ISSN 2201-2982
2023/2

IELTS Research Reports Online Series

Investigating the effect of interactive videos on test-takers' performance on the listening section of IELTS

Ruslan Suvorov and Zhi Li

IELTS™
BRITISH COUNCIL idp Cambridge Assessment English

Investigating the effect of interactive videos on test-takers' performance on the listening section of IELTS

This study compared test-takers' performance on the IELTS Listening test delivered in the traditional paper-based audio-only format vs. a computer-based format that uses interactive videos with embedded test items. The study also explored test-takers' perceptions and preferences of the two formats.

Funding

This research was funded by the IELTS Partners: British Council, Cambridge Assessment English and IDP: IELTS Australia. Grant awarded 2022.

Publishing details

Published by the IELTS Partners: British Council, Cambridge Assessment English and IDP: IELTS Australia © 2023.

How to cite this report

Suvorov, R., & Li, Z. (2023). Investigating the effect of interactive videos on test-takers' performance on the listening section of IELTS. *IELTS Research Reports Online Series, No. 2/23*. British Council, Cambridge Assessment English and IDP: IELTS Australia. Available at https://www.ielts.org/teaching-and-research/research-reports

Topic tags: IELTS Test - Listening; Technology; Test comparisons

Introduction

This study by Suvorov and Li was conducted with support from the IELTS Partners (British Council, IDP: IELTS Australia, and Cambridge University Press & Assessment), as part of the IELTS joint-funded research program. Research funded by the British Council and IDP: IELTS Australia under this program complement those conducted or commissioned by Cambridge University Press & Assessment, and together inform the ongoing validation and improvement of IELTS.

A significant body of research has been produced since the joint-funded research program started in 1995, with over 200 empirical studies receiving grant funding. After undergoing a process of peer review and revision, many of the studies have been published in academic journals, in several IELTS-focused volumes in the *Studies in Language Testing* series (http://www.cambridgeenglish.org/silt), and in the *IELTS Research Reports* series. Since 2012, to facilitate timely access, the research reports have been published on the IELTS website immediately after completing the peer review and revision process.

The inclusion of video input in listening tasks has long been an area of debate. While being able to process visual or non-verbal information is often seen as embedded within listening comprehension, language assessments typically avoid the inclusion of visuals. An important aspect of this is the interaction demanded of the candidate with the input. In the context of IELTS Listening, which can be defined as a 'while-listening' performance test, the candidate is required to listen concurrent to answering items, and there is the risk of cognitive overload and a potential impact on performance.

One way to ameliorate issues of split attentional resources is to utilise interactive video input where the items are embedded in the input. The use of interactive videos has been relatively extensive in wider educational contexts, yet within language testing these studies have been notably rare. With this gap in mind, Suvorov and Li seek to explore candidate performance in both an online, interactive video input format IELTS Listening test and the standard, audio-only, paper-based test, and how the test-takers' response processes interacted with item characteristics such as item difficulty and item type. This study also investigates test-taker perceptions concerning the interactive video input version.

Through a counter-balanced, mixed-methods approach this study raises interesting insights into the potential impacts interactive videos could have on test-taker performance. The findings indicate that a statistically-significant difference in the challenge of items existed between formats; item-type and length were key factors in response times; and candidate preferences tended towards the more familiar audio-only format.

While a note of caution is warranted (materials used were originally written for an audio-only format), this study provides evidence of the pedagogic value interactive video input could bring to assessments. It raises the timely issue that audio-only listening may represent a very different construct from multimodal listening. It also highlights the potentially distinct constructs targeted by while-listening and post-listening assessments, and the influences these test design choices may have on candidate behaviours, perceptions and ultimately their performance.

Nick Glasson
Senior Research Manager
Cambridge University Press & Assessment

Investigating the effect of interactive videos on test-takers' performance on the listening section of IELTS

Abstract

In the area of second language (L2) listening assessment, videos as input materials have been studied for several decades. However, with the exception of He's (2022) study, there appears to be no L2 assessment research on interactive videos (i.e., videos with tasks embedded in the video timeline), despite the growing utility of such videos in online instructional settings and reported benefits for learners (Ketsman et al., 2018).

To explore the potential of using interactive videos for assessing L2 listening comprehension, we converted two original IELTS Listening tests into interactive video listening tests with embedded test items. A counterbalanced design (2 tests by 2 formats) was employed to administer the original audio-only paper-based listening tests and the interactive video-based listening tests to
65 L2 English-speaking participants at two Canadian universities.

In this mixed-methods study, participants' performance on the two test formats was compared quantitatively using ANOVA and many-facet Rasch model (MFRM) analysis. Item-level response time on the interactive video listening tests was analysed using a mixed-effects model. Participants' perceptions and preferences of both formats were elicited via focus group interviews and an end-of-the-experiment questionnaire.

The results revealed a statistically significant difference in total scores between the two IELTS Listening tests, with the interactive video version found to be more difficult than the audio-only version. Bias analyses in MFRM identified the items with differential item difficulties between the formats, with most of such items being more difficult in the interactive video format than in the audio-only format. Meanwhile, participants' response time was found to be related to item difficulty, item length, and item types. Lastly, the qualitative data analyses demonstrated that while more than half of the participants preferred the audio-only format, many participants also valued specific features that were unique to the interactive videos.

Implications are discussed in relation to the design of interactive video-based listening tests and the construct(s) of L2 listening comprehension.

Topic tags: IELTS Test - Listening; Technology; Test comparisons

.

Authors' biodata

Ruslan Suvorov

Ruslan Suvorov is an associate professor in applied linguistics at the University of Western Ontario, Canada, where he teaches courses in second language assessment and computer-assisted language learning (CALL). His research interests lie at the intersection of language assessment, CALL, and instructional technology and design, with a focus on second language listening and eye tracking. Ruslan has given numerous conference presentations and workshops and published in peer-reviewed journals (e.g., *CALICO Journal, International Journal of Listening, Language Testing*), edited volumes, conference proceedings, encyclopaedias, and research reports. He is a co-author of *Blended Language Program Evaluation* (Palgrave Macmillan, 2016).

Zhi Li

Zhi Li is an assistant professor in the Department of Linguistics at the University of Saskatchewan (UoS), Canada. Before joining UoS, he worked as a Language Assessment Specialist at Paragon Testing Enterprises, Canada, and as a sessional instructor in the Department of Adult Learning at the University of the Fraser Valley, Canada. Zhi Li holds a doctoral degree in applied linguistics and technology from Iowa State University, USA. His research interests include language assessment, technology-supported language teaching and learning, corpus linguistics, and computational linguistics. His research papers have been published in *System, CALICO Journal*, and *Language Learning & Technology*.

Acknowledgements

We would like to thank and acknowledge our outstanding graduate research assistants, Shanshan He (PhD student, University of Western Ontario) and Olga Kriukova (PhD student, University of Saskatchewan) who have made a significant contribution to every stage of this project, including participant recruitment, data collection, and data analysis.

Table of contents

List of figures

List of tables

1 Introduction

The ability to understand and process visual information is considered indispensable for second language (L2) listening comprehension, as evidenced from various definitions of this skill (Field, 2019). As a result, visuals are widely used in L2 listening instruction as they are believed to increase the authenticity of listening tasks. Meanwhile, paradoxically, the integration of visuals in L2 listening assessment contexts remains a controversial point among researchers (e.g., Pusey & Lenz, 2014; Suvorov, 2015; Wagner, 2013), with most major listening tests, including IELTS Listening, relying on the audio-only delivery format (Kang et al., 2019).

The avoidance of videos in existing L2 listening tests can be partly attributed to the multiplicity of factors affecting how, and to what extent, test-takers watch videos. Studies investigating L2 test-takers' viewing behaviour have demonstrated that there are significant differences among test-takers in terms of the extent to which they watch videos during L2 listening assessment, with some test-takers attending to the video most of the time and others avoiding any eye contact with the visual input (Batty, 2021; Suvorov, 2015; Wagner, 2007). In addition, note-taking has been found to be a major factor interfering with the viewing behaviour of test-takers who have to split their attention between watching the videos and taking notes that they rely on when answering test items (Suvorov, 2013). As a result, during while-listening performance (WLP) tests such as IELTS Listening that require reading and answering test items while listening to the stimuli (Aryadoust, 2012), L2 test-takers must attend to watching videos, taking notes, and deciding when to answer each test item, which oftentimes leads to cognitive overload and arguably has a detrimental impact on test performance, especially for less proficient test-takers (Suvorov & He, 2022).

Seeking a solution to the cognitive overload and split attention problem, this study explores the extent to which interactive videos can obviate the need for test-takers to attend to watching the videos, taking notes, and answering test items simultaneously while completing the listening section of IELTS. Unlike regular videos, interactive videos contain test items embedded at specific intervals directly into the playback. When a test-taker gets to the point in the video where a specific test item is located, the video is automatically paused, and the test item appears on top of the video. Thus, interactive videos can be deemed a variant of post-listening performance (PLP) tests, in which listening to the stimulus and answering a test item are consecutive rather than concurrent processes (Aryadoust, 2012).

We hypothesise that an interactive video format can:

1. encourage test-takers to watch the video more consistently (and thus avoid switching their attention between video-watching and note-taking)
2. minimise the need to take notes due to short intervals between test items in the video (in order to focus on listening comprehension)
3. provide a direct indication to test-takers when to answer a specific test item (so that there is no need for test-takers to expend their cognitive resources on deciding when to give an answer).

2 Literature review

2.1 Theoretical framework

This study is informed by two main theoretical frameworks: cognitive load theory (Sweller et al., 2011) and cognitive theory of multimedia learning (Mayer, 2020). The cognitive load theory provides instructional design principles that can be used to analyse and explain the relationship between test-takers' cognitive load and working memory resources, as well as to reduce the cognitive load during the completion of a listening test. According to one of these principles, separating different sources of information (e.g., a video prompt and a test item) spatially or temporally may lead to the split-attention effect that imposes an additional cognitive load on test-takers and has a detrimental effect on their test performance.

Similar to the cognitive load theory, Mayer's (2020) cognitive theory of multimedia learning recognises that test-takers have limited capacity for processing information from different sources. According to the limited capacity assumption underlying Mayer's (2020) theory, language assessment tasks must be designed in a way that limits the amount of information presented to test-takers simultaneously and eliminates the need for them to split attention between different sources of information. Thus, the cognitive load theory and the cognitive theory of multimedia learning suggest that in order to prevent test-takers from experiencing the split-attention effect and reduce their cognitive load, all essential components (e.g., video prompts and test items) should be integrated in the design of L2 listening assessment tasks. We leverage both theories to provide a conceptual justification for using interactive videos in an L2 listening test that force test-takers to attend to only one component at a time: watching the video or responding to test items embedded in the paused video.

2.2 Interactive videos: Definition, uses, and benefits

Also known as hypervideos (Bakla & Mehdiyev, 2022), interactive videos are a type of videos that contain embedded elements such as text, images, questions, and hyperlinks designed to increase the viewers' interactivity with the video content. In the context of this study, interactivity refers to "the involvement of users in the exchange of information with computers and the degree to which this happens" (*Cambridge Business English Dictionary*, 2023). It should also be noted that unlike linear videos with questions displayed outside of the video player window, interactive videos present questions inside the video player window at certain intervals, with the video playback being automatically paused each time a question appears. Multiple applications such as Edpuzzle (https://edpuzzle.com/) and H5P (https://h5p.org/) can be leveraged to design and create interactive videos. H5P, for instance, allows for creating activities or quiz questions that can be built into an interactive video, thus minimising learners' need to move between a screen and test paper. More importantly, locations of activities or questions can be optimised to help learners concentrate on the tasks at hand and reduce their cognitive load (Casañ Núñez, 2017).

A review of the literature suggests that interactive videos have been utilised fairly extensively in the teaching of natural sciences such as biology (Haagsman et al., 2020) and physics (Ketsman et al., 2018), as well as other disciplines such as business leadership (Campbell et al., 2019) and podiatry and marketing (Rice et al., 2019). Recent studies on video-based learning in higher education have demonstrated a number of benefits of embedding activities or questions in instructional videos.

Such benefits include improved learner engagement and better course performance (Kleftodimos & Evangelidis, 2016; Sablić et al., 2020). Haagsman et al. (2020), for example, studied the effect of pop-up questions in interactive videos on learning in a molecular biology course. The findings indicated that students benefited from the interactions with pop-up questions embedded in videos as they adjusted their viewing behaviour accordingly and improved learner engagement. Both Rice et al.'s (2019) study and Tweissi's (2016) doctoral dissertation revealed that university students performed significantly better on the questions embedded in interactive videos vs. the questions presented after linear videos. The embedded questions "helped participants to raise self-efficacy and gain more confidence, enhance existing knowledge with new information, rehearse memory, and achieve better learning outcome" (Tweissi, 2016, p. 4). In van der Meij and Böckmann's (2021) study, students performed significantly better and engaged significantly more with the open-ended questions (without feedback) that were embedded in video-recorded lectures than on the questions that were not embedded. Similarly, Delen et al. (2014) found that compared to common/traditional videos, videos with enhanced interactivity were more effective for engaging students in self-regulated learning and improving their learning performance. While introducing embedded questions in an interactive video-based listening comprehension test may bring similar benefits to test-takers, language testers should be cognisant of the fact that pausing the videos frequently to display the embedded questions may be interruptive for the test-taking process (Bakla & Mehdiyev, 2022). This is because these planned pauses and embedded questions may demand the types of meta-cognitive strategies that differ from the ones required in paper-based listening comprehension tests. Such differences in meta-cognitive strategies can inevitably have a differential effect on the test-takers' listening test performance (Xu, 2017).

In addition to research examining the effectiveness of interactive videos for students' learning, some studies have explored students' perceived benefits of interactive videos. For example, the participants in Ketsman et al.'s (2018) study expressed stronger preference for interactive videos because of enhanced engagement with the video content and immediate feedback on embedded questions. Interestingly, the students perceived interactive videos to be beneficial for deeper learning and better retention of new information, even though the study did not reveal any statistically significant differences between their performance on questions embedded in interactive videos vs. questions presented after videos. Similarly, Rice et al. (2019) reported that students overwhelmingly preferred videos with embedded questions and found them to be more beneficial than non-interrupted videos with questions grouped together at the end. As perceived by students, the benefits of interactive videos with embedded questions included immediate feedback on their responses to the questions, increased attention and engagement with the content, and increased understanding and retention of new information. Preference for interactive videos was also reported in Campbell et al. (2019), whose participants favoured a non-linear fashion of watching videos, especially videos that were longer than five minutes.

The use of interactive videos has been also gaining some momentum in language learning contexts (Taslibeyaz, 2020), including flipped learning contexts (Bakla & Mehdiyev, 2022; Zou & Xie, 2019). Taslibeyaz (2020), for instance, compared the effectiveness of scenario-based vs. non-scenario-based interactive videos on the achievement scores of learners in an English grammar course. She found that the use of scenario-based interactive videos resulted in statistically significantly higher achievement scores and higher levels of students' self-regulation operationalised as the frequency of using additional videos and answering practice questions. In Zou and Xie's (2019) study, the positive impact of the flipped learning model on the development of upper-intermediate L2 learners' writing skills was partly attributed to the presence of interactive videos that enabled learners to focus on understanding one piece of key information at a time by pausing the video and answering the embedded questions.

2.3 Videos in second language listening assessment

While questions embedded in interactive videos have been utilised in a variety of online courses as effective learning tools or formative assessment instruments, they are rarely used in summative or high-stakes language testing contexts. To our knowledge, the only study that used interactive videos in an L2 listening assessment context is that of He (2022). In her study, He (2022) investigated 30 L2 English speakers' performance on a video-based listening test in two formats: one that uses traditional linear videos and another that uses interactive videos. The results of her comparative study revealed that while the participants performed better on the listening test with traditional videos, they expressed a strong preference for interactive videos, mostly because they found this video type to be useful for maintaining their focus and reducing distractions during the listening test.

Even though videos have been used to assess listening for several decades, there is still a lot of debate about how to operationalise and measure the construct of video-based listening, also known as multimodal listening (Gruba, 2020). Previous studies on multimodal listening tests have suggested that test-takers may benefit from the use of video stimuli in such tests as they have access to more resources to facilitate comprehension (Brett, 1997; Wagner, 2010). Meanwhile, researchers have also raised concerns about the impact of videos on the test construct (Gruba & Suvorov, 2020; Ockey, 2007) as the degree of test-takers' engagement with visuals tends to vary greatly. In other words, test-takers may not watch a video input if the competition of attention resources for different modalities becomes too intense (Wagner, 2010). This may be largely related to how videos and items are integrated in a listening test as many of the researched video-based listening tests are paper-based, with few exceptions (e.g., Brett, 1997; Lesnov, 2022; Suvorov, 2015). Unlike the listening tests that are based on traditional linear videos, listening tests utilising interactive videos may prompt test-takers to use different viewing behaviours and response strategies. Therefore, it is important to monitor test-takers' response processes in order to understand the possible impact of the test formats on L2 listening comprehension (Li et al., 2017).

The current study aims to provide insights into this issue by investigating how test-takers' performance on the traditional paper-based audio-only version of the IELTS Listening test compares to their performance on a computer-based version of the IELTS Listening test that uses interactive videos. It also aims to explore test-takers' perceptions and preferences of interactive videos, as well as their response processes during the completion of the IELTS Listening tests mediated by such videos.

3 Research questions

This study was guided by the following three research questions.

1. **To what extent does the test-takers' performance on a computer-based interactive video IELTS Listening test differ from their performance on a paper-based audio-only IELTS Listening test?**
2. **How are the test-takers' response processes (operationalised as response time on individual items embedded in interactive videos) associated with item characteristics (such as item difficulty, item length, and item type) in the computer-based interactive video IELTS Listening test?**
3. **What are the test-takers' perceptions and preferences regarding the effectiveness of interactive videos for measuring their L2 listening ability?**

4 Methodology

4.1 Research design

This study used a within-participants design. The dependent variables were test scores and item-level response time. The independent variables were the test format (i.e., paper-based audio-only test vs. computer-based interactive video test), item difficulty (facility) index, item discrimination index, item type (i.e., multiple-choice, multiple-answer, drag-and-drop, and fill-in-the-blank), item length in terms of word count, testlet (i.e., a specific video with associated ten items), and test version (i.e., Test A and Test B, see Section 4.3, Materials below).

