Article

LANGUAGE TESTING

Language Testing
2015, Vol. 32(4) 463–483

DOI: 10.1177/0265532214562099
ltj.sagepub.com
SAGE

The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos

Ruslan Suvorov
University of Hawai'i at Mānoa, USA

Abstract

Investigating how visuals affect test takers' performance on video-based L2 listening tests has been the focus of many recent studies. While most existing research has been based on test scores and self-reported verbal data, few studies have examined test takers' viewing behavior (Ockey, 2007; Wagner, 2007, 2010a). To address this gap, in the present study I employ eye-tracking technology to record the eye movements of 33 test takers during the Video-based Academic Listening Test (VALT). Specifically, I aim to explore test takers' oculomotor engagement with two types of videos – context videos and content videos – from the VALT, and the relationship between the test takers' viewing behavior and test performance. Eye-tracking measures comprising fixation rate, dwell rate, and the total dwell time for context and content videos were compared using paired-samples *t*-tests. Additionally, each measure was correlated with test scores for items associated with each video type. Results revealed statistically significant differences between fixation rates and between total dwell time values, but no difference between the dwell rates for context and content videos. No statistically significant relationship was found between the three eye-tracking measures and the test scores. Directions for future research on video-based L2 listening assessment are discussed.

Keywords

Content video, context video, eye tracking, L2 listening assessment, video-based listening test, visuals

Corresponding author:
Ruslan Suvorov, Center for Language and Technology, University of Hawai'i at Mānoa, 1890 East-West Rd, Moore Hall 256, Honolulu, HI 96822, USA.
Email: rsuvorov@hawaii.edu

Visuals are widely considered to be an important aspect of listening because in most target language use situations listeners integrate information both from the aural modality and from the visual modality (Ockey, 2007; Picou, Ricketts, & Hornsby, 2011; Sueyoshi & Hardison, 2005; Wagner, 2007, 2010b). In this study, visuals are defined as multimodal texts (such as videos) constructed to carry semantic information that is perceptible to the human ocular system (i.e., vision). In applied linguistics, such information is traditionally referred to as "visual information" and is differentiated from "auditory information" that is perceived by the human auditory system (i.e., hearing). Although L2 listening has traditionally been assessed using audio-only tests (Brindley, 1998; Buck, 2001), the implementation of visuals in L2 listening assessment has been steadily garnering attention, as evident from the growing number of publications on this topic (e.g., Batty, 2014; Coniam, 2001; Cubilo & Winke, 2013; Ginther, 2002; Londe, 2009; Ockey, 2007; Suvorov, 2009; Wagner, 2002, 2007, 2008, 2010a, 2010b, 2013). The bulk of these studies have focused on analyzing test takers' scores in connection with self-reported data to investigate whether visuals make any difference in performance on video-mediated L2 listening tests, as opposed to the test takers' performance on audio-only tests. Surprisingly, however, little attention has been directed towards exploring how test takers *interact* with visuals in such tests, with an exception of Ockey's (2007) and Wagner's (2007, 2010a) studies, in which video recordings of the participants taking the tests were used to measure the amount of time they had made eye contact with the screen.

Furthermore, to my knowledge, none of the existing studies has explored how different types of videos – such as content videos and context videos (Bejar, Douglas, Jamieson, Nissan, & Turner, 2000; Ginther, 2002) – would affect test takers' performance on video-mediated L2 listening tests. To address this gap in research, in the present study I utilized an eye-tracking system to investigate how L2 learners of English taking a Video-based Academic Listening Test (VALT) interacted with context videos and content videos from the test. Drawing on the eye-tracking data, in this study I also examine the relationship between the L2 learners' test scores and the measures of their oculomotor engagement with the two types of video used in the test.

Background

The construct of L2 listening

The construct of L2 listening is difficult to describe and define (Batty, 2014). This difficulty is caused by the complexity of processes and factors involved in listening (Bloomfield et al., 2010; Rubin, 1994), as well as a multitude of definitions of this language skill (Glenn, 1989; Rubin, 1995; Wolvin & Coakley, 1996). Moreover, developing a construct definition is complicated by the existence of different models of listening comprehension (Bejar et al., 2000; Flowerdew & Miller, 2010; Friederici, 2002; Nagle & Sanders, 1986; Wolvin & Coakley, 1996), with Wolvin and Coakley's (1996) model appearing to be the only one that takes into account both the acoustic signal *and* the visual input when describing the process of listening comprehension.

Although the ability to process visual input is generally recognized to be part of L2 listening proficiency (Baltova, 1994; Rost, 2011; Rubin, 1995; Taylor, 2013; Weir, 2005), there appears to be no consensus as to whether this ability should be included in the construct of L2 listening measured by L2 listening tests. Some researchers assert that visuals should be used in L2 listening tests to avoid construct underrepresentation (Wagner, 2008; 2010b) and increase the authenticity of L2 listening assessment (Ockey, 2007), although Ockey (2007) advised that the decision to include or exclude visuals from the L2 listening construct should depend on the type of visuals used in the test. Buck (2001), on the other hand, argued against the inclusion of visuals in the construct of L2 listening because "we are usually interested in the test-takers' language ability, rather than the ability to understand subtle visual information" (Buck, 2001, p. 172) that can give unfair advantage to those test takers who are particularly good at utilizing non-verbal information. The poor understanding of how exactly different aspects of non-verbal information affect test takers' scores on visually enhanced L2 listening tests should, according to Batty (2014), warrant "the definition of a new construct of 'visual listening comprehension'" (p. 16).

