (2014). Using Unscripted Spoken Texts in the Teaching of Second Language Listening. *TESOL Journal*, 5(2), 288–311. <https://doi.org/10.1002/tesj.120>
- Woolfitt, Z. (2015). *The effective use of video in higher education*. Retrieved August 22, 2019, from <https://www.inholland.nl/media/10230/the-effective-use-of-video-in-higher-education-woolfitt-october-2015.pdf>