4.2 Participants

The participants were 65 L2 speakers of English recruited among international students at the University of Saskatchewan (n = 33) and the University of Western Ontario (n = 32) in Canada. They were 53 female, 11 male, and one participant who self-identified as non-binary. The average age of the participants was 27 years (SD = 4.80, min = 20, max = 41). Almost half of all participants were native speakers of Chinese (n = 32), followed by Persian (n = 8), Spanish (n = 6), Portuguese (n = 3), Bengali (n = 2), and Tamil (n = 2). Thirteen additional languages were spoken by one participant each, with Tamil and Malay being reported as two L1s by one of the participants. Most of the participants were graduate students (i.e., 37 Master-level students and 18 doctoral students), with 10 participants being undergraduate students. They majored in a variety of academic disciplines, including education (n = 21), social science (n = 16), engineering (n = 9), medical science (n = 7), computer science (n = 4), and natural science (n = 3).
The participants' experience with English as a medium of instruction varied from half a year to 20 years (M = 8.78, SD = 6.12). Two-thirds of the participants (n = 44) reported that prior to this study they had rarely, if ever, encountered interactive videos, with only five participants claiming to have used interactive videos frequently or very frequently.

With regard to their previous test-taking experience, the majority of the participants had taken either IELTS Academic (n = 43) or IELTS General (n = 2). The remaining tests included TOEFL iBT (n = 11), CELPIP (n = 3), Duolingo English Test (n = 2), and TOEIC (n = 1). Based on their self-reported test scores, 43 participants were at the C1 level, 15 participants were at B2, three participants at C2, one participant at B1, with three participants not reporting any test scores.

In addition to the 65 participants who completed the main study, six more participants were recruited for the pilot study to test the data collection instruments and procedure.

4.3 Materials

4.3.1 Listening tests

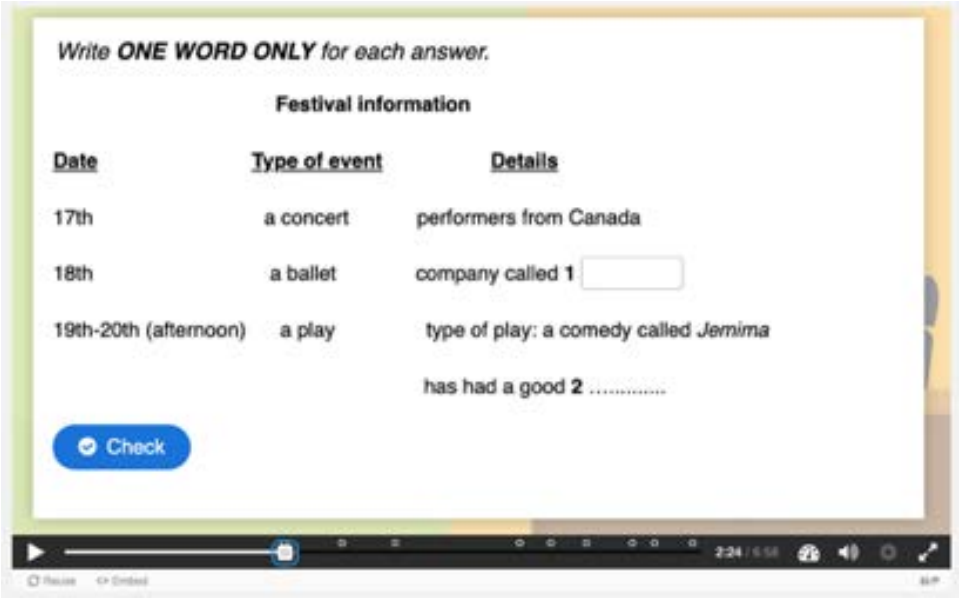
This study utilised two retired IELTS Academic Listening tests taken from the IELTS Academic 15 Practice Test Series published by Cambridge University Press (2020). Using the content from these paper-based listening tests, we created a computer-based version of these tests with interactive videos. In particular, we used the original audio stimuli to create animated videos in Vyond, which is a web application for creating video animation (https://www.vyond.com/). Because the recording of traditional videos would have presented us with a plethora of design choices related to video content, videography, and video composition (i.e., the arrangement of visual elements within a frame), we opted for animated videos. One of the key benefits of animated videos is that they can be used to convey complex ideas or processes in a simple way. Animations are also easier and cheaper to create as they do not require actors, access to special locations, or expensive recording equipment. Finally, when creating the animated videos

for the listening test, we did not need to reproduce the audio and could simply use the original audio tracks as a voiceover. Using the original audio in the animated videos allowed us to control all aspects of the auditory input such as accent, pauses, rate of speech, and intonation that were identical in the paper-based audio-only listening tests and the computer-based listening tests with interactive videos.

In total, eight animated videos were produced in Vyond. We subsequently used the animated videos to design interactive videos in HTML5 Package (H5P), which is an open-source application for multimodal content creation (https://h5p.org/). In this study, interactive videos were defined as animated video stimuli in which test items were embedded at certain intervals directly into the video playback (typically, right after the information that was relevant to the question asked). When a test-taker watching the video (Figure 1) reached a specific test item, the video would automatically pause and the test item would appear in a pop-up window over the video (Figure 2). After the test-taker answered the item, the video playback would resume automatically and continue until the next embedded test item. As shown in Figure 1, the location of each item is marked with a small hollow circle in the progress bar underneath the video. Once an item is answered, the circle turns into a solid dot to indicate the completion of the item. This feature allows test-takers to monitor their listening process and anticipate upcoming items.

Figure 1: *An interactive video (Listening Test B, Part 1)*

Figure 2: *A test item embedded in an interactive video (Listening Test B, Part 1)*



When designing the computer-based version of the listening test with interactive videos, we aimed to preserve the original format of the test as much as possible. Specifically, we used the same instructions (with slight modifications), provided the same amount of time for item previewing, and allowed paper-based note-taking. Similar to the original audio-only IELTS Listening test, each computer-based interactive video listening test comprised four parts and 40 items, with each part (testlet) containing one animated video stimulus and 10 items. As each item was scored dichotomously (i.e., 1 for the correct answer and 0 for the incorrect answer), the maximum score that a participant could obtain on a particular listening test was 40.

While we also tried our best to keep the original format of the items (i.e., item types), some of them had to be re-purposed for computer delivery. In particular, the original matching items that required test-takers to choose the answer from the list of options and write the correct letter next to the question were converted into drag-and-drop matching items. Similarly, the map labelling items were converted into the drag-and-drop matching items. Finally, note and table completion items were presented as fill-in-the-blank items. Larger notes and tables that did not fit on a single video frame were split into several parts. Overall, four item types were used in the computer-based interactive video listening tests: fill-in-the-blank, multiple-choice multiple-answer, multiple-choice (single-answer), and drag-and-drop items.

We also had to make a number of design choices related to the interactive video format. First, when designing interactive videos, we purposefully avoided any visuals that would provide explicit cues or answers to the test items. We did this to ensure that the test items could be answered primarily by using auditory rather than visual information. Second, we embedded the questions shortly (typically within a few seconds) after the relevant information was presented in the video, which was usually the end of an idea unit defined as "a message segment consisting of a topic and comment that is separated from contiguous units syntactically and/or intonationally" (Ellis & Barkhuizen, 2005, p. 154). We made this design choice for two reasons: (a) to reduce the role of memory and cognitive load in our participants' performance; and (b) to support the same types of response processes that test-takers typically demonstrate during the original IELTS Listening test (i.e., answering the questions as soon as the relevant information for the answers is heard in the input). Third, due to the settings of the H5P application, we were not able to hide item-level feedback that showed whether a particular response was correct. As this feedback was presented automatically after each item, the test-takers knew immediately whether they had answered the item correctly or not. Fourth, all the playback controls were disabled so that the participants could not pause, rewind, or fast forward the video once they started watching it. Finally, while it was not possible to enforce time for answering individual test items, we did ask the participants to complete each listening test within 40 minutes. This time limit was the same as the duration of the original IELTS Listening test that lasted approximately 30 minutes, with additional ten minutes provided for transferring the answers from the test booklet to the answer sheet.

After designing the computer-based version of the two listening tests with interactive videos in H5P, we embedded both tests in Moodle (v. 3.11), which is a learning management system (LMS). To familiarise the participants with the computer-based interactive-video format, we created a short Practice Test (comprising one interactive video and five items) and made it available in Moodle as well. We chose Moodle to deliver the interactive video listening tests for two main reasons. First, this LMS allowed for participant authentication that could be used to identify which test scores belonged to which participant. Second, Moodle could be connected to GrassBlade Learning Record Store (LRS), which is a cloud-based repository of learning data. Using GrassBlade LRS enabled us to automatically capture not only individual responses on test items, but also the amount of time spent by each participant on those items.

In total, this study used two comparable versions of the IELTS Listening test (i.e., Listening Tests A and B), each of which was available in two formats: a paper-based audio-only format (i.e., the original version used in IELTS) and a computer-based interactive video format. As a result, there were four listening test forms: audio-only Listening Test A (paper-based), interactive video Listening Test A (computer-based), audio-only Listening Test B (paper-based), and interactive video Listening Test B (computer-based). Table 1 provides a brief overview of the listening tests used in the study. The audio transcripts, test items, and answer keys are provided in Appendix A.

Table 1: *Overview of the listening tests used in the study*

	Audio-only format (paper-based)	Interactive video format (computer-based)
Practice Test **Topic: Employment agency: Possible jobs**	none	Stimulus: 1 interactive video Tasks: 5 items Duration: 5 minutes
Listening Test A **Part 1 topic: Bankside recruitment agency** **Part 2 topic: Matthews Island holidays** **Part 3 topic: Personality traits** **Part 4 topic: The eucalyptus tree in Australia**	Stimulus: 4 audio tracks Tasks: 40 items Duration: 30 minutes for listening and 10 minutes for transferring the answers from the booklet to the answer sheet	Stimulus: 4 interactive videos Tasks: 40 items Duration: 40 minutes for listening and answering the items embedded in the videos
Listening Test B **Part 1 topic: Festival information** **Part 2 topic: Minster Park** **Part 3 topic: Charles Dickens** **Part 4 topic: Agricultural program in Mozambique**	Stimulus: 4 audio tracks Tasks: 40 items Duration: 30 minutes for listening and 10 minutes for transferring the answers from the booklet to the answer sheet	Stimulus: 4 interactive videos Tasks: 40 items Duration: 40 minutes for listening and answering the items embedded in the videos

4.3.2 Questionnaire

We created a questionnaire to gather background information about the participants, including their age, gender, L1, academic discipline, and L2 proficiency level. In addition, the questionnaire was used to elicit the participants' perceptions and preferences regarding the use of the two listening test formats (i.e., listening tests with interactive videos and audio-only listening tests). The questionnaire comprised 14 open-ended questions and 12 statements that participants were asked to rate using a 6-point Likert scale (see Appendix B). The Likert-scale statements asked the participants to rate the perceived difficulty of the two test formats, the key features in interactive video-based tests, and their preferences regarding the two formats. Qualtrics, which is a web-based survey application, was utilised to design and administer the questionnaire.

4.3.3 Focus group interviews

To complement the self-reported questionnaire data with more in-depth information about the participants' experience using interactive videos in the IELTS Listening test, we conducted focus group interviews with small groups of participants (2–5 people per group). During the focus group interviews, we asked the participants questions about their preferences of the interactive video format vs. the audio-only format, strengths and limitations of each format, reasons for their preferences, and their perceptions regarding the effectiveness of the interactive video format for measuring L2 listening ability (see Appendix C for a focus group interview guide outlining the procedure and guiding questions).

4.4 Data collection procedure

After receiving approval for the study from research ethics boards at the University of Saskatchewan and at the University of Western Ontario, we recruited participants among the students who were L2 speakers of English at the respective universities. All the data were collected in May–June 2022.

Study participation entailed two in-person data collection sessions with small groups of participants. Thirty-three (33) participants at the University of Saskatchewan and 32 participants at the University of Western Ontario were assigned to one of the eight groups at each university, with four participants per group. (Note: due to the odd number of participants at the University of Saskatchewan, one of the groups comprised five rather than four people.) All data collection sessions took place in a classroom or in a computer lab on campuses of the above-mentioned two universities.

During the first session (Day 1) that lasted for approximately 40–50 minutes, each group of participants completed the questionnaire and one of the listening tests (either Listening Test A or B) in one of the two formats: the paper-based audio-only format or the computer-based interactive video format. The groups that were assigned a listening test in the computer-based interactive video format spent slightly more time because they were also required to complete the Practice Test to familiarise themselves with this format. The groups that were assigned a listening test in the paper-based audio-only format followed the standard test-taking procedure used for the operational IELTS Listening test. This procedure entailed listening to the audio stimuli that we played as mp3 files on a computer, taking notes in a test booklet, and then transferring the answers to the answer sheet.

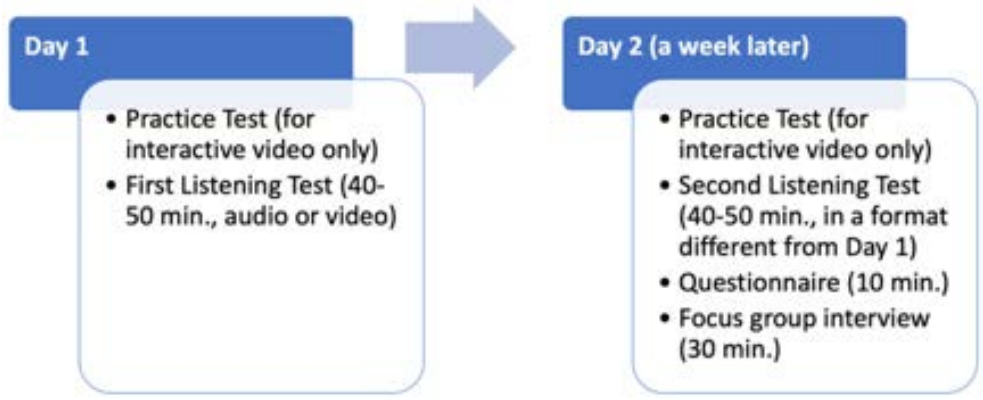
During the second session (Day 2) that was typically scheduled within a week after the first session, the same groups were asked to complete the second listening test in a format different from the format of the first listening test. Paper-based note-taking was allowed for all formats on both days. The order of the listening tests (i.e., Listening Tests A and B) and their formats were counterbalanced for each group (see Table 2).

Table 2: *Counterbalanced groups (N = 65)*

	Day 1	Day 2
Group 1	Listening Test A, audio-only	Listening Test B, interactive video
Group 2	Listening Test A, interactive video	Listening Test B, audio-only
Group 3	Listening Test B, audio-only	Listening Test A, interactive video
Group 4	Listening Test B, interactive video	Listening Test A, audio-only
Group 5	Listening Test A, audio-only	Listening Test B, interactive video
Group 6	Listening Test A, interactive video	Listening Test B, audio-only
Group 7	Listening Test B, audio-only	Listening Test A, interactive video
Group 8	Listening Test B, interactive video	Listening Test A, audio-only

After the second listening test (40–50 minutes), the participants were asked to fill out the online questionnaire (10 minutes) and take part in a focus group interview (up to 30 minutes) audio-recorded in the mp3 file format. As such, the second session lasted for approximately 80–90 minutes. The entire data collection procedure is shown in Figure 3.

Figure 3: *Data collection procedure*



4.5 Data preparation

Four types of data were gathered: test score data, questionnaire data, focus group interview data, and response time on the test items from the computer-based interactive video listening tests. Focus group data were qualitative, whereas the other three types of data were quantitative. Test score data consisted of all participants' (n = 65) responses to the items associated with the computer-based interactive video IELTS Listening tests and the paper-based audio-only IELTS Listening tests (i.e., 40 items x 2 listening tests per participant). Questionnaire data comprised the participants' responses to the 12 Likert-scale statements asking the participants to rate the perceived difficulty of the two test formats, the key features in interactive video-based tests, and their preferences regarding the two formats. Focus group interview data represented the participants' verbalisations of their perceptions and preferences of the interactive video format vs. the audio-only format. Finally, response time represented the amount of time (in seconds) spent by each participant on individual items from the computer-based interactive video listening tests.

To prepare the focus group interview data for analysis, we used a two-step approach to transcribe the mp3 files containing the interview recordings. During the first step, we generated automated transcripts using Trint, which is an AI-driven audio transcription tool (https://trint.com/). As these transcripts lacked accuracy (partly due to Trint's limited ability to recognise accented speech), we had to manually edit all the inaccuracies and finalise the transcripts as part of the second step.

Test score data and response time for the items associated with interactive video listening tests were extracted from the user data report provided by GrassBlade LRS as a .csv file. To obtain test score data for the items associated with audio-only listening tests, we had to manually transfer individual participants' responses from the paper-based answer sheets to an Excel spreadsheet for subsequent analysis.

4.6 Data analysis

Both quantitative and qualitative analyses were conducted to answer the research questions. To answer the first research question regarding the comparability of the computer-based interactive video IELTS Listening test and the paper-based audio-only IELTS Listening test, we leveraged descriptive statistics analysis, reliability analysis, paired-sample t-tests, ANOVA, and correlation analysis to examine the possible differences between the two formats. We then conducted bias analysis using many-facet Rasch modelling to identify the items that showed differential item difficulty between the formats.