Visuals in L2 listening assessment

According to recent theorizing by language assessment researchers, there are two main types of visuals that can be incorporated in L2 assessment tasks: content visuals and context visuals (Bejar et al., 2000; Ginther, 2002; Ockey, 2007; Suvorov, 2009). In *content visuals*, information presented through the visual channels is semantically relevant to the verbally presented information (Ginther, 2002). *Context*, or situation, *visuals* provide information about the context of the spoken discourse, such as the setting, the speaker's role, or the text type (Bejar et al., 2000; Ginther, 2002). In addition to the context–content distinction, visuals used in L2 listening tasks can also be classified according to the mode of delivery and format, that is, (a) into single still images, (b) a series of still images, and (c) videos (Ockey, 2007). Whether or not this simple classification system of visuals for L2 listening assessment will stand the test of time, it has been the starting point for research on expanding the construct of listening (to include visuals). Thus, the distinctions and classifications are important to understand fully the way in which researchers have approached this recent expansion.

A growing number of studies have investigated how the presence of the visual medium affects test takers' performance on visually enhanced L2 listening tests, with the bulk of research in this area focusing on the video-mediated assessment of L2 listening (Batty, 2014; Coniam, 2001; Cubilo & Winke, 2013; Gruba, 1993, 1999; Londe, 2009; Progosh, 1996; Suvorov, 2009; Wagner, 2007, 2008, 2010a, 2010b, 2013). Most of these studies compared test takers' performance on video-based L2 listening tests with their performance on audio-only tests, with mixed results. While some researchers have found that the use of videos could help L2 listeners achieve higher test scores (Wagner, 2010b, 2013), the results of other studies have demonstrated that visuals have either no effect (Batty, 2014; Coniam, 2001; Gruba, 1993; Londe, 2009) or a debilitating effect on L2 test takers' performance (Suvorov, 2009).

Differing results can be attributed to differing methodologies and procedures, noncomparable stimuli used in assessment instruments, and different groups of participants used in the studies (Wagner, 2010b), which makes the interpretation of research in this area a challenge (Batty, 2014). Research on visuals in L2 listening assessment has primarily been concerned with the video–audio distinction, whereas the content–context distinction has been disregarded, with the exception of Ginther's (2002) study. Although not all researchers reported the types of videos used in their L2 listening assessment instruments, it appears that context videos have predominated in the testing scene (Batty, 2014; Coniam, 2001; Cubilo & Winke, 2013; Londe, 2009; Suvorov, 2009; Wagner, 2007, 2008, 2010a, 2010b, 2013). It is interesting to note that, to my knowledge, no research has directly compared the effects of context and content videos on L2 learners' listening test performance.

The prevalence of comparative studies appears to suggest the pervasiveness of the following assumption: If there is a statistically significant difference between test takers' scores on a video-mediated L2 listening test and the same test takers' scores on an audio-only L2 test, then this difference can be attributed to the effect of the visual input. The main problem with this assumption, however, is that it does not take into consideration L2 test takers' viewing behavior and their interaction with the visual stimuli during the test. Since test takers are not forced to watch a screen during the visually enhanced assessment of their L2 listening skills, they vary in the extent to which they interact with the visual input and utilize visual information from the test (Ockey, 2007; Wagner, 2007, 2010a), with some of them closing their eyes and not looking at the visuals (Cubilo & Winke, 2013) or watching the visuals to a very limited extent, as reported in previous studies (e.g., Brett, 1997; Coniam, 2001; Gruba, 1993; Ockey, 2007). Consequently, for those test takers who do not watch videos – or barely watch them – but who obtain scores that are statistically significantly different from their scores on the audio-only version of the listening test, the difference between the scores cannot be attributed to the effect of visual information. Wagner (2007) acknowledged this issue and suggested that it would be "premature to advocate for or argue against the use of video texts on L2 listening tests" (p. 69) without exploring the extent to which test takers actually utilize the visual information and the amount of time they orient towards looking at the screen.

Only two researchers have attempted to explicitly investigate L2 test takers' viewing behavior. Ockey (2007) and Wagner (2007, 2010a) used a video camera to record their participants and to measure the amount of time they spent watching the screen during an L2 listening test with visuals. Wagner (2007, 2010a) used a video-based listening test with context videos, whereas the listening tests in Ockey's (2007) study were enhanced by context images and context videos. Both researchers discovered significant individual differences among their participants in terms of the amount of time they looked at the screen during the test. The average viewing rate also varied: While the 36 participants in Wagner's (2007) study spent 69% of the time looking at the screen, the 56 participants in his 2010 study watched the videos only 47.9% of the time.

Although videotaping can supply some information about test takers' gaze, such data can give only a "crude indication" (Boraston & Blakemore, 2007, p. 894) of the test takers' gaze duration and gaze direction. More importantly, the data from video recordings can generally yield information about *how long* the participants look at the screen, not *what exactly* they look at, *how long* they focus on specific elements of the video, or *why* they look at them. More specialized technology such as an eye-tracking system can provide much more precise and detailed data that include not only the information about the test takers' gaze duration and gaze direction, but also about their eye movements and oculomotor engagement with different elements of the visual input from an L2 listening test.

Therefore, for the construct of L2 listening to be expanded so that it would include the ability to understand and use visual information (Ockey, 2007; Wagner, 2006, 2008), it is essential to collect new types of validity evidence derived not only from L2 test takers' scores, but also from their viewing behavior during the video-mediated L2 listening assessment.

Eye tracking in L2 assessment research

Eye-tracking research allows an investigator to record and examine the eye movements of participants. Such research requires an eye-tracker, which is an electronic device capable of recording eye movements, and software for recording and analyzing the data obtained with the help of such a device (Duchowski, 2007; Holmqvist et al., 2011). Even though the past decade has witnessed a surge of eye-tracking studies in various disciplines, including psycholinguistics, especially with the focus on reading, the use of eye tracking in research on L2 assessment appears to be inchoate.

The first – and still only – eye-tracking study published in *Language Testing* was that of Bax (2013) who investigated cognitive processing of test takers completing onscreen reading items from International English Language Testing System (IELTS). The use of eye-tracking data from 38 Malaysian undergraduates allowed Bax (2013) to discover differences between proficient and less proficient test takers in terms of their ability to read expeditiously and their focus on specific aspects of reading prompts and test items. In a related study, Bax and Weir (2012) employed an eye-tracking system to explore the extent to which items from a computer-based CAE Reading test elicited test takers' cognitive processes that were appropriate and necessary for advanced-level academic reading. Specifically, the analysis of eye-tracking data consisting of 28 eye-movement recordings from ten participants revealed that the test items were capable of eliciting a range of cognitive processes in reading, as well as different levels of processing, and these findings were in line with how these items were expected to perform.

While the researchers of the aforementioned studies have implemented eye-tracking technology to explore test takers' cognitive processes during their (the test takers') interaction with text-based L2 reading tasks, there appears to be no eye-tracking research investigating test takers' viewing behavior during a video-based test of L2 listening comprehension (also mentioned by Cubilo & Winke, 2013). With the present study I aim to address this gap.

Research questions

This study is guided by the following two research questions:

RQ1: To what extent do L2 test takers watch context videos differently from content videos in the Video-based Academic Listening Test (VALT), as indicated by eye-tracking measures?

RQ2: To what extent do L2 test takers' viewing patterns, as indicated by eye-tracking measures, correlate with their scores on the context-video subtest and on the content-video subtest?

Methodology

Research design

For this study I used a within-subjects experimental design. The independent variable was video type (i.e., context video versus content video). The dependent variables were test takers' scores for items associated with context videos and with content videos, and test takers' eye movements during the VALT represented by three eye-tracking measures (i.e., fixation rate, dwell rate, and the percentage of total dwell time), which are defined below in the data analysis section.

Participants

The participants were 33 international undergraduate and graduate students at a large public university in the Midwest of the United States. Out of these, 25 participants were undergraduate and graduate students enrolled in an English-as-a-second-language (ESL) listening course that is mandatory for all incoming international students who have not passed the in-house English Placement Test and have not achieved the required TOEFL score (640 for PBT, 270 for CBT, or 105 for iBT). The remaining eight participants were graduate students in an applied linguistics program at the same university and were more advanced non-native speakers of English with TOEFL iBT scores of 111 or above. Thirteen were male and 20 were female, with most being native speakers of Mandarin Chinese. Their ages were between 18 and 35, with one third of the participants being 18 years old.

Materials and instruments

The Video-based Academic Listening Test (VALT) I used in this study was developed as part of my dissertation project. As defined in test specifications, the purpose of the VALT was to measure L2 learners' ability to process and understand verbal and visual information from academic lectures enhanced by context and content videos.

The VALT consisted of six short video prompts and 30 multiple-choice questions. The video prompts were taken from YouTube videos of real-life university lectures and varied in length from 188 to 234 seconds. The videos came from six different disciplines and covered topics such as neurons, exoplanets, Enlightenment, rent control, p-functions, and mushrooms. Three prompts were context videos (see Figure 1) and the other three prompts were content videos (see Figure 2). Each video prompt was followed by five four-option, multiple-choice questions asking about the main idea or various details pertaining to the prompt.

Table 1 summarizes the main characteristics of context videos and content videos selected for the study. As evident from the table, Videos 1, 3, and 5 labeled as "context"

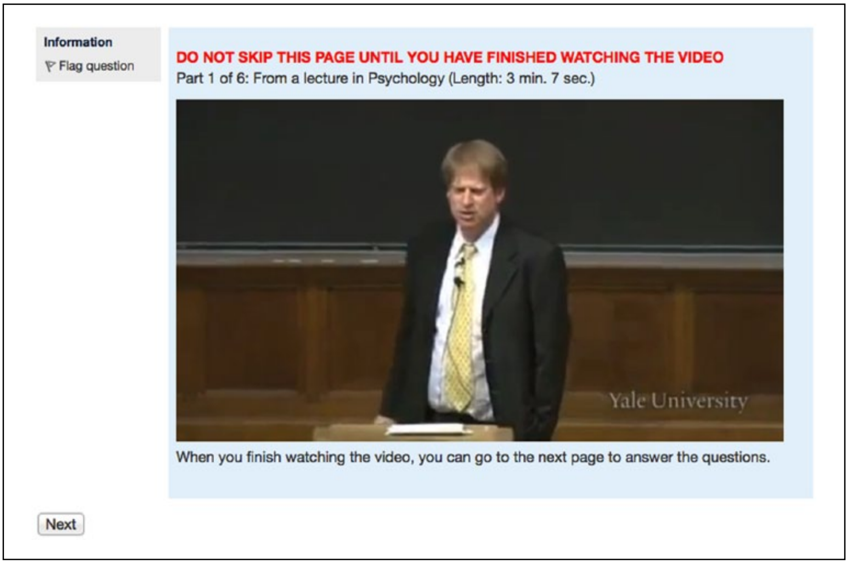

Figure 1. An example of a context video from the VALT (printed with permission from Dr. Paul Bloom, Yale University).