To answer the second research question investigating the relationship between participants' response time and item-related characteristics, we conducted a mixed-effects model analysis with mean response time as a dependent variable. The independent variables of interest included item difficulty (facility) index, item discrimination index, item type (i.e., fill-in-the-blank, multiple-choice multiple-answer, multiple-choice single-answer, and drag-and-drop), item length in terms of word count, and testlet (topic). Among the independent variables, testlet was treated as a random effect to account for possible variations within these conditions. The lmer function in R package *lme4* (Bates et al., 2015) was used to carry out mixed-effects model analysis.

To answer the third research question enquiring about the participants' perceptions and preferences related to the use of interactive videos in the IELTS Listening test, we conducted qualitative analysis of the participants' responses to the focus group interviews and quantitative analysis of their responses to the Likert-scale items in the questionnaire. To analyse the participants' responses to the 12 Likert-scale items in the questionnaire, we calculated descriptive statistics.

To analyse the focus group interview data, we employed both deductive and inductive coding to develop a coding scheme and code the data in NVivo 12, which is software for qualitative data analysis. Deductive codes were informed by the guiding questions from the focus group interviews, whereas inductive codes were based on the themes found in the data. To ensure the reliability of coding, we followed Loewen and Plonsky's (2015) recommendation to double-code 10 to 20% of the qualitative data. Specifically, two trained coders who were research assistants for this project double-coded the interview data from two groups of participants (i.e., Group 5 at the University of Western Ontario and Group 8 at the University of Saskatchewan), which represented approximately 12% of the data. It should be noted that even though the two coders generally assigned the same codes to the data, they differed in the amount of text they were selecting to apply to each code, with one coder selecting more contextual information than the other coder. Because of these differences, we calculated Cohen's kappa at the paragraph level rather than at the sentence level. Based on the paragraph-level analysis, unweighted Cohen's kappa was .81 for Group 5 at the University of Western Ontario and .82 for Group 8 at the University of Saskatchewan, whereas the overall interrater agreement for both groups was 99%, indicating very good agreement. The final coding scheme comprised two levels of codes, with first-level codes corresponding to the main themes from the guiding questions used during the focus group interviews and the second-level codes representing the main themes that emerged from the data analysis. In total, there were 14 first-level codes and 120 second-level codes (see Appendix D). The qualitative analysis of the focus group interviews was used to supplement the quantitative analysis of the questionnaire data.

5 Results

5.1 RQ1: Comparison of test formats

Research Question 1 concerned the comparison of the listening tests in two formats (audio vs. interactive video) in terms of participants' performance. Table 3 presents the descriptive statistics of the raw scores from each of the four test forms. Among these four forms, Listening Test A in the interactive video format appeared to be the most difficult test with a mean of 28.2 and a relatively large standard deviation (SD = 6.57), whereas the easiest test was Listening Test A in the audio-only format (M = 32.5, SD = 4.41). The four forms also varied in terms of internal consistency reliability measured by Cronbach's alpha. Listening Test B in the interactive video format had the lowest reliability (α = 0.669).

Table 3: *Descriptive statistics and reliability for the listening test scores (N = 65)*

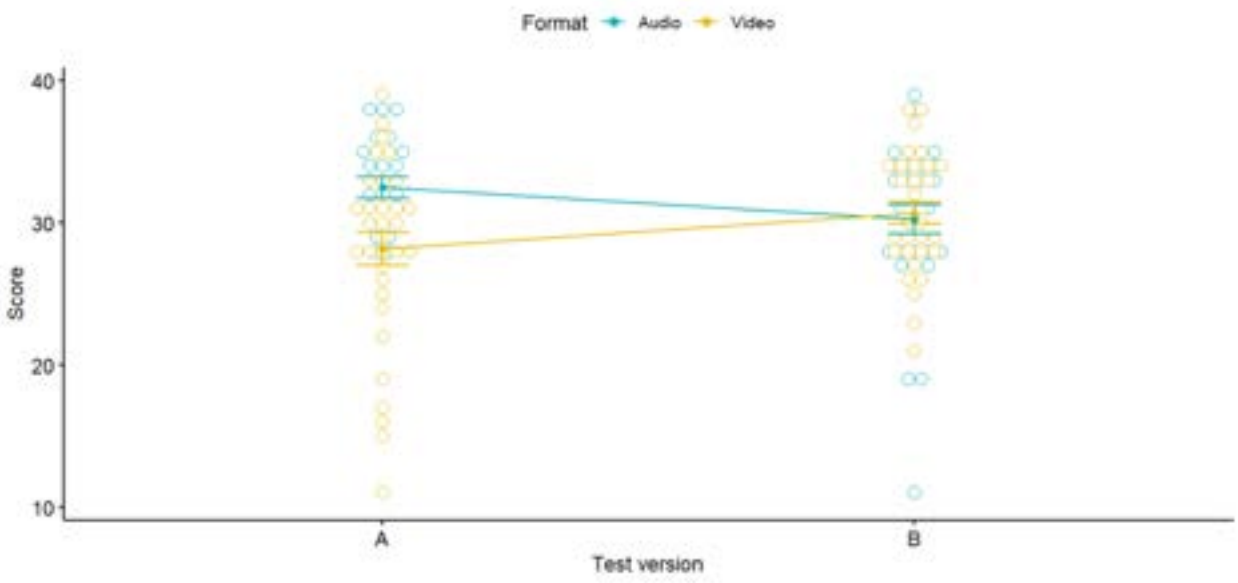
Listening test version and format	Mean (SD)	Min/Max	Cronbach's α	*n*
Test A (audio)	32.5 (4.41)	16/39	0.721	32
Test B (video)	30.7 (4.18)	21/38	0.669	32
Test A (video)	28.2 (6.57)	11/39	0.843	33
Test B (audio)	30.2 (5.86)	11/39	0.825	33

Prior to the comparison of the participants' performance on the tests of different formats, we examined the comparability of raw scores on Listening Test A and Listening Test B first using paired-sample t-tests and Pearson's correlations. Paired-sample t-tests revealed statistically significant differences between the scores on the audio-only Listening Test A (M = 32.5, SD = 4.41) and the interactive video Listening Test B (M = 30.7, SD = 4.18, t = 2.606, df = 31, p = .014, Cohen's d = 0.40), as well as between the scores on the interactive video Listening Test A (M = 28.2, SD = 6.57) and the audio-only Listening Test B (M = 30.2, SD = 5.86, t = -2.174, df = 32, p = .037,

Cohen's *d* = -0.38). The overall Pearson correlation coefficient for the scores from the two test versions (i.e., Listening Tests A and B) was 0.580 ($p < .001$, $N = 65$). Regarding the tests in different formats, the correlation of raw scores from the audio-only Listening Test A and the interactive video Listening Test B was 0.596 ($p < .001$, $n = 32$), whereas the correlation for the interactive video Listening Test A and the audio-only Listening Test B was 0.610 ($p < .001$, $n = 33$). Such relatively strong and statistically significant correlation coefficients suggested a strong relationship among these test forms.

The comparability of raw scores across the test formats was firstly investigated using ANOVA with an interaction between test version and format. Both residual normality and variance equality assumptions were checked. The results indicated a statistically significant interaction effect ($F = 6.400$, $df = 1$, $p = .013$, $\eta^2 = 0.05$) as well as a statistically significant main effect of the formats ($F = 4.231$, $df = 1$, $p = .042$, $\eta^2 = 0.03$), while the effect of test version was not statistically significant ($F = 0.038$, $df = 1$, $p = .845$). The result of a post-hoc pairwise comparison (Tukey HSD test) showed that, overall, the interactive video format was statistically significantly more difficult than the audio-only format (95% CI: -1.94, -3.80, $p = .042$). Figure 4 illustrates the raw score distributions for each test version and format, showing that Listening Test A was the only test version with a statistically significant difference in raw scores.

Figure 4: *Line graph for raw test scores by test version and format (N = 65)*



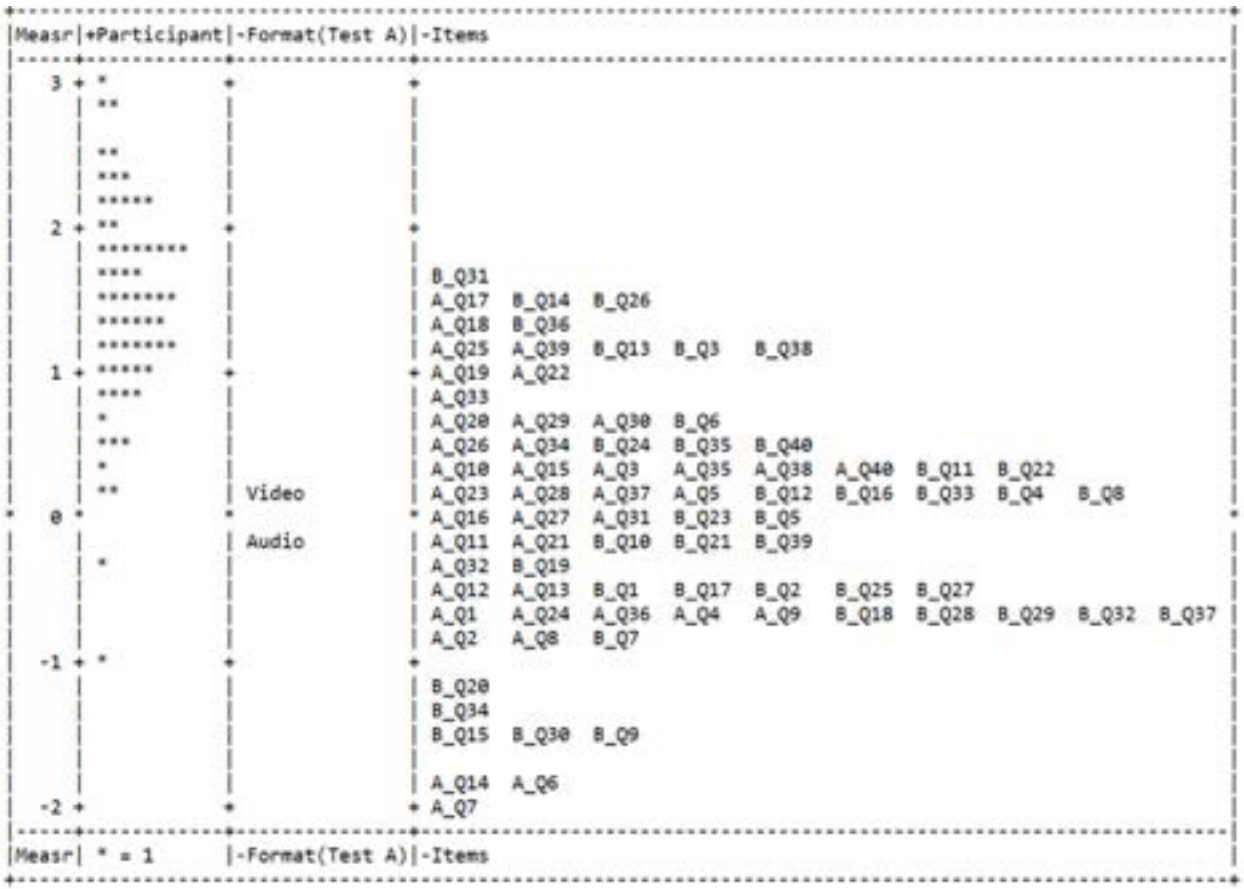
To better understand the effect of the formats at both the test level and the item level, we conducted a many-facet Rasch model (MFRM) analysis with three facets: participants, test format, and items. Table 4 reports the summary statistics for the three facets in the MFRM analysis. Overall, based on the infit information (mean, SD, and range) for the three facets, the many-facet Rasch model fits the data well.

Table 4: *Summary statistics for the MFRM analysis*

Facets	Mean measure in logits (SD)	Measure range	Mean infit MnSq (SD)	Infit MnSq range	Separation index	Strata
Participants (N = 65)	1.43 (0.74)	-1.01, 3.24	1.00 (0.08)	0.88, 1.21	2.22	3.29
Items (k = 40)	0.00 (0.86)	-2.23, 1.61	1.00 (0.11)	0.80, 1.34	2.21	3.28
Formats (k = 2)	0.00 (0.14)	-0.10, 0.10	1.01 (0.01)	0.99, 1.02	1.75	2.67

Figure 5 shows the Wright map of the MFRM output on the logit scale. The facet of 'Participant' confirms that the participants performed well on the tests, with an average measure of 1.43. The facet of 'Format' indicates that the format of interactive video was more difficult than the audio-only format (0.10 vs. -0.10 logit). The 'Items' facet shows that most of the items are centred around zero.

Figure 5: *Wright map of the facets of person measure, format and items*



A follow-up bias analysis in the MFRM indicated that 12 out of 40 items in Listening Test A had differential difficulty levels across the test formats (audio-only vs. interactive video), as indicated by measure differences of the value above 1 (see Table 5). The values in the column 'Contrast' are the outcomes of subtracting interactive video item measures from audio-only item measures. A negative sign of a 'Contrast' value indicates that the item in the audio-only format was easier than its video-based counterpart. Nine (9) out of the 12 flagged items in Table 5 were easier in the audio-only format, whereas only three items were more difficult in the audio-only format compared to the interactive video format. However, the results of the Rasch-Welch t-test indicated that only two items in Listening Test A were statistically significantly different across the two formats, with Item 19 being easier in the audio-only format and Item 29 being easier in the interactive video format. With the opposite direction of differences, the differential difficulty effect may be cancelled out or mitigated to some extent at the test level.

It is noteworthy that the topics of individual parts of the listening test (i.e., testlets) seem to have had some impact on the direction of differential difficulty. In Listening Test A, the flagged items in "Bankside recruitment agency" (Part 1) and "Matthew Island Holidays" (Part 2) were easier in the audio-only format than in the interactive video mode, while the two items in "Personality traits" (Part 3) were more difficult in the audio-only format. The last topic "Eucalyptus tree in Australia" (Part 4) had a mixture of items that had either negative or positive contrast values (see Table 5).

Table 5: *Interaction analysis (format by items) for Listening Test A*

Topic	Item number (Item type)	Audio measure (SE)	Interactive video measure (SE)	Contrast (audio-video)	Rasch-Welch *t* (df)	*p*
Bankside recruitment agency (Part 1)	Q2 (FB)	-2.01 (1.03)	-0.44 (0.48)	-1.57	-1.39 (43)	.173
	Q6 (FB)	-2.74 (1.43)	-1.32 (0.63)	-1.42	-0.90 (42)	.374
	Q8 (FB)	-2.01 (1.03)	-0.44 (0.48)	-1.57	-1.39 (43)	.173
Matthew Island Holidays (Part 2)	Q12 (MC)	-1.26 (0.74)	-0.03 (0.43)	-1.23	-1.44 (50)	.157
	Q14 (MC)	-2.74 (1.43)	-1.32 (0.63)	-1.42	-0.90 (42)	.374
	Q19 (FB)	0.24 (0.44)	1.47 (0.37)	-1.23	-2.12 (60)	.038*
	Q20 (FB)	0.04 (0.47)	1.2 (0.37)	-1.16	-1.94 (59)	.057
Personality traits (Part 3)	Q24 (DD)	0.04 (0.47)	-1.32 (0.63)	1.36	1.73 (58)	.089
	Q29 (MA)	1.35 (0.37)	-0.23 (0.45)	1.58	2.70 (61)	.009**
Eucalyptus trees in Australia (Part 4)	Q36 (FB)	-0.20 (0.5)	-1.32 (0.63)	1.13	1.40 (60)	.168
	Q38 (FB)	-0.47 (0.55)	0.77 (0.38)	-1.25	-1.87 (55)	.067
	Q39 (FB)	0.61 (0.41)	1.61 (0.37)	-1.00	-1.81 (62)	.075

Note. FB = fill-in-the-blank; MC = multiple-choice (single-answer); DD = drag-and-drop; MA = multiple-choice multiple-answer. *$p < .05$. **$p < .01$.

In Listening Test B, when only the difference magnitude was considered, 21 out of 40 items were flagged as showing differential difficulty levels across the test formats, with almost two thirds of these items ($k = 13$) being easier in the audio-only test format (see Table 6). This was especially obvious in the topic "Festival" (Part 1) as seven out of ten questions were flagged as easier in the audio-only format. The flagged items were more balanced in terms of contrast directions in the other three topics of Listening Test B. If we consider the results of the Rasch-Welch *t*-test, the number of items that displayed statistically significant differential difficulty levels across the two formats is reduced to 11, with ten of them being easier in the audio-only format.

Table 6: *Interaction analysis (format by items) for Listening Test B*

Topic	Item number (Item type)	Audio measure (SE)	Interactive video measure (SE)	Contrast (audio-video)	Rasch-Welch *t* (df)	*p*
Festival (Part 1)	Q1 (FB)	-1.32 (0.63)	0.24 (0.44)	-1.57	2.03 (57)	.047*
	Q2 (FB)	-1.79 (0.75)	0.43 (0.42)	-2.23	2.58 (50)	.013*
	Q3 (FB)	-0.03 (0.43)	2.19 (0.38)	-2.22	3.85 (62)	.001***
	Q5 (FB)	-0.97 (0.56)	0.77 (0.4)	-1.74	2.53 (57)	.014*
	Q6 (FB)	0.15 (0.42)	1.21 (0.38)	-1.06	1.9 (62)	.063
	Q7 (FB)	-1.79 (0.75)	-0.2 (0.5)	-1.6	1.77 (55)	.083
	Q9 (FB)	-2.56 (1.04)	-0.81 (0.62)	-1.75	1.45 (52)	.154
Minster Park (Part 2)	Q14 (MC)	0.92 (0.38)	2.05 (0.38)	-1.13	2.12 (62)	.038*
	Q16 (DD)	-0.69 (0.51)	0.92 (0.39)	-1.61	2.5 (59)	.015*
	Q18 (DD)	-0.23 (0.45)	-1.26 (0.74)	1.04	-1.2 (51)	.237
	Q20 (DD)	-0.44 (0.48)	-2.73< (1.43)	2.29	-1.51 (37)	.138
Charles Dickens (Part 3)	Q21 (MA)	0.32 (0.4)	-0.81 (0.62)	1.13	-1.52 (53)	.134
	Q25 (DD)	-1.79 (0.75)	0.43 (0.42)	-2.23	2.58 (50)	.013*
	Q26 (DD)	2.04 (0.39)	0.92 (0.39)	1.12	-2.04 (62)	.046*
Agricultural program in Mozambique (Part 4)	Q32 (FB)	-0.03 (0.43)	-2.73< (1.43)	2.70	-1.8 (36)	.080
	Q33 (FB)	0.63 (0.39)	-0.47 (0.55)	1.10	-1.64 (56)	.107
	Q35 (FB)	-0.23 (0.45)	1.07 (0.38)	-1.30	2.19 (61)	.032*
	Q36 (FB)	0.77 (0.38)	1.91 (0.37)	-1.13	2.12 (62)	.038*
	Q37 (FB)	-0.23 (0.45)	-1.26 (0.74)	1.04	-1.2 (51)	.237
	Q40 (FB)	-0.44 (0.48)	1.35 (0.37)	-1.80	2.96 (59)	.004**

Note. FB = fill-in-the-blank; MC = multiple-choice (single-answer); DD = drag-and-drop; MA = multiple-choice multiple-answer. * $p < .05$. **$p < .01$. ***$p < .001$.