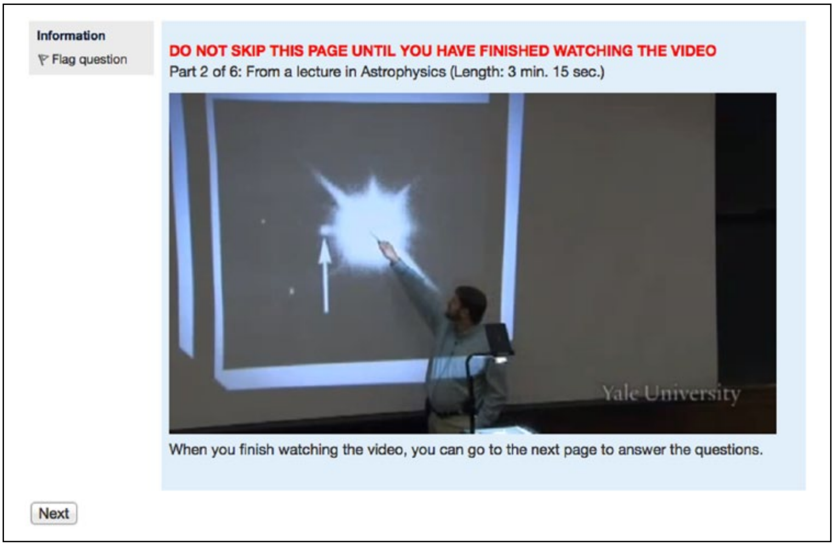

Figure 2. An example of a content video from the VALT (printed with permission from Dr. Charles Bailyn, Yale University).

showed only the speaker and the front of the classroom with a podium, a table, or a board/ screen. Context videos did not contain any visual aids (such as graphs or images), except

Table 1. Summary of the characteristics of context videos and content videos in the VALT.

Video type	Topic of video in the VALT	Main characteristics
Context	Video 1: Neurons	Professor giving a lecture behind a podium or while walking in front of the empty board; no visual aids
	Video 3: Enlightenment	Professor giving a lecture behind a podium with a picture of John Locke projected on the screen
	Video 5: P-functions	Professor giving a lecture while sitting on a table; no visual aids
Content	Video 2: Exoplanets	Professor projecting a photograph of an exoplanet, and pointing to different elements of the photograph during the lecture
	Video 4: Rent control	Professor giving a lecture behind a podium, and showing slides with a graph and a floor map
	Video 6: Mushrooms	Professor showing a mushroom, and drawing the structure of a mushroom on the board during the lecture

for Video 3 that showed a slide with a picture of John Locke. Since the professor did not refer to the picture and did not allude to John Locke during the lecture (and thus the visual information in the slide was not semantically relevant to the verbally presented information), Video 3 was labeled as "context" following Ginther's (2002) definition of context visuals. Meanwhile, the three videos labeled as "content" displayed not only the speaker, but also visual aids such as a photograph, a drawing, and PowerPoint slides. The visual aids in these videos were directly mentioned and referred to by the speakers, thereby making the visual information semantically relevant to the information presented verbally – and making Videos 2, 4, and 6 fall under Ginther's (2002) definition of content visuals.

I developed the 45-minute VALT using the Quiz module in the Moodle course management system (version 2.2), and test takers took the test in the Moodle platform as well. Each video prompt was played automatically one time. Owing to the set-up restrictions of the Quiz module in Moodle, the actual video occupied only about a quarter of the computer screen, with the rest of the screen showing the instructions, the "Next" button, and the user's log-in status. After watching each video, test takers saw one question at a time. Each question was presented in a written format on a computer screen. Test takers had no control over the browser or video play controls; they could only choose an answer for each question and move forward by clicking on the "Next" button (see Figure 3).

The overall test structure is represented in Figure 4. Detailed information about the validation process for the VALT can be found in Suvorov (2013).

I created and administered a post-test questionnaire in Moodle to collect background information about the participants. Seven open-ended questions asked about the participants' age, gender, L1, home country, current student status, major, and the number of years they had spent learning English.

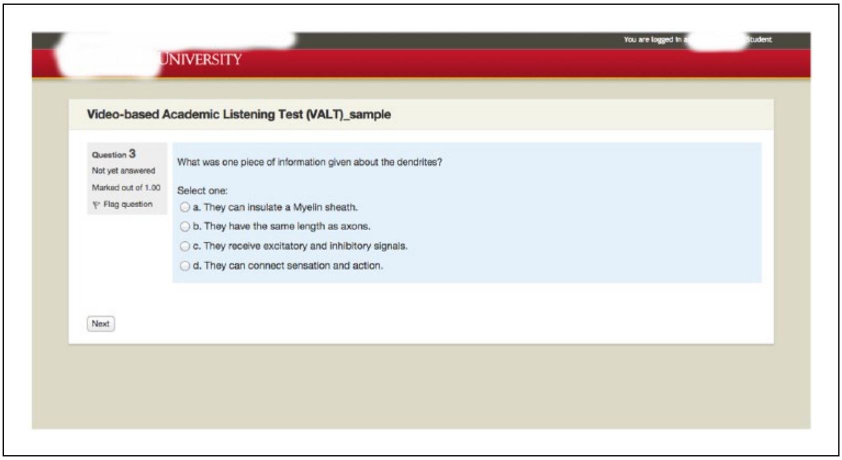

Figure 3. An example of a multiple-choice question from the VALT.