In sum, our analyses revealed that the interactive video-based listening tests were more difficult than their audio-only counterparts. The results of bias analyses partially explain why Listening Test A in the interactive video format was more difficult than in the audio-only format. These differences will be further discussed in the context of the participants' responses to the focus group interviews and questionnaire.

5.2 RQ2: Response time on the interactive video listening tests

Research Question 2 examined the participants' response time for individual items in the interactive video listening tests. The results of the mixed-effects model analysis are summarised in Table 7. Overall, word count had a statistically significant positive effect on the participants' response time (β = 0.35, t = 4.83, p < .001), item difficulty (facility) had a statistically significant negative effect (β = -17.74, t = -4.52, p < .001), whereas item discrimination had a non-significant positive effect (β = 1.42, t = 0.475, p = .637). As for the item types, both multiple-choice multiple-answer (MA) items and multiple-choice single-answer (MC) items had a statistically significant negative effect on the response time when compared to the time spent on the drag-and-drop items (default for comparison). Taken together, our model indicates that the participants' response time was affected positively by the word count (including the options in multiple-choice items) but negatively by item difficulty (that is, easier items required less time). Meanwhile, selected-response item types (i.e., multiple-choice multiple-answer items and multiple-choice single-answer items) took participants less time to answer compared to other item types.

Table 7: *Summary of mixed-effects model for response time in interactive video listening tests*

Effect	Response time (mean)		
	Estimates	95% CI	*p*
Predictors			
Intercept	24.72	17.15 – 32.29	0.001***
Word count	0.35	0.20 – 0.49	0.001***
Item difficulty	-17.74	-25.56 – -9.91	0.001***
Item discrimination	1.42	-4.55 – 7.40	0.636
Item type (FB)	-0.31	-3.17 – 2.54	0.828
Item type (MA)	-9.75	-15.49 – -4.02	0.001***
Item type (MC)	-8.45	-13.34 – -3.56	0.001***
Random Effects			
σ^2	20.96		
$\tau_{00\ Part\_video}$	0.34		
ICC	0.02		
N_{Part_video}	4		
Observations	80		
Marginal R^2 / Conditional R^2	0.415 / 0.424		

Note. FB = fill-in-the-blank; MA = multiple-choice multiple-answer; MC = multiple-choice (single-answer). Drag-and-drop (DD) was used as the reference item type. ***p < .001.

5.3 RQ3: Test-takers' perceptions and preferences

With the qualitative data from the focus group interviews and the quantitative data from the questionnaire, the last research question probed test-takers' perceptions and preferences regarding the effectiveness of interactive videos for measuring L2 listening ability.

5.3.1 Perceptions

Participants' responses to the 12 Likert-scale statements in the questionnaire are tabulated in Table 8. Overall, the participants shared a mixed perception of the difficulty of the interactive video listening tests: Slightly more than half of the participants (38 out of 65, or 58%) believed that the listening tests in the interactive video format were more difficult than in the audio-only format (Statement 1), whereas the majority of participants (47 out of 65, or 72%) did not find the items in the interactive video listening tests to be more difficult than the items in the audio-only listening tests (Statement 7). Meanwhile, only 24 participants (37%) agreed that they watched the videos all the time during the test (Statement 9). These response patterns suggest that it was mainly the format that made the interactive video listening tests more challenging for the participants.

Table 8: *Summary of responses to 12 Likert-scale statements in the questionnaire (N = 65)*

	Likert scale points						M	*SD*
Statements	**1**	**2**	**3**	**4**	**5**	**6**		
1. The interactive video-based listening test is more difficult than the audio-only listening test.	3	15	9	15	13	10	3.77	1.51
2. The animation used in the interactive video-based listening test facilitated my listening comprehension.	8	10	13	13	16	5	3.52	1.51
3. The gestures used by the animated characters helped me answer some questions on the video-based listening test.	6	19	12	11	16	1	3.23	1.39
4. The questions embedded in the interactive videos distracted me from listening.	11	16	8	11	9	10	3.32	1.39
5. The visuals in the animation helped me better understand the content of the video.	2	10	8	22	18	5	3.91	1.27
6. The animation in the interactive videos helped me predict what may happen next in the video.	8	13	15	14	12	3	3.28	1.42
7. The questions in the interactive video-based listening test were more difficult than the questions in the audio-only listening test.	10	23	14	7	8	3	2.83	1.42
8. I felt more confident completing the audio-only listening test than the interactive video-based listening test.	4	10	10	11	17	13	4.02	1.56
9. I watched the interactive videos during the listening test all the time.	9	19	13	10	8	6	3.11	1.53
10. It was easier for me to take notes during the audio-only listening test than during the video-based listening test.	3	12	8	9	16	17	4.14	1.61
11. The audio-only listening test provided a more accurate measurement of my listening comprehension skills.	2	5	10	19	18	11	4.22	1.29
12. Overall, I preferred the interactive video-based listening test.	9	19	9	15	10	3	3.11	1.46

Note. 6-point Likert scale: 1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Somewhat agree, 5 = Agree, 6 = Strongly agree.

Similar responses were found in the focus group interviews. Overall, the interview data confirmed that the interactive video format was not widely known to the participants as only 19 participants had encountered interactive videos prior to the experiment at either university-level courses or non-university courses (e.g., videos from Coursera or a MOOC website). Regarding the advantages of the interactive video-based listening tests, the most frequently mentioned advantage was the amount of information contained in the videos (36 participants, or 55%). For example, Participant 204 stated that she had been able to extract more information from the interactive video than from the audio. Similarly, Participant 227 recognised that, compared to the audio-only format, the interactive video format had provided him additional information:

> So the main difference [between the audio and interactive video input] for me was the body language of the people in the video, the setting and the environment that they're in. So, I found that that was useful in my listening and understanding the context of the information that has been given.

In addition, the participants mentioned some unique features of H5P interactive videos that they found useful. One of these features entailed using the dots in the progress bar of a video to identify the location of test items. As claimed by 22 participants, seeing the dots helped them—to some extent—prepare for the upcoming questions. Some participants (n = 15) further stated that they had even adopted some strategies for this feature. For example, Participant 113 made the following statement:

> I think also there's one advantage of video-based question is that you cannot miss a point. It stops right after the question, so you can see that there's a question, so you are waiting for this question, so it's coming right now. So, video gives you some orientation, so you can feel that you're inside that paragraph and you can find the keywords. This is my strategy for this. I don't know if it's working or not.

Another unique feature brought up by some participants was automated instant feedback on the embedded items. On the one hand, eight participants saw it as an advantage of the interactive video format, mainly because it could confirm their correct responses or help them pay attention to their errors. On the other hand, five participants explicitly complained about this feature, saying that instant feedback was distracting. When the participants were asked to evaluate the effects of instant feedback, 39 participants held a negative view of this feature, mainly because it caused confusion, induced anxiety, and/or impacted one's mood or concentration. The experience of Participant 106 illustrates this issue:

> Sometimes I was confused because I was sure it was a right answer. So, in the other test, I didn't know if it was right or not. So, it was alright. Just sometimes I said to myself: 'Maybe I did add 's' the end, and it wasn't' [referring to a fill-in-the-blank item]. So, the answer was just without 's'.

Meanwhile, 14 participants supported this feature for its motivating effect. The reaction of Participant 115, for example, highlighted this view: "I think like really knowing that I did correctly kind of motivates me to, like, doing well. And if I notice that I did something wrong, like it will help me to like, be more cautious." Only eight participants remained neutral regarding this feature because, according to them, they did not care much about the instant feedback.

The insights from the focus group interviews corroborated the responses to the questionnaire. For example, among the key features of the interactive video listening tests (see Table 8), the embeddedness of questions (Statement 4) was seen as distracting by almost half of all the participants (n = 30, or 46%).

Animations were found facilitative to the listening comprehension (Statement 2) by 34 participants (or 52%), and 29 participants (45%) thought that they helped them make predictions (Statement 6). Visuals in the animated videos (Statement 5) appeared to be helpful to the majority of the participants (45 participants, or 69%), whereas gestures were perceived as beneficial (Statement 3) by only 28 participants (43%).

5.3.2 Preferences

Four Likert scale-based statements in the questionnaire were related to the participants' preference of the test format. Thirty-seven (37) out of 65 participants (57%) preferred the audio-only listening tests (Statement 12). This finding was aligned with the participants' responses to the statements that compared the two formats, for example, Statement 8 (confidence level), Statement 10 (ease of note-taking), and Statement 11 (perceived construct validity). Approximately two-thirds of the participants felt more confident completing the audio-only listening test (n = 41, or 63%) and found it easier to take notes in that format (n = 42, or 65%), with three quarters of the participants perceiving the audio-only listening test as a more accurate measurement of their listening comprehension skills (n = 48, or 74%).

During the focus group interviews, we invited the participants to reflect on their preferences first from the perspective of a test-taker and then from the perspective of a language learner. Interestingly, a clear distinction in the preferences emerged. As test-takers, most of the participants (41 out of 65, or 63%) preferred the audio-only format, with 10 participants advocating for a hybrid format that would combine the video input with the test items accessible in a printed test booklet. For example, Participant 118 said, "I would choose audio-based, but if, maybe, there is a design, like a hybrid, it would be interesting to take." Only 16 participants expressed preference for the interactive video format from the perspective of a test-taker. Participant 116 offered the following explanation:

> I think, having had this experience now, I would choose the video-based even though it was a little bit harder to really focus, I would just try to focus more. Because it does help out, like, you have turn in conversation, then you have a question. Well, if you're taking just the audio-based, you have all the questions there and you have no idea when the answer will come up. So, I might take, although I would be more comfortable with the audio, I still think I would go for the video.

By contrast, when wearing the language learner's hat, nearly half of the participants (n = 32) appreciated the interactive video format and preferred it over the audio-only one. In addition, 18 participants stated that they would accept both formats for language learning purposes. In other words, while the participants recognised the benefits of interactive videos as a language learning resource, the use of interactive videos as a testing format was viewed less favourably.

We also invited the participants to offer their suggestions for improving future designs of interactive video-based listening tests. Four major themes emerged in the participants' suggestions: visibility of questions, video content design, availability of feedback, and an on-screen timer. First, 40 participants called for keeping the questions visible during the listening test either as part of the video or outside the video frame (e.g., by using the split screen). As Participant 116 stated, "I think that was the only problem I had, I didn't have the questions. So, if I have questions, it would be perfect." The second group of suggestions concerned the content of interactive videos. Four participants advocated for using real people in the video instead of animated characters to make the videos more realistic. As pointed out by one of the participants, some visual effects such as thought bubbles applied to some characters in our animated videos did not appear to be very relevant to the content of the conversation.

Third, 18 participants suggested (a) replacing instant feedback on each item with delayed feedback at the end of the test to avoid undue pressure on the test-takers, or (b) making instant feedback optional so that test-takers could decide whether they need to see it or not. Participant 209's comment conveyed this shared opinion:

> I'm really struggling with the prompt like feedback. So, I would suggest maybe they just, we can submit the answer but without getting any feedback because I think for high stakes exams, like I told you, if you get some feedback, it might influence your later performance. Because you're already very nervous and then you think, you know, you got the answer wrong.

Lastly, nine participants wanted to have an on-screen timer for answering the test items so that they could monitor and better manage their time during the test. The following comment made by Participant 234 was echoed by several other participants:

> Maybe we can set a time limit for each question, but actually it will increase the anxiety. Yes, but if I don't have a timer, I might take too many time to previous questions so I don't have enough time for the last questions.

Overall, the audio-only listening tests appeared to be more favoured by the participants, primarily due to the fact that they were more accustomed to this traditional format. Nevertheless, the interactive videos were also appreciated, mainly from the language learning perspective, suggesting their potential future use as a testing format.

6. Discussion and conclusion

The main purpose of this study was to compare L2 test-takers' performance on a computer-based version of the IELTS Listening test that uses interactive videos with their performance on the traditional paper-based audio-only version of this listening test. In addition, the study aimed to examine how these test-takers' perceptions and preferences related to the two formats, as well as the test-takers' response processes during the completion of the IELTS Listening test in the interactive video format. In doing so, the study was guided by three research questions. In this section, we summarise and interpret the findings, mention some limitations, and conclude with suggestions for further research.

6.1 RQ1: Comparison of test formats

In response to Research Question 1 that enquired about the extent to which test-takers' performance on a computer-based interactive video IELTS Listening test differed from their performance on a paper-based audio-only IELTS Listening test, we found a statistically significant difference between the scores on the audio-only Listening Test A and the interactive video Listening Test B, as well as between the scores on the interactive video Listening Test A and the audio-only Listening Test B. The results of ANOVA revealed a statistically significant interaction effect between test version and format and a statistically significant main effect of the format, with the interactive video format being statistically significantly more difficult than the audio-only format of the listening test. A follow-up bias analysis in the MFRM demonstrated that out of the 13 items (i.e., two items in Listening Test A and 11 items in Listening Test B) that exhibited statistically significant differential difficulty levels between the two formats, 11 items were more difficult in the interactive video format. These findings suggest that, overall, the interactive video format was more difficult than the audio-only format, resulting in lower listening test scores.

To better understand why the test-takers received lower scores on the listening tests in the interactive video format than on the audio-only listening tests, we took a closer look at the items that exhibited statistically significant differential difficulty levels between the two formats. First, we noticed that some of the visual cues in the animated videos seemed to have misled the test-takers and resulted in incorrect answers. For instance, when answering Item 20 in Listening Test A, which required a one-word response 'capital', some participants were seemingly affected by the picture of a castle in the video as they provided responses that appeared to describe the castle.

Second, for several items, there was a fairly long interval between the point of the video that presented the information subsequently asked in an item and the time when the item popped up on the screen. For example, the word 'review' which was the correct answer to Item 2 in Listening Test B appeared in the video 14 seconds before the item was displayed on the screen. Even though we tried to embed all the items immediately after the relevant information was presented in the video (which was usually the end of an idea unit, as described in the Materials section above), in some cases we were not able to avoid longer intervals between the relevant input and the item without breaking up the idea unit. It appears that these intervals were too taxing for the test-takers' memory, increasing their cognitive load and having a detrimental impact on their answers.

Another reason that can explain why our participants received lower scores on the listening tests in the interactive video format concerns item access. In the audio-only format, the test-takers had access to the items in the test booklet throughout the entire listening test. In the interactive video format, however, the test-takers could preview the items before starting each part of the listening test, but they did not have access to them while watching the video. Even though the test-takers had initial access to the items and could preview them before beginning each part of the listening test, it is likely that they did not remember all the items due to memory decay. As a result, the test-takers did not always know which exact information they should be looking for or focusing on while watching the video, which might have increased the cognitive load for these tasks and, as a result, led to lower scores on the items in the interactive video format.

It should also be acknowledged that the items used in this study were originally designed for a paper-based audio-only while-listening-performance test, in which the test-takers could read and answer the questions as they were listening to the audio input. Adopting these items for a computer-based interactive video-mediated post-listening-performance test could have had an impact on the construct of listening measured by the new test format. It is possible that answering such items in the video-based format required test-takers to use a larger amount of mental processing power (i.e., heavier cognitive load). Specifically, it can be hypothesised that the type of "interrupted viewing" that test-takers were engaged in (i.e., watching a segment of the video that was automatically paused and interrupted by a pop-up question) could have been disruptive for their listening comprehension flow and posed a heavier cognitive load on their working memory, resulting in lower scores on the items associated with interactive videos.

Several possible reasons can also be put forward to explain statistically significantly higher scores on some items in the interactive video format. First, the automated immediate feedback provided by H5P on each item seemed to have helped some test-takers eliminate incorrect options in drag-and-drop items (which, for instance, was the case with Item 24 in Listening Test A). Second, certain visuals in the interactive videos contained minor cues that seem to have helped some test-takers (e.g., Item 32 in Listening Test B). Third, the visibility of item locations in the form of circles or dots, in conjunction with question previewing, may have promoted test-takers' use of meta-cognitive strategy, thus facilitating their comprehension of details relevant to upcoming items.

Finally, the test-takers tended to perform better in the interactive video format on the items presented immediately after the relevant information was mentioned in the video. For instance, the word 'preservation', which was the correct answer to Item 37 in Listening Test B, was the last word in the video before Item 37 was shown.

6.2 RQ2: Response time on the interactive video listening tests

In Research Question 2, we examined the association between test-takers' response time on the items from the interactive video listening tests and item characteristics such as word count, item difficulty, item discrimination, and item type. The results of the mixed-effects model analysis showed that the participants' response time was positively affected by the word count (i.e., the more words a test item contained, the more time the participants would spend answering it). Meanwhile, the response time was negatively affected by item difficulty and item type, namely, multiple-choice multiple-answer items and multiple-choice single-answer items. In other words, the participants appeared to spend less time answering easier items and selected-response items in the interactive video format.