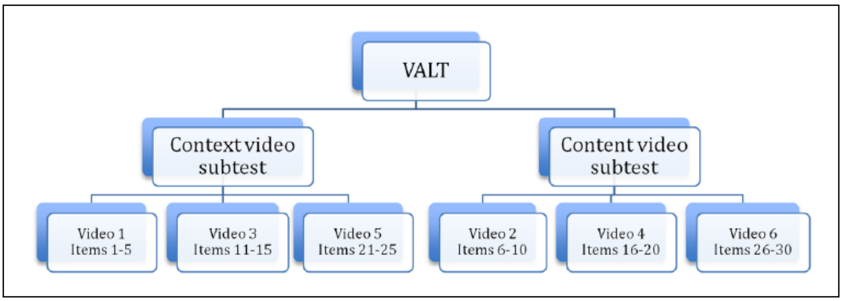

Figure 4. Overall structure of the VALT.

I used a remote eye-tracking system called the EyeTech Vision Tracker 2 (0.5° accuracy, 80 fps data sampling rate, 65–100 cm operating range, 1680 × 1050 display). The eye-tracker was physically connected to a computer display and ran on an iMac station (27 inches, 3.7 GHz) using Windows 7 64-bit OS. The display was also equipped with a web camera Logitech Webcam Pro 9000. In addition, a second display was used to monitor the data collection process. The eye-tracking data were recorded and processed using Attention Tool's Usability Module (version 4.8), which is an eye-tracking software application developed by iMotions for market research, scientific research, and website usability (iMotions, 2013a). I used Dynamic Media Module, an add-on module in Attention Tool for analyzing dynamic media such as videos, for the subsequent analysis of eye-tracking data (iMotions, 2013b).

Data collection

After obtaining an approval from the Institutional Review Board (IRB) to conduct the study, I recruited the participants among students enrolled in two sections of the ESL listening course and international graduate students in the applied linguistics program. Each partook in a 2-hour data collection session and received a small monetary compensation (25 US dollars).

I collected data with one participant at a time in an eye-tracking laboratory. Upon arriving in the laboratory, each participant received a headset, a pen, several sheets of paper for note-taking, and instructions on what he or she would need to do during the study. The participant was then seated approximately 65 cm from the computer screen with a connected eye-tracker, and asked to complete a 16-point eye calibration to ensure the accuracy of the recorded eye movements. After completing the calibration, I asked each participant to log into Moodle and proceed to take the VALT while his or her eye movements were being recorded. Owing to the test set-up in Moodle, the participants did not have an option to preview questions before watching each video clip. The participants were allowed to take notes on paper during the test. Finally, I asked each participant to complete the post-test questionnaire in Moodle.

Data analysis

Research question 1. To answer the first research question, I used the recordings of the participants' eye movements during the VALT to calculate three eye-tracking measures for content videos and context videos, which I subsequently compared using a paired-samples *t*-test. The analysis of the eye-tracking data recorded by Attention Tool consisted of three steps: (a) segmentation of the eye-tracking data into scenes, (b) identification of an area of interest (AOI) in each scene, and (c) calculation of eye-tracking measures per each AOI.

Since the eye-tracking data comprised the recordings of the participants' eye movements while they were watching the video clips *and* answering the test items on the computer screen, I had to "clean" these data by selecting only the segments that represented the participants' eye movements during the video clips. Thus, during the first step of the eye-tracking data analysis, I used Attention Tool's Usability Module to select six segments called "scenes" for each participant, with each scene representing the participant's eye movements during one of the six video clips from the VALT. Each of the six scenes had the same length across all 33 participants and each was labeled to describe the video it contained (e.g., Video1_context, Video2_content, etc.).

During the second step, I used Attention Tool's Dynamic Media Module to identify a specific area of interest (called "video AOI") for each scene, which contained the area of the screen where the actual video was played during the VALT. Since each video clip occupied only about a quarter of the computer screen, using the video AOI for each scene allowed me to calculate eye-tracking measures that represented the participants' oculomotor engagement with the actual video in the VALT (see Figure 5), as opposed to eye-tracking measures associated with the test takers' looking at the computer screen in general.

In total, I created six video AOIs (i.e., one AOI per video) for each participant. Table 2 summarizes all scenes and video AOIs created in Attention Tool during the eye-tracking data segmentation.