Two explanations can be offered in relation to these findings. First, when encountering easy items, the test-takers did not need to spend extra time thinking about the correct answer as they apparently knew it. This relationship between response time and item difficulty was not unexpected. It is also likely that, unlike fill-in-the-blank items, selected-response items took less time to answer because the test-takers did not need to compose their answer, which would generally take more time than selecting the best option.

6.3 RQ3: Test-takers' perceptions and preferences

Research Question 3 explored the test-takers' perceptions and preferences regarding the effectiveness of interactive videos for measuring their L2 listening ability. In their responses to the 12 Likert-scale statements in the questionnaire, the participants expressed somewhat mixed perceptions of the interactive video format. Specifically, while over half of all the participants found the interactive video format to be more difficult than the audio-only format, the vast majority of them considered visuals in the animated videos to be helpful and perceived the items embedded in the videos to be less difficult than the audio-only items. The focus group interviews revealed a number of additional features of the interactive video format, such as the visual indication of the upcoming items embedded in the video and instant feedback, that many participants perceived to be valuable (in line with He, 2022).

Regarding their preferences, slightly more than half of the participants indicated in the questionnaire that they preferred the audio-only format, with a larger percentage viewing it as a more accurate measure of their listening skills and feeling more confident about taking a listening test in this format. This preference for the audio-only format over the interactive video format can largely be attributed to their lack of prior experience with interactive videos, as evidenced from many participants' comments during the focus group interviews. It is also possible that the lack of prior experience with interactive videos resulted in lower scores on the listening tests delivered in this format. However, as we did provide the Practice Test that comprised a short video with five embedded items, the extent to which the reported lack of familiarity with the new format really affected test-takers' performance might have been fairly limited. The focus group interviews pointed to similar findings, namely, the strong preference for the audio-only listening test. However, when asked to reflect on their preferences from the language learning perspective, half of the participants chose interactive videos as their preferred format.

The key factor that can explain the participants' preferences and perceptions of the two formats concerns the participants' close familiarity with the audio-only listening tests. Had all the participants in our study had the same amount of experience with the interactive videos as they did with the audio-only format, it is likely that their perceptions of the interactive video format would have been different.

6.4 Implications and limitations

This study has several important implications. First, it supports the notion that videos increase the pedagogic value of L2 listening assessment instruments, as evidenced from the participants' comments about the benefits of the interactive video format. Including videos in a test of L2 academic listening ability can provide test-takers with visual information that complements the auditory input and creates listening conditions that better resemble the characteristics of the target language use domain compared to an audio-only condition (Ockey & Wagner, 2018). Second, our findings suggest that the integration of interactive videos in language assessment instruments can have a facilitative impact on test-takers' performance. In particular, certain features of the interactive videos such as embedded questions, as well as instant feedback and the visual indication of the upcoming questions, can help some test-takers during the listening assessment process.

The results of our study also have important implications for the construct of L2 listening. Specifically, they imply that the constructs measured by the audio-only format and the interactive video format might be quite different. As discussed in Section 4, Methodology, the listening tests that we used to gather performance data from the 65 participants differed not only in terms of the input format (i.e., audio vs. interactive video), but also in terms of the delivery mode (i.e., paper-based vs. computer-based). More importantly, there were also some key differences in the test methods: The audio-only listening tests were primarily while-listening performance tests (WLP), whereas the interactive video listening tests had the features of post-listening performance tests (PLP). As defined by Aryadoust (2012), WLP tests require the test-takers to read and respond to the items while they are listening to the stimulus, whereas in PLP tests, the test-takers complete those tasks sequentially rather than simultaneously (i.e., first listen to the stimulus and then read and respond to the test items). Statistically significant differences between the scores for the audio-only items and the items associated with interactive videos (as described in Section 5.1) offer further empirical evidence pointing to the existence of different constructs in our study. Given these differences, it is likely that the cognitive load imposed by the interactive video format was higher than the cognitive load imposed by the audio-only format, even though we hypothesised that interactive videos would be less taxing for our participants as this format would obviate the need for them to decide when each item should be answered.

Another implication of this study concerns note-taking practices during L2 listening assessments. Even though the explicit examination of note-taking was not the focus of our study, we noticed the differences in our participants' note-taking behaviour during the tests. During the interactive video listening tests, the participants took fairly extensive notes at the global level, jotting down all key points mentioned in the videos and attempting to copy entire questions during the preview stage. On the other hand, when completing the audio-only listening tests, the participants took notes at a more local level, jotting down only information related to specific questions in the test booklet. Anecdotal evidence also suggests that in our study, paper-based note-taking had a detrimental impact on the extent to which the participants watched the interactive videos (which, in turn, could have affected their performance on the items embedded in the interactive videos). It remains unclear whether the scores on the interactive video

listening tests would have been different, had the study allowed for computer-based note-taking (e.g., by splitting the screen into two areas: one area to display the interactive video with the embedded test items and another area to take notes).

The study has three main limitations. First, despite our intention to use two comparable versions of the IELTS Listening test, comparability appeared to be an issue. Even though we used two retired IELTS Listening tests, Listening Test B turned out to be noticeably less reliable than Listening Test A in both formats.

We do not have a satisfying explanation for this discrepancy in reliability and can only speculate about the potential underlying reasons, such as the possible effect of prior knowledge of the topic or the use of test-deviousness strategies such as random guessing. Administering the listening tests to a larger sample of participants could have potentially increased the reliability of both tests. Overall, the lack of comparability between Listening Test A and B might have skewed some of our findings, especially in relation to Research Question 1.

The second limitation concerns the comparison of the two test formats. As discussed above, the audio-only listening tests and the interactive video listening tests differed not only in the format of their stimuli (i.e., audio vs. interactive video), but also in their delivery modes (i.e., paper-based vs. computer-based) as well as the test methods (i.e., WLP vs. PLP tests). As a result, the constructs measured by the two formats might have been different, although the extent of this difference remains unknown and warrants further investigation. It should also be noted that, due to practicality reasons, we used animated videos rather than live videos, which limits the generalisability of our findings. As mentioned earlier, a number of design decisions were made when developing the interactive video versions. One of them involved balancing the need for including visuals that provided general background while avoiding visuals with specific cues that could be used to answer the questions. Undoubtedly, such a concern restricted the number and types of visuals that we were able to include in the animated videos, which, to some extent, reduced the authenticity of the visual content. In addition, several H5P-specific features such as enforced provision of instant feedback and item location markers may have prompted our test-takers to use different meta-cognitive strategies, which, in turn, could have altered their test-taking experiences and impacted their test scores.

The final limitation pertains to the data analysis for Research Question 2. When extracting the data from GrassBlade for multiple-choice multiple-answer items, we realised that the response time recorded by GrassBlade applied to item pairs rather than individual items (e.g., combined response time for Items 29 and 30 in Listening Test A rather than individual response time for each of these items). As there was no way for us to know exactly how much time the participants spent on each of these items, we simply split the response time for item pairs in half. In our view, this limitation was fairly minor as there were only two pairs of items in Listening Test B (i.e., Items 21–22 and 23–24) and one pair of items in Listening Test A (Items 29–30).

6.5 Directions for further research

Based on the results of this study, a series of recommendations for future research on interactive videos in L2 assessment contexts can be made.

Despite somewhat mixed findings of this study, test developers should consider the possibility of integrating interactive live videos in their assessment instruments as they seem to be more authentic for assessing L2 academic listening and are generally perceived more positively by L2 learners (cf. He, 2022). If live videos were to be used in the IELTS Listening test, it would be essential to design test items that would measure not only the ability to comprehend auditory information, but also the ability to understand and process visual information from such videos.

In particular, there is a need for multimodal listening tests designed to explicitly measure L2 test-takers' ability to comprehend and interpret gestures, facial expressions, and other multimodal elements that, combined with the auditory input, comprise semiotic resources (Gruba, 2020). One way to analyse how L2 test-takers make use of the vast repertoire of semiotic resources available to them during multimodal listening tests is through the systemic functional approach to multimodal discourse analysis, or SF-MDA (O'Halloran, 2008; O'Halloran & Fei, 2014). Viewing language as a social-semiotic system, the SF-MDA approach allows for examining the meaning of various semiotic resources in multimodal texts.

In addition to SF-MDA, eye-tracking studies, as well as other types of process-oriented validation research, can be leveraged to scrutinise the extent to which L2 test-takers actually watch the interactive videos. Finally, by providing their participants with ample opportunities to familiarise themselves with interactive video-based testing, researchers would be able to control for the effect that the familiarity with this format may have on their participants' performance, perceptions, and preferences.

In terms of materials design, test developers should consider the option of giving the test-takers continuous access to the items throughout the interactive video listening test, as suggested by our participants, and conduct empirical research on the effect of this option. Determining the most suitable timing for embedding the test items and investigating the effect of computer-based note-taking vs. paper-based note-taking during the listening tests with interactive videos would further expand this line of research and better inform the design of such tests.

REFERENCES

Aryadoust, V. (2012). Differential item functioning in while-listening performance tests: The case of the International English Language Testing System (IELTS) listening module. *International Journal of Listening*, 26(1), 40–60. https://doi.org/10.1080/10904018.2012.639649

Bakla, A., & Mehdiyev, E. (2022). A qualitative study of teacher-created interactive videos versus YouTube videos in flipped learning. *E-Learning and Digital Media*, 19(5). https://doi.org/10.1177/20427530221107789

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1) 1–48. https://doi.org/10.18637/jss.v067.i01

Batty, A. O. (2021). An eye-tracking study of attention to visual cues in L2 listening tests. *Language Testing*, 38(4), 511–535. https://doi.org/10.1177/0265532220951504

Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. *System*, 25(1), 39–53. https://doi.org/10.1016/S0346-251X(96)00059-0

Cambridge Business English Dictionary. (n.d.) Interactivity. In *Cambridge Business English Dictionary*. Retrieved 8 May 2023, from https://dictionary.cambridge.org/dictionary/english/interactivity

Cambridge University Press. (2020). *IELTS 15 Academic: Authentic practice tests*.

Campbell, L. O., Planinz, T., Morris, K., & Truitt, J. (2019). Investigating undergraduate students' viewing behaviors of academic video in formal and informal settings. *College Teaching*, 67(4), 211–221. https://doi.org/10.1080/87567555.2019.1650703

Casañ Núñez, J. C. (2017). Testing audiovisual comprehension tasks with questions embedded in videos as subtitles: A pilot multimethod study. *The EuroCALL Review*, 25(1), 36–60. https://doi.org/10.4995/eurocall.2017.7062

Delen, E., Liew, J., & Willson, V. (2014). Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments. *Computers and Education*, 78, 312–320. https://doi.org/10.1016/j.compedu.2014.06.018

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing Learner Language*. Oxford University Press.

Field, J. (2019). Second language listening: Current ideas, current issues. In J. W. Schwieter & A. Benati (Eds.), *The Cambridge Handbook of Language Learning* (pp. 283–319). Cambridge University Press. https://doi.org/10.1017/9781108333603.013

Gruba, P. (2020). What does language testing have to offer to multimodal listening? In G. J. Ockey & B. A. Green (Eds.), *Another Generation of Fundamental Considerations in Language Assessment: A festschrift in honor of Lyle F. Bachman* (pp. 43–57). Springer. https://doi.org/10.1007/978-981-15-8952-2_4

Gruba, P., & Suvorov, R. (2020). Technology and second language listening. In M. A. Peters & R. Heraud (Eds.), *Encyclopedia of Educational Innovation* (pp. 1–7). Springer. https://doi.org/10.1007/978-981-13-2262-4_142-2

Haagsman, M. E., Scager, K., Boonstra, J., & Koster, M. C. (2020). Pop-up questions within educational videos: Effects on students' learning. *Journal of Science Education and Technology*, 29(6), 713–724. https://doi.org/10.1007/s10956-020-09847-3

He, S. (2022). *Exploring the use of interactive videos in an L2 listening test* [Unpublished master's thesis]. University of Western Ontario.

Kang, T., Arvizu, M. N. G., Chaipuapae, P., & Lesnov, R. O. (2019). Reviews of academic English listening tests for non-native speakers. *International Journal of Listening,* 33(1), 1–38. https://doi.org/10.1080/10904018.2016.1185210

Ketsman, O., Daher, T., & Colon Santana, J. A. (2018). An investigation of effects of instructional videos in an undergraduate physics course. *E-Learning and Digital Media*, 15(6), 267–289. https://doi.org/10.1177/2042753018805594

Kleftodimos, A., & Evangelidis, G. (2016). Using open source technologies and open internet resources for building an interactive video based learning environment that supports learning analytics. *Smart Learning Environments*, 3(1), 1–23. https://doi.org/10.1186/s40561-016-0032-4

Lesnov, R. O. (2022). Furthering the argument for visually inclusive L2 academic listening tests: The role of content-rich videos. *Studies in Educational Evaluation*, 72, 101087. https://doi.org/10.1016/J.STUEDUC.2021.101087

Li, Z., Banerjee, J., & Zumbo, B. D. (2017). Response time data as validation evidence: Has it lived up to its promise, and if not, what would it take to do so. In B. D. Zumbo & A. M. Hubley (Eds.), *Understanding and Investigating Response Processes in Validation Research* (pp. 159–177). Springer International Publishing. https://doi.org/10.1007/978-3-319-56129-5_9

Linacre, J. M. (2022). Winsteps® (Version 5.3.20). Winsteps.com. https://www.winsteps.com/

Loewen, S., & Plonsky, L. (2015). *An A–Z of Applied Linguistics Research Methods*. Palgrave Macmillan.

Mayer, R. E. (2020). *Multimedia Learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355

Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, 24(4), 517–537. https://doi.org/10.1177/0265532207080771

Ockey, G., & Wagner, E. (2018). *Assessing L2 Listening: Moving towards authenticity*. John Benjamins. https://doi.org/10.1075/lllt.50

O'Halloran, K. L. (2008). Systemic functional-multimodal discourse analysis (SF-MDA): Constructing ideational meaning using language and visual imagery. *Visual Communication*, 7(4), 443–475. https://doi.org/10.1177/1470357208096210

O'Halloran, K. L., & Fei, V. L. (2014). Systemic functional multimodal discourse analysis. In S. Norris & C. D. Maier (Eds.), *Interactions, Images and Texts: A reader in multimodality* (pp. 137–154). De Gruyter Mouton.

Pusey, K., & Lenz, K. (2014). Investigating the interaction of visual input, working memory, and listening comprehension. *Language Education in Asia*, 5(1), 66–80. http://dx.doi.org/10.5746/LEiA/14/V5/I1/A06/Pusey_Lenz

Rice, P., Beeson, P., & Blackmore-Wright, J. (2019). Evaluating the impact of a quiz question within an educational video. *TechTrends*, 63(5), 522–532. https://doi.org/10.1007/s11528-019-00374-6

Sablić, M., Mirosavljević, A., & Škugor, A. (2020). Video-based learning (VBL)—past, present and future: An overview of the research published from 2008 to 2019. *Technology, Knowledge and Learning,* 1–17. https://doi.org/10.1007/s10758-020-09455-5

Suvorov, R. (2013). *Interacting with visuals in L2 listening tests: An eye-tracking study* (Publication No. 3597397) [Doctoral dissertation, Iowa State University]. ProQuest Dissertations Publishing. https://doi.org/10.31274/etd-180810-667

Suvorov, R. (2015). 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,* 32(4), 463–483. https://doi.org/10.1177/0265532214562099

Suvorov, R., & He, S. (2022). Visuals in the assessment and testing of second language listening: A methodological synthesis. *International Journal of Listening*, 36(2), 80–99. https://doi.org/10.1080/10904018.2021.1941028

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Taslibeyaz, E. (2020). The effect of scenario-based interactive videos on English learning. *Interactive Learning Environments*, 28(7), 808–820. https://doi.org/10.1080/10494820.2018.1552870

Tweissi, A. (2016). *The effects of embedded questions strategy in video among graduate students at a middle Eastern university.* [Publication No. 10393019]. [Doctoral dissertation, Ohio University]. ProQuest Dissertations Publishing.

van der Meij, H., & Böckmann, L. (2021). Effects of embedded questions in recorded lectures. *Journal of Computing in Higher Education*, 33(1), 235–254. https://doi.org/10.1007/s12528-020-09263-x

Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning & Technology,* 11(1), 67–86. http://dx.doi.org/10125/44089

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

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

Xu, J. (2017). The mediating effect of listening metacognitive awareness between test-taking motivation and listening test score: An expectancy-value theory approach. *Frontiers in Psychology*, 8, 2201. https://doi.org/10.3389/FPSYG.2017.02201

Zou, & Xie, H. (2019). Flipping an English writing class with technology-enhanced just-in-time teaching and peer instruction. *Interactive Learning Environments*, 27(8), 1127–1142. https://doi.org/10.1080/10494820.2018.1495654

Appendix A: Audio transcripts, test items, and answer keys

Practice Test

Transcript

SALLY: Good morning. Thanks for coming in to see us here at the agency, Joe. I'm one of the agency representatives, and my name's Sally Baker.

JOE: Hi Sally. I think we spoke on the phone, didn't we?

SALLY: That's right, we did. So thank you for sending in your CV. We've had quite a careful look at it and I think we have two jobs that might be suitable for you.

JOE: OK.

SALLY: The first one is in a company based in North London. They're looking for an administrative assistant.

JOE: OK. What sort of company is it?

SALLY: They're called Home Solutions and they design and make **furniture. (Q1)**

JOE: Oh, I don't know much about that, but it sounds interesting.

SALLY: Yes, well as I said, they want someone in their office, and looking at your past experience it does look as if you fit quite a few of the requirements. So on your CV it appears you've done some data entry?

JOE: Yes.

SALLY: So that's one skill they want. Then they expect the person they appoint to attend **meetings (Q2)** and take notes there …

JOE: OK. I've done that before, yes.

SALLY: And you'd need to be able to cope with general admin.

JOE: Filling, and keeping records and so on? That should be OK. And in my last job I also had to manage the **diary. (Q3)**

SALLY: Excellent. That's something they want here too. I'd suggest you add it to your CV – I don't think you mentioned that, did you?