During the third step, I calculated eye-tracking measures for each scene's video AOI per participant. Because all six scenes varied in duration (as indicated in Table 2), I based

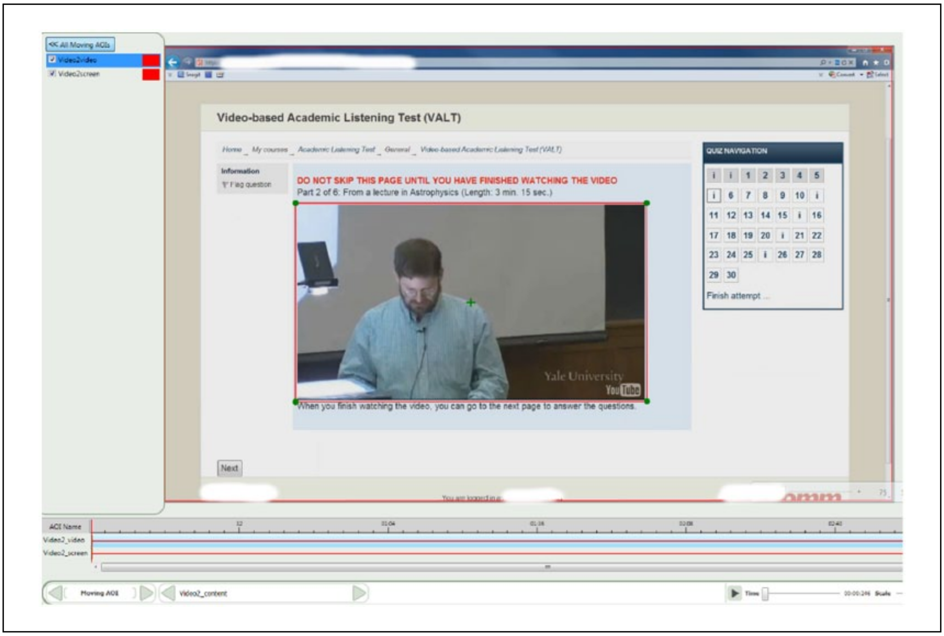

Figure 5. A video AOI for Video 2 created using Attention Tool's Dynamic Media Module (printed with permission from Dr. Charles Bailyn, Yale University).

Table 2. Overview of scenes and AOIs created in Attention Tool.

Scene label	Scene duration (sec.)	Area of interest (AOI)	Associated video
Video1_context	188	Video1_AOI	Video 1
Video2_content	196	Video2_AOI	Video 2
Video3_context	212	Video3_AOI	Video 3
Video4_content	226	Video4_AOI	Video 4
Video5_context	223	Video5_AOI	Video 5
Video6_content	234	Video6_AOI	Video 6

the eye-tracking data analysis on relative rather than absolute eye-tracking measures (Borsboom, Mellenbergh, & Van Heerden, 2002): fixation rate, dwell rate, and the percentage of total dwell time. Only the latter measure was automatically generated by Attention Tool, whereas I had to manually calculate the fixation rate and dwell rate based on the absolute eye-tracking measures provided by Attention Tool. In the following section I give a brief overview of the three eye-tracking measures, provide my rationale for using them, and explain how I calculated them.

Fixation rate is the number of fixations – that is "eye movements that stabilize the retina over a stationary object of interest" (Duchowski, 2007, p. 46) – per second. As has been indicated by previous research, the more important an object is, the more it will be fixated on (Henderson, Weeks, & Hollingworth, 1999; Jacob & Karn, 2003; Yarbus, 1967). Thus, I chose this measure for the analysis in order to determine how semantically important the participants would find each video clip while viewing it. In this study, I calculated the fixation rate using the following formula:

$$F_{rate} = \frac{N_F}{\sum_{i=1}^{N} Dt_i}$$

where F_{rate} is the fixation rate per second (s^{-1}), N_F is the number of fixations, and $\sum_{i=1}^{N} Dt_i$ is the total dwell time in seconds.

Dwell rate, which is also known as gaze rate, is "the number of entries (or visits) into a specific AOI per minute" (Holmqvist et al., 2011, p. 419). Since dwell rate can reflect the importance of a specific AOI for the completion of a task (Jacob & Karn, 2003), I chose it for the analysis to determine how frequently the participants visited (and re-visited) each scene's video AOI containing a context or content video from the VALT. In this study, I calculated the dwell rate using the following formula:

$$D_{rate} = \frac{N_D}{\sum_{i=1}^{N} Dt_i}$$

where D_{rate} is the dwell rate per minute (m^{-1}), N_D is the number of dwells, and $\sum_{i=1}^{N} Dt_i$ is the total dwell time in minutes.

Finally, the percentage of total dwell time is the percentage of time that a viewer spends looking at a specific AOI (Holmqvist et al., 2011). In this study, a dwell time measure comprised the durations of all eye movements, including fixations, non-fixations (e.g., blinks, saccades, and glissades), as well as fixations that are shorter than the algorithmic operational definition of a fixation in Attention Tool (i.e., 100 milliseconds). According to Holmqvist et al. (2011), dwell time can indicate interest in the visual or higher informativeness of the visual (i.e., the amount of meaningful information that the visual carries for the viewer). Previous research has also shown that objects that are less likely to occur in a visual field have higher informativeness, and as such they are looked at longer than objects that are more common (Friedman & Liebelt, 1981; Pieters, Rosbergen, & Hartog, 1996). Thus, I expected the analysis of this measure to reveal any potential differences between the levels of informativeness in context and content videos from the VALT.

I calculated each of the three eye-tracking measures per each individual video AOI, per total three video AOIs associated with context videos, and per total three video AOIs associated with content videos. Finally, I compared the three measures for the total context video AOIs and the total content video AOIs using three paired-samples *t*-tests. This analysis allowed me to determine whether there was any statistically significant difference between the extent to which the participants watched context videos and the extent to which they watched content videos in the VALT.

Research question 2. To answer the second research question, I correlated each of the three eye-tracking measures with scores for the context-video subtest and with scores for the content-video subtest using Pearson product-moment correlation coefficient (r). The use of this correlation coefficient was appropriate because both variables were interval and had relatively normal distributions (Carr, 2011).