JOE: No.

SALLY: So as far as the requirements go, they want good computer skills, of course, and they particularly mention spreadsheets.

JOE: That should be fine.

SALLY: And interpersonal skills – which would be something they'd check with your references.

JOE: I think that should be OK, yes.

SALLY: Then they mention that they want someone who is careful and takes care with **details (Q4)** – just looking at your CV, I'd say you're probably alright there.

JOE: I think so, yes. Do they want any special experience?

SALLY: I think they wanted some experience of teleconferencing.

JOE: I've got three years' experience of that.

SALLY: Let's see, yes, good. In fact they're only asking for at least **one year (Q5)**, so that's great. So is that something that might interest you?

JOE: It is, yes. The only thing is, you said they were in North London so it would be quite a long commute for me.

SALLY: OK.

Test items

PART 1 Questions 1-5

Complete the notes below.

Write **ONE WORD AND/OR A NUMBER** for each answer.

Employment Agency: Possible Jobs

First Job

Administrative assistant in a company that produces **(1)** ..... (North London)

Responsibilities

- data entry
- go to **(2)** ..... and take notes
- general admin
- management of **(3)** .....

Requirements

- good computer skills including spreadsheets
- good interpersonal skills
- attention to **(4)** .....

Experience

- need a minimum of **(5)** ..... of experience of teleconferencing

Answer keys

1 furniture 2 meetings 3 diary 4 detail(s) 5 1 / one year

Listening Test A

Transcript

Part 1

AMBER: Hello William. This is Amber – you said to phone if I wanted to get more information about the job agency you mentioned. Is now a good time?

WILLIAM: Oh, hi Amber. Yes. Fine. So the agency I was talking about is called Bankside – they're based in Docklands – I can tell you the address now – 497 Eastside.

AMBER: OK, thanks. So is there anyone in particular I should speak to there?

WILLIAM: The agent I always deal with is called Becky Jamieson.

AMBER: Let me write that down – Becky …

WILLIAM: **Jamieson (Q1)** J-A-M-I-E-S-O-N.

AMBER: Do you have her direct line?

WILLIAM: Yes, it's in my contacts somewhere – right, here we are: 078 double 6, 510 triple 3. I wouldn't call her until the **afternoon (Q2)** if I were you – she's always really busy in the morning trying to fill last-minute vacancies. She's really helpful and friendly so I'm sure it would be worth getting in touch with her for an informal chat.

AMBER: It's mainly clerical and admin jobs they deal with, isn't it?

WILLIAM: That's right. I know you're hoping to find a full-time job in the media eventually – but Becky mostly recruits temporary staff for the finance sector – which will look good on your CV – and generally pays better too.

AMBER: Yeah – I'm just a bit worried because I don't have much office experience.

WILLIAM: I wouldn't worry. They'll probably start you as a receptionist, or something like that. So what's important for that kind of job isn't so much having business skills or knowing lots of different computer systems – it's **communication (Q3)** that really matters – so you'd be fine there. And you'll pick up office skills really quickly on the job. It's not that complicated.

AMBER: OK good. So how long do people generally need temporary staff for? It would be great if I could get something lasting at least a month.

WILLIAM: That shouldn't be too difficult. But you're more likely to be offered something for a **week (Q4)** at first, which might get extended. It's unusual to be sent somewhere for just a day or two.

AMBER: Right, I've heard the pay isn't too bad – better than working in a shop or a restaurant.

WILLIAM: Oh yes – definitely. The hourly rate is about **£10 (Q5)**, 11 if you're lucky.

AMBER: That's pretty good. I was only expecting to get eight or nine pounds an hour.

WILLIAM: Do you want me to tell you anything about the registration process?

AMBER: Yes, please. I know you have to have an interview.

WILLIAM: The interview usually takes about an hour and you should arrange that about a week in advance.

AMBER: I suppose I should dress smartly if it's for office work – I can probably borrow a **suit (Q6)** from Mum.

WILLIAM: Good idea. It's better to look too smart than too casual.

AMBER: Will I need to bring copies of my exam certificates or anything like that?

WILLIAM: No – they don't need to see those, I don't think.

AMBER: What about my **passport? (Q7)**

WILLIAM: Oh yes – they will ask to see that.

AMBER: OK.

WILLIAM: I wouldn't get stressed about the interview though. It's just a chance for them to build a relationship with you – so they can try and match you to a job which you'll like. So there are questions about **personality (Q8)** that they always ask candidates – fairly basic ones. And they probably won't ask anything too difficult like what your plans are for the future.

AMBER: Hope not.

WILLIAM: Anyway, there are lots of benefits to using an agency – for example, the interview will be useful because they'll give you **feedback (Q9)** on your performance so you can improve next time.

AMBER: And they'll have access to jobs which aren't advertised.

WILLIAM: Exactly – most temporary jobs aren't advertised.

AMBER: And I expect finding a temporary job this way takes a lot less **time (Q10)** – it's much easier than ringing up individual companies.

WILLIAM: Yes indeed. Well I think …

Part 2

Good morning. My name's Erica Matthews, and I'm the owner of Matthews Island Holidays, a company set up by my parents. Thank you for coming to this presentation, in which I hope to interest you in what we have to offer. We're a small, family-run company, and we believe in the importance of the personal touch, so we don't aim to compete with other companies on the number of customers. What we do is build on our **many years' experience – more than almost any other rail holiday company (Q11)** – to ensure we provide perfect holidays in a small number of destinations, which we've got to know extremely well.

I'll start with our six-day Isle of Man holiday. This is a fascinating island in the Irish Sea, with Wales to the south, England to the east, Scotland to the north and Northern Ireland to the west. Our holiday starts in **Heysham, where your tour manager will meet you (Q12)**, then you'll travel by ferry to the Isle of Man. Some people prefer to fly from Luton instead, and another popular option is to go by train to Liverpool and take a ferry from there.

You have five nights in the hotel, and the price covers five breakfasts and dinners, and **lunch on the three days when there are organised trips (Q13)**: day four is free, and most people have lunch in a café or restaurant in Douglas.

The price of the holiday includes the ferry to the Isle of Man, all travel on the island, the hotel, and the meals I've mentioned. Incidentally, we try to make booking our holidays as simple and fair as possible, so unlike with many companies, the price is the same whether you book six months in advance or at the last minute, and there's no supplement for single rooms in hotels. **If you make a booking then need to change the start date, for example because of illness, you're welcome to change to an alternative date or a different tour, for a small administrative fee. (Q14)**

OK, so what does the holiday consist of? Well, on day one you'll arrive in time for a short introduction by your tour manager, followed by dinner in the hotel. The dining room looks out at the **river (Q15)**, close to where it flows into the harbour, and there's usually plenty of activity going on.

On day two you'll take the coach to the small town of Peel, on the way calling in at the Tynwald Exhibition. The Isle of Man isn't part of the United Kingdom, and it has its own parliament, called Tynwald. It's claimed that this is the world's oldest parliament that's still functioning, and that it dates back to 979. However, the earliest surviving reference to it is from **1422 (Q16)**, so perhaps it isn't quite as old as it claims!

Day three we have a trip to the mountain Snaefell. This begins with a leisurely ride along the promenade in Douglas in a horse-drawn tram. Then you board an electric train which takes you to the fishing village of Laxey. From there it's an eight-kilometre ride in the Snaefell Mountain Railway to the **top (Q17)**. Lunch will be in the café, giving you spectacular views of the island.

Day four is free for you to explore, using the **pass (Q18)** which we'll give you. So you won't have to pay for travel on local transport, or for entrance to the island's heritage sites. Or you might just want to take it easy in Douglas and perhaps do a little light shopping.

The last full day, day five, is for some people the highlight of the holiday, with a ride on the **steam (Q19)** railway, from Douglas to Port Erin. After some time to explore, a coach will take you to the headland that overlooks the Calf of Man, a small island just off the coast. From there you continue to Castletown, which used to be the **capital (Q20)** of the Isle of Man, and its mediaeval castle.

And on day six it's back to the ferry – or the airport, if you flew to the island – and time to go home.

Now I'd like to tell you …

Part 3

RUTH: Ed, how are you getting on with the reading for our presentation next week?

ED: Well, OK, Ruth – but there's so much of it.

RUTH: I know, I hadn't realised birth order was such a popular area of research.

ED: But the stuff on birth order and personality is mostly unreliable. From what I've been reading a lot of claims about how your position in the family determines certain personality traits are just stereotypes, with no robust evidence to support them.

RUTH: OK, but that's an interesting point – we could start by outlining what previous research has shown. There are studies going back over a hundred years.

ED: Yeah – so we could just run through some of the typical traits. Like the consensus seems to be that oldest children are generally less well-adjusted because they never get over the arrival of a younger sibling.

RUTH: Right, but on a positive note, some studies claimed that **they were thought to be good at nurturing – certainly in the past when people had large families they would have been expected to look after the younger ones. (Q21)**

ED: There isn't such a clear picture for middle children – but one trait that a lot of the studies mention is that they are easier to get on with than older or younger siblings.

RUTH: **Generally eager to please and helpful (Q22)** – although that's certainly not accurate as far as my family goes – my middle brother was a nightmare – always causing fights and envious of whatever I had.

ED: As I said – none of this seems to relate to my own experience. I'm the youngest in my family and I don't recognise myself in any of the studies I've read about. I'm supposed to have been **a sociable and confident child who made friends easily (Q23)** – but I was actually terribly shy.

RUTH: Really? That's funny. There have been hundreds of studies on twins but mostly about nurture versus nature…

ED: There was one on personality, which said that a twin is likely to be **quite shy in social situations (Q24)** because they always have their twin around to depend on for support.

RUTH: My cousins were like that when they were small – they were only interested in each other and found it hard to engage with other kids. They're fine now though.

ED: Only children have had a really bad press – a lot of studies have branded them as **loners who think the world revolves around them (Q25)** because they've never had to fight for their parents' attention.

RUTH: That does seem a bit harsh. One category I hadn't considered before was children with much older siblings – a couple of studies mentioned that these children **grow up more quickly and are expected to do basic things for themselves – like getting dressed. (Q26)**

ED: I can see how that might be true – although I expect they're sometimes the exact opposite – playing the baby role and clamouring for special treatment.

RUTH: What was the problem with most of these studies, do you think?

ED: I think it was because in a lot of cases data was collected from only one sibling per family, who rated him or herself and his or her siblings at the same time.

RUTH: Mmm. Some of the old research into the relationship between birth order and academic achievement has been proved to be accurate though. Performances in intelligence tests decline slightly from the eldest child to his or her younger siblings. This has been proved in lots of recent studies.

ED: Yes. **Although what many of them didn't take into consideration was family size (Q27)**. The more siblings there are, the likelier the family is to have a low socioeconomic status – which can also account for differences between siblings in academic performance.

RUTH: The oldest boy might be given more opportunities than his younger sisters, for example.

ED: Exactly.

RUTH: But the main reason for the marginally higher academic performance of oldest children is quite surprising, I think. It's not only that they benefit intellectually from extra attention at a young age – which is what I would have expected. **It's that they benefit from being teachers for their younger siblings, by verbalising processes. (Q28)**

ED: Right, and this gives them status and confidence, which again contribute, in a small way, to better performance. So would you say sibling rivalry has been a useful thing for you?

RUTH: I think so – my younger brother was incredibly annoying and we fought a lot but I think this has made me a stronger person. **I know how to defend myself (Q29/Q30)**. We had some terrible arguments and I would have died rather than apologise to him – but **we had to put up with each other (Q29/Q30)** and most of the time we co-existed amicably enough.

ED: Yes, my situation was pretty similar. But I don't think having two older brothers made me any less selfish – I was never prepared to let me brothers use any of my stuff …

RUTH: That's perfectly normal, whereas …

Part 4

Today I'm going to talk about the eucalyptus tree. This is a very common tree here in Australia, where it's also sometimes called the gum tree. First I'm going to talk about why it's important, then I'm going to describe some problems it faces at present.

Right, well the eucalyptus tree is an important tree for lots of reasons. For example, it gives **shelter (Q31)** to creatures like birds and bats, and these and other species also depend on it for food, particularly the nectar from its flowers. So it supports biodiversity. It's useful to us humans too, because we can kill germs with a disinfectant made from **oil (Q32)** extracted from eucalyptus leaves.

The eucalyptus grows all over Australia and the trees can live for up to four hundred years. So it's alarming that all across the country, numbers of eucalyptus are falling because the trees are dying off prematurely. So what are the reasons for this?

One possible reason is disease. As far back as the 1970s the trees started getting a disease called Mundulla Yellows. The trees' leaves would gradually turn yellow, then the tree would die. It wasn't until 2004 that they found the cause of the problem was lime, or calcium hydroxide to give it its proper chemical name, which was being used in the construction of **roads (Q33)**. The lime was being washed away into the ground and affecting the roots of the eucalyptus trees nearby. What it was doing was preventing the trees from sucking up the iron they needed for healthy growth. When this was injected back into the affected trees, they immediately recovered.

But this problem only affected a relatively small number of trees. By 2000, huge numbers of eucalyptus were dying along Australia's East Coast, of a disease known as Bell-miner Associated Die-back. The bell-miner is a bird, and the disease seems to be common where there are high populations of bell-miners. Again it's the leaves of the trees that are affected. What happens is that **insects (Q34)** settle on the leaves and eat their way round them, destroying them as they go, and at the same time they secrete a solution which has sugar in it. The bell-miner birds really like this solution, and in order to get as much as possible, they keep away other creatures that might try to get it. So these birds and insects flourish at the expense of other species, and eventually so much damage is done to the leaves that the tree dies.

But experts say that trees can start looking sick before any sign of Bell-miner Associated Die-back. So it looks as if the problem might have another explanation. One possibility is that it's to do with the huge bushfires that we have in Australia. A theory proposed over 40 years ago by ecologist William Jackson is that the frequency of bushfires in a particular region affects the type of vegetation that grows there. If there are very frequent bushfires in a region, this encourages **grass (Q35)** to grow afterwards, while if the bushfires are rather less frequent, this results in the growth of eucalyptus forests.

So why is this? Why do fairly frequent bushfires actually support the growth of eucalyptus? Well, one reason is that the fire stops the growth of other species which would consume **water (Q36)** needed by eucalyptus trees. And there's another reason. If these other quick-growing species of bushes and plants are allowed to proliferate, they harm the eucalyptus in another way, by affecting the composition of the **soil (Q37)**, and removing nutrients from it. So some bushfires are actually essential for the eucalyptus to survive as long as they are not too frequent. In fact there's evidence that Australia's Indigenous people practised regular burning of bush land for thousands of years before the arrival of the Europeans.

But since Europeans arrived on the continent, the number of bushfires has been strictly controlled. Now scientists believe that this reduced frequency of bushfires to low levels had led to what's known as **'dry (Q38)** rainforest', which seems an odd name as usually we associate tropical rainforest with wet conditions. And what's special about this type of rainforest? Well, unlike tropical rainforest which is a rich ecosystem, this type of ecosystem is usually a **simple (Q39)** one. It has very thick, dense vegetation, but not much variety of species. The vegetation provides lots of shade, so one species that does find it ideal is the bell-miner bird, which builds its **nests (Q40)** in the undergrowth there. But again that's not helpful for the eucalyptus tree.

Test items

PART 1 Questions 1-10

Complete the notes below.

Write **ONE WORD AND/OR A NUMBER** for the answer.

Bankside Recruitment Agency

- Address of agency: 497 Eastside, Docklands
- Name of agent: Becky **(1)** .....
- Phone number: 07866 510333
- Best to call her in the **(2)** .....

Typical jobs

- Clerical and admin roles, mainly in the finance industry
- Must have good **(3)** ..... skills
- Jobs are usually for at least one **(4)** .....
- Pay is usually **(5)** £ ..... per hour

Registration process

- Wear a **(6)** ..... to the interview
- Must bring your **(7)** ..... to the interview
- They will ask questions about each applicant's **(8)** .....

Advantages of using an agency

- The **(9)** ..... you receive at interview will benefit you
- Will get access to vacancies which are not advertised
- Less **(10)** ..... is involved in applying for jobs

PART 1 Questions 11-14

Choose the correct letter A, B or C

11. According to the speakers, the company:
    A) has been in business for longer than most of its competitors.
    B) arranges holidays to more destinations than its competitors.
    C) has more customers than its competitors.

12. Where can customers meet the tour manager before travelling to the Isle of Man?
    A) Liverpool
    B) Heysham
    C) Luton

13. How many lunches are included in the price of the holiday?
    A) three
    B) four
    C) five

14. Customers have to pay extra for:
    A) guaranteeing themselves a larger room.
    B) booking at short notice.
    C) transferring to another date.

Questions 15-20

Complete the table below. Write **ONE WORD AND/OR A NUMBER** for each answer

Timetable for Isle of Man holiday		
	Activity	**Notes**
Day 1	Arrive	Introduction by manager. Hotel dining room has view of the **(15)** .....
Day 2	Tynwald Exhibition and Peel	Tynwald may have been founded in **(16)** ..... not 979.
Day 3	Trip to Snaefell	Travel along promenade in a tram; train to Laxey; train to the **(17)** ..... of Snaefell.
Day 4	Free day	Company provides a **(18)** ..... for local transport and heritage sites.
Day 5	Take the **(19)** ..... railway train from Douglas to Port Erin	Free time, then coach to Castletown – former **(20)** ..... has old castle.
Day 6	Leave	Leave the island by ferry or plane

Part 3 Questions 21-26

What did the findings of previous research claim about the personality traits a child is likely to have because of their position in the family?

Choose SIX answers from the box and match the correct answer, A-H, to Questions 21-26.

Position in family

21 the oldest child .....

22 a middle child .....

23 the youngest child .....

24 a twin .....

25 an only child .....

26 a child with much older siblings .....

Personality Traits

A outgoing
B selfish
C independent
D attention-seeking
E introverted
F co-operative
G caring
H competitive

Questions 27 and 28

Choose the correct letter A, B or C.

27 What do the speakers say about the evidence relating to birth order and academic success?

A) There is conflicting evidence about whether oldest children perform best in intelligence tests.

B) There is little doubt that birth order has less influence on academic achievement than socio-economic status.

C) Some studies have neglected to include important factors such as family size.

28 What does Ruth think is surprising about the difference in oldest children's academic performance?

A) It is mainly thanks to their roles as teachers for their younger siblings.

B) The advantages they have only lead to a slightly higher level of achievement.

C) The extra parental attention they receive at a young age makes little difference.

Questions 29 and 30

Choose TWO letters, A-E. Which TWO experiences of sibling rivalry do the speakers agree has been valuable for them?

A) learning to share
B) learning to stand up for oneself
C) learning to be a good loser
D) learning to be tolerant
E) learning to say sorry

PART 4 Questions 31-40.

Complete the notes below. Write **ONE WORD ONLY** for the answer.

The Eucalyptus Tree in Australia

Importance

- it provides **(31)** ..... and food for a wide range of species
- its leaves provide **(32)** ..... which is used to make a disinfectant

Reasons for present decline in number

A) Diseases

(i) 'Mundulla Yellows'

- Cause — lime used for making **(33)** ..... was absorbed
  — trees were unable to take in necessary iron through their roots

(ii) 'Bell-miner Associated Die-back'

- Cause — **(34)** *insects* feed on eucalyptus leaves
  — they secrete a substance containing sugar
  — bell-miner birds are attracted by this and keep away other species

B) Bushfires

William Jackson's theory:

- high-frequency bushfires have impact on vegetation, resulting in the growth of **(35)** .....
- mid-frequency bushfires result in the growth of eucalyptus forests, because they:
  —make more **(36)** ..... available to the trees
  —maintain the quality of the **(37)** .....
- low-frequency bushfires result in the growth of **(38)** ' ..... rainforest', which is:
  —a **(39)** ..... ecosystem
  —an ideal environment for the **(40)** ..... of the bell-miner

Answer keys

1 Jamieson
2 afternoon
3 communication
4 week
5 10/ten
6 suit
7 passport
8 personality
9 feedback
10 time
11 A
12 B
13 A
14 C
15 river
16 1422
17 top
18 pass
19 steam
20 capital
21 G
22 F
23 A
24 E
25 B
26 C
27 C
28 A
29&30 B, D
31 shelter
32 oil
33 roads
34 insects
35 grass(es)
36 water
37 soil
38 dry
39 simple
40 nest(s)

Listening Test B

Transcript

Part 1

TIM: Good morning. You're through to the tourist information office, Tim speaking. How can I help you?

JEAN: Oh hello. Could you give me some information about next month's festival, please? My family and I will be staying in the town that week.

TIM: Of course. Well it starts with a concert on the afternoon of the 17th.

JEAN: Oh I heard about that. The orchestra and singers come from the USA, don't they?

TIM: They're from Canada. They're very popular over there. They're going to perform a number of well-known pieces that will appeal to children as well as adults.

JEAN: That sounds good. My whole family are interested in music.

TIM: The next day, the 18th, there's a performance by a ballet company called **Eustatis. (Q1)**

JEAN: Sorry?

TIM: The name is spelt E-U-S-T-A-T-I-S. They appeared in last year's festival, and went down very well. Again, their program is designed for all ages.

JEAN: Good. I expect we'll go to that. I hope there's going to be a play during the festival, a comedy, ideally.

TIM: You're in luck! On the 19th and 20th a local amateur group are performing one written by a member of group. It's called Jemima. That'll be on in the town hall. They've already performed it two or three times. I haven't seen it myself, but the **review (Q2)** in the local paper was very good.

JEAN: And is it suitable for children?

TIM: Yes, in fact it's aimed more at children than at adults, so both performances are in the afternoon.

JEAN: And what about **dance (Q3)**? Will there by any performances?

TIM: Yes, also on the 20th, but in the evening. A professional company is putting on a show of modern pieces, with electronic music by young composers.

JEAN: Uh-huh.

TIM: The show is about how people communicate, or fail to communicate, with each other, so it's got the rather strange name, **Chat. (Q4)**

JEAN: I suppose that's because that's something we do both face-to-face and online.

TIM: That's right.

TIM: Now there are also some workshops and other activities. They'll all take place at least once every day, so everyone who wants to take part will have a chance.

JEAN: Good. We're particularly interested in cookery – you don't happen to have a cookery workshop, do you?

TIM: We certainly do. It's going to focus on how to make food part of a **healthy (Q5)** lifestyle, and it'll show that even sweet things like cakes can contain much less sugar than they usually do.

JEAN: That might be worth going to. We're trying to encourage our children to cook.

TIM: Another workshop is just for children, and that's on creating **posters (Q6)** to reflect the history of the town. The aim is to make children aware of how both the town and people's lives have changed over the centuries. The results will be exhibited in the community centre. Then the other workshop is in toy-making, and that's for adults only.

JEAN: Oh, why's that?

TIM: Because it involves carpentry – participants will be making toys out of **wood (Q7)**, so there'll be a lot of sharp chisels and other tools around.

JEAN: It makes sense to keep children away from it.

TIM: Exactly. Now let me tell you about some of the outdoor activities. There'll be supervised wild swimming …

JEAN: Wild swimming? What's that?

TIM: It just means swimming in natural waters, rather than a swimming pool.

JEAN: Oh OK. In a **lake (Q8)**, for instance.

TIM: Yes, there's a beautiful one just outside the town, and that'll be the venue for the swimming. There'll be lifeguards on duty, so it's suitable for all ages. And finally, there'll be a walk in some nearby woods every day. The leader is an expert on **insects (Q9)**. He'll show some that live in the woods, and how important they are for the environment. So there are going to be all sorts of different things to do during the festival.

JEAN: There certainly are.

TIM: If you'd like to read about how the preparations for the festival are going, the festival organiser is keeping a **blog (Q10)**. Just search online for the festival website, and you'll find it.

JEAN: Well, thank you very much for all the information.

TIM: You're welcome. Goodbye.

JEAN: Goodbye.

Part 2

WOMAN: I'm very pleased to welcome this evening's guest speaker, Mark Logan, who's going to tell us about the recent transformation of Minster Park. Over to you, Mark.

MARK: Thank you. I'm sure you're all familiar with Minster Park. It's been a feature of the city for well over a century, and has been the responsibility of the city council for most of that time. What perhaps isn't so well known is the origin of the park: **unlike many public parks that started in private ownership, as the garden of a large house, for instance, Minster was some waste land, which people living nearby started planting with flowers in 1892 (Q11).** It was unclear who actually owned the land, and this wasn't settled until 20 years later, when the council took possession of it.

You may have noticed the statue near one of the entrances. It's of Diane Gosforth, who played a key role in the history of the park. Once the council had become the legal owner, it planned to sell the land for housing. **Many local people (Q12)** wanted it to remain a place that everyone could go to, to enjoy the fresh air and natural environment – remember the park is in a densely populated residential area. **Diane Gosforth was one of those people, and she organised petitions and demonstrations (Q12)**, which eventually made the council change its mind about the future of the land.

Soon after this the First World War broke out, in 1914, and most of the park was dug up and **planted with vegetables (Q13)**, which were sold locally. At one stage the army considered taking it over for troop exercises and got as far as contacting the city council, then decided the park was too small to be of use. There were occasional public meetings during the war, in an area that had been retained as grass.

After the war, the park was turned back more or less to how it had been before 1914, and continued almost unchanged until recently. Plans for transforming it were drawn up at various times, most recently in 2013, though they were revised in 2015, before any work had started. **The changes finally got going in 2016 (Q14)**, and were finished on schedule last year.

OK, let me tell you about some of the changes that have been made – and some things that have been retained. If you look at this map, you'll see the familiar outline of the park, with the river forming the northern boundary, and a gate in each of the other three walls. The statue of Diane Gosforth has been moved: it used to be close to the south gate, but it's now **immediately to the north of the lily pond, almost in the centre of the park (Q15)**, which makes it much more visible.

There's a new area of wooden sculptures, which are **on the river bank, where the path from the east gate makes a sharp bend. (Q16)**

There are two areas that are particularly intended for children. The playground has been enlarged and improved, and that's **between the river and the path that leads from the pond to the river. (Q17)**

Then there's a new maze, a circular series of paths, separated by low hedges. That's **near the west gate – you go north from there towards the river and then turn left to reach it. (Q18)**

There have been tennis courts in the park for many years, and they've been doubled, from four to eight. They're still **in the south-west corner of the park, where there's a right-angle bend in the path. (Q19)**

Something else I'd like to mention is the new fitness area. This is **right next to the lily pond on the same side as the west gate. (Q20)**

Now, as you're all gardeners, I'm sure you'll like to hear about the plants that have been chosen for the park.

Part 3

CATHY: OK, Graham, so let's check we both know what we're supposed to be doing.

GRAHAM: OK.

CATHY: So, for the university's open day, we have to plan a display on British life and literature in the mid-19th century.

GRAHAM: That's right. But we'll have some people to help us find the materials and set it up, remember – for the moment, we just need to plan it.

CATHY: Good. So have you gathered who's expected to come and see the display? Is it for the people studying English, or students from other departments? I'm not clear about it.

GRAHAM: Nor me. That was how it used to be, but it didn't attract many people, so this year it's going to be part of an open day, to raise the university's profile. **It'll be publicised in the city, to encourage people to come and find out something of what does on here (Q21/Q22).** And it's included in the information that's sent to **people who are considering applying to study here next year. (Q21/Q22)**

CATHY: Presumably some current students and lecturers will come?

GRAHAM: I would imagine so, but we've been told to concentrate on the other categories of people.

CATHY: Right. We don't have to cover the whole range of 19th-century literature, do we?

GRAHAM: No, it's entirely up to us. I suggest just using Charles Dickens.

CATHY: That's a good idea. **Most people have heard of him, and have probably read some of his novels, or seen films based on them (Q23/Q24),** so that's a good lead-in to life in his time.

GRAHAM: Exactly. **And his novels show the awful conditions that most people had to live in, don't they: he wanted to shock people into doing something about it. (Q23/Q24)**

CATHY: Did he do any campaigning, other than writing?

GRAHAM: Yes, he campaigned for education and other social reforms, and gave talks, but I'm inclined to ignore that and focus on the novels.

CATHY: Yes, I agree.

CATHY: OK, so now shall we think about a topic linked to each novel?

GRAHAM: Yes. I've printed out a list of Dickens' novels in the order they were published, in the hope you'd agree to focus on him!

CATHY: You're lucky I did agree! Let's have a look. OK, the first was *The Pickwick Papers*, published in 1836. It was very successful when it came out, wasn't it, and was adapted for the theatre straight away.

GRAHAM: There's an interesting point, though, that there's **a character who keeps falling asleep, and that medical condition was named after the book – Pickwickian Syndrome. (Q25)**

CATHY: Oh, so why don't we use that as the topic, and include some quotations from the novel?

GRAHAM: Right, Next is *Oliver Twist*. There's a lot in the novel about poverty. But maybe something less obvious …

CATHY: Well Oliver is taught how to steal, isn't he? We could use that to illustrate the fact that **very few children went to school, particularly not poor children, so they learnt in other ways. (Q26)**

GRAHAM: Good idea. What's next?

CATHY: Maybe *Nicholas Nickleby*. Actually he taught in a really cruel school, didn't he?

GRAHAM: That's right. But there's also the **company of touring actors that Nicholas joins. We could do something on theatres and other amusements of the time. (Q27)** We don't want only the bad things, do we?

CATHY: OK.

GRAHAM: What about *Martin Chuzzlewit*? He goes to the USA, doesn't he?

CATHY: Yes, and **Dickens himself had been there a year before, and drew on his experience there in the novel. (Q28)**

GRAHAM: I wonder, though … The main theme is selfishness, so we could do something on social justice? No, too general, let's keep to your idea – I think it would work well.

CATHY: He wrote *Bleak House* next – that's my favourite of his novels.

GRAHAM: Yes, mine too. His satire of the legal system is pretty powerful.

CATHY: That's true, but think about Esther, t**he heroine. As a child she lives with someone she doesn't know is her aunt, who treats her very badly. Then she's very happy living with her guardian, and he puts her in charge of the household. And at the end she gets married and her guardian gives her and her husband a house, where of course they're very happy. (Q29)**

GRAHAM: Yes, I like that.

CATHY: What shall we take next? *Little Dorrit*? Old Mr Dorrit has been in a debtors' prison for years …

GRAHAM: So was Dicken's father, wasn't he?

CATHY: That's right.

GRAHAM: What about focusing on **the part when Mr Dorrit inherits a fortune, and he starts pretending he's always been rich? (Q30)**

CATHY: Good idea.

GRAHAM: OK, so next we need to think about what materials we want to illustrate each issue. That's going to be quite hard.

Part 4

I'm going to report on a case study of a program which has been set up to help rural populations in Mozambique, a largely agricultural country in South-East Africa.

The program worked with three communities in Chicualacuala district, near the Limpopo River. This is a dry and arid region, with unpredictable rainfall. Because of this, people in the area were unable to support themselves through agriculture and instead they used the forest as a means of providing themselves with an income, mainly by selling charcoal. However, this was not a sustainable way of living in the long term, as they were rapidly using up this resource.

To support agriculture in this dry region, the program focused primarily on making use of existing water resources from the Limpopo River by setting up systems of **irrigation (Q31)**, which would provide a dependable water supply for crops and animals. The program worked closely with the district government in order to find the best way of implementing this. The region already had one farmers' association, and it was decided

to set up two more of these. These associations planned and carried out activities including water management, livestock breeding and agriculture, and it was notable that in general, **women (Q32)** formed the majority of the workforce.

It was decided that in order to keep the crops safe from animals, both wild and domestic, special areas should be fenced off where the crops could be grown. The community was responsible for creating these fences, but the program provided the necessary **wire (Q33)** for making them.

Once the area had been fenced off, it could be cultivated. The land was dug, so that vegetables and cereals appropriate to the climate could be grown, and the program provided the necessary **seeds (Q34)** for this. The program also provided pumps so that water could be brought from the river in pipes to the fields. However, the labour was all provided by local people, and they also provided and put up the **posts (Q35)** that supported the fences around the fields.

Once the program had been set up, its development was monitored carefully. The farmers were able to grow enough produce not just for their own needs, but also to sell. However, getting the produce to places where it could be marketed was sometimes a problem, as the farmers did not have access to **transport (Q36)**, and this resulted in large amounts of produce, especially vegetables, being spoiled. This problem was discussed with the farmers' associations and it was decided that in order to prevent food from being spoiled, the farmers needed to learn techniques for its **preservation. (Q37)**

There was also an additional initiative that had not been originally planned, but which became a central feature of the program. This was when farmers started to dig holes for tanks in the fenced-off areas and to fill these with water and use them for breeding **fish (Q38)** – an important source of protein. After a time, another suggestion was made by local people which hadn't been part of the program's original proposal, but which was also adopted later on. They decided to try setting up colonies of **bees (Q39)**, which would provide honey both for their own consumption and to sell.

So what lessons can be learned from this program? First of all, it tells us that in dry, arid regions, if there is access to a reliable source of water, there is great potential for the development of agriculture. In Chicualacuala, there was a marked improvement in agricultural production, which improved food security and benefited local people by providing them with both food and income. However, it's important to set realistic timelines for each phase of the programme, especially for its **design (Q40)**, as mistakes made at this stage may be hard to correct later on.

The program demonstrates that sustainable development is possible in areas where …

Test items

Part 1 Question 1-10

Questions 1-4

Complete the table below.

Write **ONE WORD ONLY** for each answer.

Festival information		
Date	**Type of event**	**Details**
17th	a concert	performers from Canada
18th	a ballet	company called **(1)** .....
19th-20th (afternoon)	a play	type of play: a comedy called Jemima has had a good **(2)** .....
20th (evening)	a **(3)** ..... show	show is called **(4)** .....

Questions 5-10

Complete the notes below. Write **ONE WORD ONLY** for each answer.

Workshops

- Making **(5)** ..... food
- (children only) Making **(6)** .....
- (adults only) Making toys from **(7)** ..... using various tools

Outdoor activities

- Swimming in the **(8)** .....
- Walking in the woods, led by an expert on **(9)** .....

Seeing the festival organiser's **(10)** ..... for more information

PART 2 Questions 11-20

Questions 11-14

Choose the correct letter, A, B or C.

Minster Park

11 The park was originally established:

A) as an amenity provided by the city council.

B) as land belonging to a private house.

C) as a shared area set up by the local community.

12 Why is there a statue of Diane Gosforth in the park?

A) She was a resident who helped to lead a campaign.

B) She was a council member responsible for giving the public access.

C) She was a senior worker at the park for many years.

13 During the First World War, the park was mainly used for:

A) exercises by troops.

B) growing vegetables.

C) public meetings.