Results

Research question 1

The first research question inquired about the difference between L2 test takers' viewing of context videos versus content videos in the VALT, as indicated by eye-tracking measures. Table 3 provides the results of three paired-samples t-tests that I carried out to compare the three eye-tracking measures (i.e., fixation rate, dwell rate, and the percentage of total dwell time) associated with the total context video AOIs and the total content video AOIs.

The results of the first paired-samples t-test revealed that the fixation rate for content videos (M = .87, SD = .42) was higher than the fixation rate for context videos (M = .71, SD = .40), and this difference was statistically significant, $t(32) = 4.73$, $p < .01$. As shown by the effect size (eta squared = .41), the magnitude of the difference between the means was moderate (Cohen, 1988), with the video type explaining about 41% of the variation in the fixation rate. Similarly, the results of the third paired-samples t-test demonstrated that the percentage of total dwell time for content videos (M = 57.99, SD = 19.79) was higher than the percentage of total dwell time for context videos (M = 50.70, SD = 22.49), with the difference also being statistically significant, $t(32) = 5.02$, $p < .01$. A medium effect size (eta squared = .44) signaled that the magnitude of the difference between the means was moderate, as 44% of the variation in the percentage of total dwell time could be attributed to the video type.

Meanwhile, no statistically significant difference was detected between the dwell rate for content videos (M = 29.40, SD = 15.49) and the dwell rate for context videos (M = 29.07, SD = 17.26), as shown by the results of the second paired-samples t-test, $t(32)$ = .38, p = .71. A very small effect size (eta squared = .01) suggested that the magnitude of the difference between the means was negligible, with the video type explaining less than 1% of the variance in dwell rate.

Research question 2

The second research question concerned the extent to which L2 test takers' viewing patterns represented by the eye-tracking measures correlated with their scores on the context-video subtest and on the content-video subtest from the VALT. Table 4 presents the values for the Pearson product-moment correlation coefficient (r) that I calculated to measure the degree of the relationship between the three eye-tracking measures (i.e., fixation rate, dwell rate, and the percentage of total dwell time) and the participants' scores on the context- and content-video subtests from the VALT.

Table 3. Results of three paired-samples *t*-tests comparing three eye-tracking measures for context videos and content videos (*n* = 33).

T-test	Eye-tracking measures	*M*	*SD*	*df*	*t*	*p*	Effect size (eta squared)
1	**Fixation rate**			32	4.73	.01*	.41
	Context videos	.71	.40				
	Content videos	.87	.42				
2	**Dwell rate**			32	.38	.71	.01
	Context videos	29.07	17.26				
	Content videos	29.40	15.49				
3	**Total dwell time**			32	5.02	.01*	.44
	Context videos	50.70	22.49				
	Content videos	57.99	19.79				

Note: Statistically significant *p* values of < .05 are marked with an asterisk (*).

Table 4. Correlations between the three eye-tracking measures and the VALT subtest scores (*n* = 33).

Scores	Fixation rate		Dwell rate		Total dwell time	
	r	*p*	*r*	*p*	*r*	*p*
Context-video Subtest	.32	.07	.04	.81	.23	.21
Content-video Subtest	.02	.93	.15	.41	−.01	.96

The results of the correlation analysis revealed a weak relationship between the context-video subtest scores and the fixation rate for context videos ($r = .32$), but this relationship was not statistically significant at $p < .05$ (perhaps because of the low number of participants), limiting the generalizability of the findings beyond the current data set. An even weaker relationship was found between the context-video subtest scores and the total dwell time for context videos ($r = .23$), and it was also not statistically significant ($p = .21$). All other Pearson product-moment correlation coefficients were close to 0, demonstrating no relationship between the participants' viewing patterns and their scores on the two video subtests from the VALT.

Discussion

Oculomotor engagement with context videos and content videos

The results of the eye-tracking data analysis demonstrated that the L2 test takers spent statistically significantly more time watching content videos (58% of the total time the videos were shown) than context videos (51%) during the VALT. Since dwell time is believed to indicate interest in the visual or higher informativeness of the visual (Holmqvist et al., 2011), these findings imply that the participants in this study were more interested in content videos and found them more informative than context videos in the VALT. The amount of time that the participants spent watching context videos in this study (51%) appears to be similar to the amount of time the participants viewed context videos in Wagner's (2010a) study (48%), but smaller than in his 2007 study that used the same L2 listening test (69%). It should be noted, however, that while in Wagner's (2007, 2010a) studies the participants had access to test questions when the stimuli were being played, no previewing of test items was allowed in the present study.

Next, a statistically significant difference between the fixation rate means for content and context videos evinced that the participants fixated their eyes on content videos more frequently than on context videos in the VALT (i.e., .87 s^{-1} for content videos vs. .71 s^{-1} for context videos). Considering that the fixation rate can indicate semantic importance of a visual (Henderson et al., 1999; Jacob & Karn, 2003), the higher fixation rate for content videos suggested that the participants found content videos more important (more deserving of their attention) than context videos.