14 When did the physical transformation of the park begin?

A) 2013

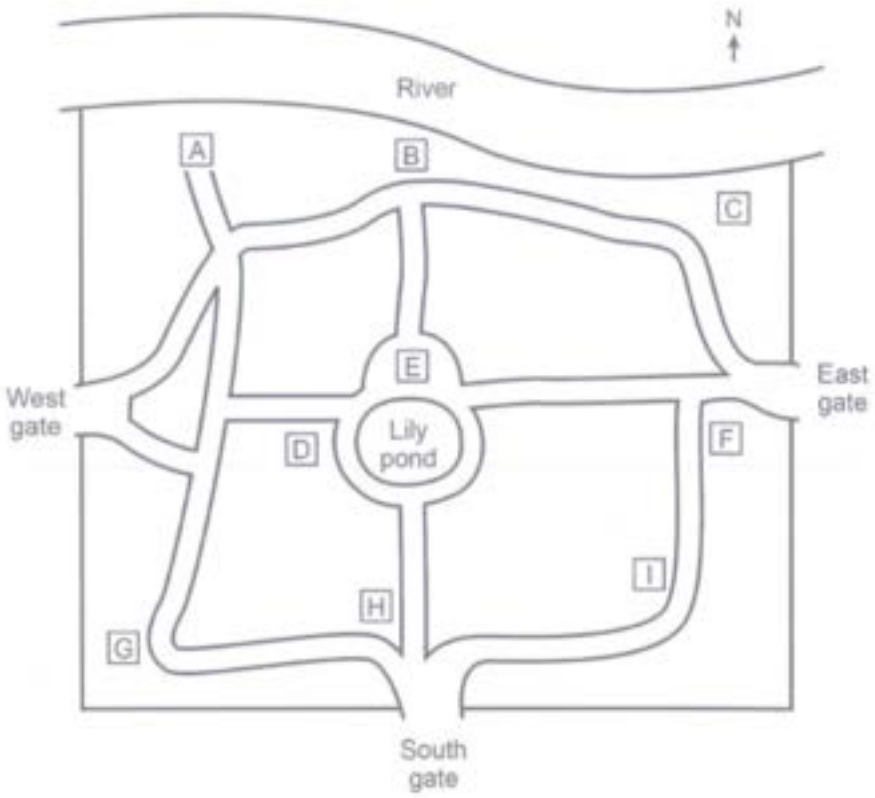
B) 2015

C) 2016

Questions 15-20

Label the map below. Write the correct letter, A-I, next to Questions 15-20.

Minister Park



15 statue of Diane Gosforth .....

16 wooden sculptures .....

17 playground .....

18 maze .....

19 tennis courts .....

20 fitness area .....

PART 3 Questions 21-30

Questions 21 and 22

Choose TWO letters, A-E.
Which TWO groups of people is the display primarily intended for?

A) students from the English department
B) residents of the local area
C) the university's teaching staff
D) potential new students
E) students from other departments

Questions 23 and 24

Choose TWO letters, A-E.
What are Cathy and Graham's TWO reasons for choosing the novelist Charles Dickens?

A) His speeches inspired others to try to improve society.
B) He used his publications to draw attention to social problems.
C) His novels are well-known now.
D) He was consulted on a number of social issues.
E) His reputation has changed in recent times.

Questions 25-30

What topic do Cathy and Graham choose to illustrate with each novel?
Choose SIX answers from the box and write the correct letter, A-H,
next to Questions 25-30.

Topics

A) poverty
B) education
C) Dickens's travels
D) entertainment
E) crime and the law
F) wealth
G) medicine
H) a woman's life

Novels by Dickens

25) *The Pickwick Papers*
26) *Oliver Twist*
27) *Nicholas Nickleby*
28) *Martin Chuzzlewit*
29) *Bleak House*
30) *Little Dorrit*

PART 4 Questions 31-40.

Complete the notes below.
Write ONE WORD ONLY for each answer.

Agricultural Program in Mozambique

How the program was organised:

- It focused on a dry and arid region in Chicualacuala district, near the Limpopo River.
- People depended on the forest to provide charcoal as a source of income.
- **(31)** ..... was seen as the main priority to ensure the supply of water.
- Most of the work organised by farmers' associations was done by **(32)** ..... .
- Fenced areas were created to keep animals away from crops.
- The program provided
  - —**(33)** ..... for the fences
  - —**(34)** ..... for suitable crops
  - —water pumps.
- The farmers provided
  - —labour
  - —**(35)** ..... for the fences on their land.

Further developments

- The marketing of produce was sometimes difficult due to lack of **(36)** ..... .
- Training was therefore provided in methods of food **(37)** .....
- Farmers made special places where **(38)** ..... could be kept.
- Local people later suggested keeping **(39)** ..... .

Evaluation and lessons learned

- Agricultural production increased, improving incomes and food security.
- Enough time must be allowed, particularly for the **(40)** ..... phase of the program.

Answer keys

1 Eustatis
2 review
3 dance
4 Chat
5 healthy
6 posters
7 wood
8 lake
9 insects
10 blog
11 C
12 A
13 B
14 C
15 E
16 C
17 B
18 A
19 G
20 D
21&22 B, D
23&24 B, C
25 G
26 B
27 D
28 C
29 H
30 F
31 Irrigation
32 women
33 wire(s)
34 seed(s)
35 posts
36 transport
37 preservation
38 fish(es)
39 bees
40 design

Appendix B: Questionnaire

Instructions: Thank you very much for participating in our research project on the impact of delivery modes in listening comprehension tests. The purpose of this survey is to elicit information about your overall test-taking experience with two IELTS listening tests and your perceptions and preferences related to the use of interactive videos. Your responses to this survey will be anonymised and aggregated with responses from other participants. It is expected that this survey will take about 10 minutes to complete.

Part I. Background information

1. ID code: _____
2. Age: _____
3. Gender: _ Female _ Male
   _ I prefer to self-identify as: _____
4. First (native) language(s): _____
5. Affiliation: _ University of Saskatchewan _ Western University
6. Current education status:
   _ Bachelor's student _ Master's student
   _ PhD student _ Other (please specify): _____
7. Primary academic discipline (your major) _____
8. Your email for receiving a digital gift card after the study _____
9. Number of years you have been studying English _____
10. Which of the following standardised English language tests have you taken most recently?
    _ IELTS Academic _ TOEFL iBT
    _ Duolingo English Test _ PTE Academic
    _ CAEL _ CELPIP
    _ Other (please specify): _____
11. Please provide information about the English language test you have taken most recently:
    Date of the test (month/year): _____
    Overall test score: _____
    Listening test score (if applicable): _____
    Reading test score (if applicable): _____
    Speaking test score (if applicable): _____
    Writing test score (if applicable): _____
11. How many years have you studied in institutions with English as a medium of instruction, including the university you are currently at? _____
12. How often have you encountered any interactive instructional videos (i.e., videos with embedded questions or activities) as a student?

    Never Very rarely Rarely
    Occasionally Frequently Very frequently

Part II. Perceptions and preferences of audio-only vs. video-based listening tests

13. Recall your experience with the two versions of the listening test you took in this study and rate the following statements on a scale from 1 to 6 (with 1 being *strongly disagree* and 6 being *strongly agree*).

Statement	1 = Strongly disagree 2 = Disagree 3 = Somewhat disagree 4 = Somewhat agree 5 = Agree 6 = Strongly agree
The interactive video-based listening test is more difficult than the audio-only listening test.	1 2 3 4 5 6
The animation used in the interactive video-based listening test facilitated my listening comprehension.	1 2 3 4 5 6
The gestures used by the animated characters helped me answer some questions on the video-based listening test.	1 2 3 4 5 6
The questions embedded in the interactive videos distracted me from listening.	1 2 3 4 5 6
The visuals in the animation helped me better understand the content of the video.	1 2 3 4 5 6
The animation in the interactive videos helped me predict what may happen next in the video.	1 2 3 4 5 6
The questions in the interactive video-based listening test were more difficult than the questions in the audio-only listening test.	1 2 3 4 5 6
I felt more confident completing the audio-only listening test than the interactive video-based listening test.	1 2 3 4 5 6
I watched the interactive videos during the listening test all the time.	1 2 3 4 5 6
It was easier for me to take notes during the audio-only listening test than during the video-based listening test.	1 2 3 4 5 6
The audio-only listening test provided a more accurate measurement of my listening comprehension skills.	1 2 3 4 5 6
Overall, I preferred the interactive video-based listening test.	1 2 3 4 5 6

14. Additional comments (if any): _____

Appendix C: Focus group interview guide

Number of participants:	Max 5 interviewees with a facilitator
Location:	A small meeting room on campus
Expected duration:	30 minutes (to be audio-recorded)

Procedure:

Step 1. Ice-breaking: Facilitator's self-introduction and brief explanation of the general procedure

Step 2. Guiding interview questions

1. How many times did you take IELTS in the past? Do you have a lot of experience preparing for it?
2. What were the differences between the audio-only and video-based listening tests that you took? Specifically,
   a. differences in the listening input (audio-only vs. video)
   b. differences in question types (pop-up questions vs. questions on paper)
   c. differences in the difficulty level?
3. When responding to the pop-up questions in the video-based listening test, what strategies did you use?
4. How were these strategies similar or different from the strategies you used for the audio-only test?
5. Which version of the listening test would you prefer as a test-taker: audio-only or interactive video-based?
6. Which version of the listening test would you prefer as a learner who needs to develop the L2 listening ability: audio-only or interactive video-based?
7. In your university-level courses, have you used any interactive video content? If yes, how would you evaluate the interactive video-based listening comprehension test, compared with instructional videos in your (online or in-person) university classes?
8. Did seeing whether you have answered each item correctly or incorrectly affect your performance on the video-based listening test? Please explain why or why not.
9. What do you see as the advantages of using interactive videos and pop-up questions for the assessment of listening?
10. What do you see as the disadvantages of using interactive videos and pop-up questions for the assessment of listening?
11. In your opinion, which delivery mode (i.e., audio-only vs. interactive video) may best assess test-takers' listening ability? Why?
12. If interactive videos were to be used in a listening comprehension test, what kind of concerns and suggestions would you like to share with us?

Step 3. Closing remarks and presentation of gift cards to participants

Appendix D: Coding scheme

1st Level Codes	2nd Level Codes	Definition	Western (n=32)	USask (n=33)
Experience_IELTS	No	P have no experience with IELTS	6	13
	Yes	P have experience with IELTS	26	21
Difference_input	V is distracting	Video is distracting	14	10
	Strategies_note-taking	P write down notes in video LT	10	8
	Not seeing ques in V	P cannot see the questions all the time in video LT	7	18
	Knowing when ques pop up is easier	P can know when the questions pop up	6	2
	Anxious about instant feedback	P feel anxious about instant feedback	5	0
	Time management	P can manage how long they spend on each question	4	3
	Preference to A	P prefer A LT	2	3
	Preference to V	P prefer V LT	5	2
	V allows to focus on segments of listening	P can focus on each segment of listening	2	1
	More focused in A	P are more focused on questions in audio LTs	5	0
	More focused in V	P are more focused on questions in video LTs	1	3
	No difference between A and V	P think there is no difference between audio and video LTs	1	1
	Getting more useful info in V	Videos provide useful information to participants to answer questions	8	4
	Difficulties in multitasking	P cannot listen, watch video, and take notes at the same time	4	0
	V is difficult	Video LT/video is difficult	4	0
	Visuals sometimes are not helpful	Visuals sometimes are not helpful for understanding the listening content	3	0
	Depending on ques	The difficulty level of video/audio depends on the question type (MCQ vs. Fill-in-the-blank)	2	2
	P cannot review answers in V	P cannot review their answers after answering the questions in video LT	2	2
	V requires short-term memory	P sometimes forget info in Video LT	1	0
	V is interesting	V format is more interesting.	1	2
	V is relaxing	V format is more relaxing and less stressful than A.	0	1

Differences_ questions	Knowing when ques pop up	P can know when the questions pop up	4	3
	A is harder	Audio LT is harder than video one	1	2
	P may miss ques in A	P may miss questions in audio LT	1	9
	P can have ques all the time in A	P can read questions all the time in A LT	10	0
	Note-taking	P tended to write down questions before watching videos	6	0
	Difficulties in multitasking	P cannot listen, watch video, and take notes at the same time	4	0
	P can drag and drop in V	P can drag and drop options in V LT	3	0
	Preference to V	P prefer video LT	3	0
	Depending on ques	The difficulty level of video/audio depends on the question type (MCQ vs. Fill-in-the-blank)	2	0
	Having instant feedback	P can have instant feedback in V LT	2	0
	P can read next ques in V	P can read next question in V LT	2	0
	P get lost in A	P sometimes get lost in A	2	0
	More focused in A	P feel more focused in A LT	1	0
	Cons of V LT	P concern about their spelling mistakes when using computers	2	0
	No difference	P don't see a difference in questions.	0	10
	Pros of V LT	P think that V LT has benefits	1	1
Differences_difficulty	A is easier	Audio LT is easier	16	7
	V is easier	Video LT is easier	9	10
	No difference	P cannot find the difference between two types of LT.	2	11
	Depending on ques	The difficulty level of video/audio depends on the question type (MCQ vs. Fill-in-the-blank)	4	6
Strategies_video	Note-taking	P need to take notes in video LT	28	24
	Focusing on dots	P pay more attention when the video is approaching to dots	9	6
	Reading the following ques	P read the following questions showed up with the current question	4	3
	Paying attention to V content	P just watch the videos and answer the questions	6	0
	Using pause time to think of ques	P use pausing time to think of questions	1	0
	Using vocabulary knowledge	P recall their previous lexical knowledge about the topic of videos while watching videos	1	1
	Finding keywords	P try to find keywords in each question.	0	3
	Focusing on audio only	P didn't watch a video and focused on audio	0	2
	No particular strategy	P didn't have a particular strategy.	0	3
	POS prediction	P tried to predict POS of an answer.	0	2
	Remembering the questions	P tries to remember all the questions.	0	4

Strategies_ comparison	Taking less notes in A	P take less notes in audio LT.	13	9
	Highlighting and circling keywords in A	P can highlight and circle keywords in audio LT.	9	12
	Instant feedback can confirm answers	Instant feedback can confirm participants' answers	3	0
	Reading ques while listening in A	P can read questions while listening in audio LT.	2	12
	Previewing ques before listening	P preview questions before audio is playing in A LT	6	0
	Strategies are similar to V	P use the similar strategies in A LT.	5	0
	No particular strategy	P has no particular strategy.	4	0
	Imagining the scenario in A	P uses extralinguistic knowledge to predict an answer.	1	0
	POS prediction	Participants tried to predict POS of an answer.	0	5
Preference_test-taker	Audio	P prefers A LT as a test-taker	25	16
	Hybrid	P prefers having the video test with printed questions or all the questions shown on the screen	0	10
	Video	P prefers V LT as a test-taker	5	11
	Depending on language proficiency	P thinks that their preference depends on their language proficiency	1	0
	Both	Both test options work for the P	1	0
Preference_learner	Audio	As English learners, participants prefer audio format	3	4
	Both	P think that both test formats are good for English learning purposes	3	15
	Video	As English learners, participants prefer video format	23	9
	Depending on ques	P preference of video/audio depends on the question type (some better in video, some in audio)	1	0
	Depending on language proficiency	P thinks that learners' preference depends on their language proficiency	2	0
Experience_video	No experience	Participants had never encountered interactive videos before	13	19
	Some experience_university level	Participants had encountered interactive videos before in university classes	5	6
	Some experience_non-university level	Participants had encountered interactive videos before in non-university settings (job training etc.)	3	5
	Showing as an example	Instructors use interactive video as an example in their CALL course.	9	0
Effect_feedback	Mixed	Receiving instant feedback may have both negative and positive impact on participants' performance depending on how they answered (correctly or incorrectly).	5	6
	Negative	Receiving instant feedback has a negative impact on participants' performance.	19	20
	Neutral	Receiving instant feedback has no impact on participants' performance.	4	4
	Positive	Receiving instant feedback has a positive impact on participants' performance.	4	8

Advantages_video	Computers for completing and rating tests	P see advantages in using computers for the test	1	1
	Getting more useful info	Videos can provide clues for a better overall understanding of the audio.	20	16
	Knowing when ques pop up is easy	The question identifiers (dots) help during the test.	7	15
	More flexible time management	P think that time management is easier during V LT	3	2
	More focused on listening	The video format helps to focus on listening better.	4	6
	Seeing instant feedback	Instant feedback after each question is helpful.	4	4
	V is authentic	V is closer to real-life settings.	1	3
	V is interesting	V format is more interesting.	3	4
	V is more relaxing	V format is more relaxing and less stressful than A.	1	6
	Showing a couple of ques in V	The ability to see the next question(-s) while answering the current one.	1	0
	Suitable for formative assessment		1	0
Disadvantages_video	Difficulties in multitasking	The video format requires multitasking which can cause difficulties.	6	1
	Felt tired	The video format is tiring.	2	0
	Not seeing ques in V	Not seeing all the questions during the video test makes it harder.	8	19
	Seeing instant feedback	Instant feedback after each question is distracting.	4	1
	Taking too many notes	The video format makes participants take too many notes during the test.	4	3
	Technical issues	Technical issues and network malfunction are possible during V.	2	2
	V is distracting	The videos distract participants from listening.	12	15
	P cannot review answers	P didn't have a chance to review their answers.	1	4
	Adapting new test taking strategies	P have to adapt new test taking strategies in V LT	2	0
	V requires short term memory	V LT require short term memory	1	1
	Animation can be improved	P think that animation has to be improved	0	1
	V gives too many clues	P think that V is too helpful for the test-takers.	0	2
Effectiveness_mode	Audio	P think that the audio format is more effective for listening skills assessment.	11	5
	Depending on info perception A vs. V	P think that the effectiveness of each mode depends on what type of input a test-taker perceives better (auditory vs. visual).	1	7
	Depending on language proficiency	P think that the effectiveness of each mode depends on the listening proficiency of a test-taker.	1	6
	Not sure	P are not sure if any of the formats is better than another for listening skills assessment purposes.	2	5
	Video	P think that the video format is more effective for listening skills assessment.	12	16
	Depending on what to be tested	The effectiveness of each mode depends on what is being tested.	5	0
	Depending on memory	P think that effectiveness depends on the test-taker's overall ability to memorise.	0	4

Suggestions_video	Access to ques	P provide suggestions on improving access to the questions during the video-based test.	22	18
	Instant feedback	P provide suggestions on instant feedback.	9	8
	Time management	P provide suggestions on time management during V	3	2
	Video content	P provide suggestions on how video content can be changed/improved.	16	1
	Technical issues_computer use	P' suggestions regarding possible tech issues.	3	0
	Concerns about V LT preparation	P' concerns about V LT preparation	0	2
	Difficulty level	P provide their opinion on the current difficulty level and the possible adjustments.	0	2
	Reviewing answers	P suggest adding the possibility of reviewing the answers after each part.	0	6
	V as an option	P provide their opinion on where/in what parts of the test V would be better.	0	3