The dwell rate values, however, revealed that the number of the participants' revisits of the context and content video AOIs during the VALT did not depend on the video type. Given that the dwell rate can reflect the importance of a specific AOI to the task (Jacob & Karn, 2003), these results imply that the participants found both the context video AOIs and the content video AOIs equally important. In light of the fact that revisits occur when participants move their eyes outside an area of interest and then return back, there are two possible interpretations of this finding. First, it is possible that the participants revisited context and content video AOIs a similar number of times because there were no other semantically important areas of interest on the screen for them to attend to while they were watching each video. Second, based on my observations during the data collection process, most of the revisits appeared to occur after note-taking: When taking notes, the participants would look back and forth between the computer screen and their sheets of paper for note-taking. Thus, a similar number of revisits of the context and content video AOIs might imply that for each video the participants took approximately the same amount of notes (although an empirical investigation of the participants' notes and note-taking behavior would be needed to confirm whether this was the case).

When interpreting the results of the three *t*-tests that were carried out to compare the eye-tracking measures for context video and content video AOIs, it is also indispensable to consider the substantial variation among the participants' viewing patterns, as indicated by large values of standard deviation for all three eye-tracking measures (see Table 3). Large standard deviation values suggest that the data were widely spread and that the participants varied considerably in terms of (a) the amount of time they spent watching content and context videos during the VALT, (b) the number of times they revisited the context and content video AOIs per minute, and (c) the number of times they fixated on the context and content video AOIs per second. For example, the values for total dwell time indicated that Participants 5 and 11 spent less than 20% of the time watching context and content videos during the VALT, whereas Participants 1, 29, and 31 watched both types of videos for more than 80% of the time. Regarding the amount of time spent for watching video stimuli, these results appear to be in line with the findings of previous research by Ockey (2007) and Wagner (2007, 2010a) who also found considerable variations in the viewing rates among the participants. However, unlike the present study that employed eye-tracking technology to record L2 test takers' eye movements and calculate eye-tracking measures, in the previous studies the viewing rates were based on the crude data gathered by videotaping the participants during media-supported L2 listening assessment.

The variations among the participants' viewing patterns in the present study should be interpreted with caution, as they might have been partially affected by the quality of eye calibration and the eye-tracking data, which differed slightly among the participants. The quality of eye calibration to some extent depended on certain physical characteristics of the participants' eyes. In particular, the robustness of the eye-tracking system deteriorated when it was exposed to eyes with very dark pupils or very narrow eyes with eyelids obscuring part of the pupils, as, for example, was the case with the eye calibration for Participant 11. Obviously, when recording the eye movements of participants with such types of eyes, the eye-tracking system would occasionally fail to recognize their pupils. Furthermore, the quality of the eye-tracking data was also slightly diminished for those participants who spent a great deal of time taking notes during the test: Every time the test takers moved their gaze from the note-taking paper back to the computer screen, the eye-tracking system had to keep re-discovering and re-capturing their pupils. It should be noted, however, that the magnitude of this problem was not large as the average quality of the eye-tracking data for all participants was 94%, with the lowest quality of 85% for Participant 11. Overall, the variations among the participants' viewing patterns in this study could largely be attributed to individual viewing differences (and not their physical differences) among the participants.

Relationship between viewing patterns and test scores

The correlation results indicated that overall the L2 test takers' scores on the context-video subtest and on the content-video subtest were not strongly related to their viewing patterns–namely, the amount of time they spent watching the videos, the number of times they fixated on the videos, and the number of times they revisited each video during the video-mediated L2 listening test were not strongly associated with their test scores. However, there was a weak but non-significant relationship between the fixation rate for context videos and the context video subtest scores ($r = .32$, $p = .07$), and there was a similar weak (but non-significant) relationship between the context-video dwell time and the context-video test scores ($r = .23$, $p = .21$). These were likely statistically insignificant correlations because of this study's relatively small sample size ($n = 33$). In other words, within this study's data set, the amount and the nature of the participants' oculomotor interaction with context-video stimuli were weakly associated with their context-video subtest scores, but these findings cannot be generalized beyond the current data set. Furthermore, test takers who received higher scores on the VALT subtests varied significantly in terms of the amount of time they spent watching the videos, as did the test takers who received lower scores on the two subtests. This area of research clearly needs more attention, with larger data sets being warranted.

Limitations

This study has two main limitations. First, the eye-tracking equipment used in the study had the data-sampling rate of 80 Hz, which was relatively low. Cognitive psychologists generally contend that eye-tracking systems with low data-sampling rates are adequate only for usability studies (i.e., studies that evaluate a product by testing it on the target audience), whereas for research on reading–or other types of eye-tracking research that require more precise data–the sampling rate must be 500 Hz (Collewijn, 1999; Poole & Ball, 2006; Rayner & Pollatsek, 1989) or at least 250 Hz (Jacob & Karn, 2003). Although the research design of this study did not necessitate a very high precision of the eye-tracking data, the low data-sampling rate had a somewhat deleterious impact on the quality of eye calibration and the obtained data. Second, note-taking appeared to have a direct effect on the L2 test takers' viewing behavior as they had to move their eyes away from the computer screen when taking notes. Because the effect of note-taking was not taken into consideration during the analysis of the eye-tracking data, the results regarding the participants' viewing behavior in this study should be interpreted with caution and may be difficult to generalize.

Conclusion and directions for future work

In summary, the results of this study revealed that the L2 test takers interacted extensively both with context videos and with content videos during the VALT. The extent and nature of their interaction with the two video types, as shown by the results of the eye-tracking data analysis, indicated that the test takers chose to use both visual information and auditory information while listening. These findings, in my opinion, expose the outdatedness of Buck's (2001) view that in L2 listening assessment "the emphasis should be on processing linguistic information, not visual information" (pp. 253–254) as "the visual information may serve to increase the cognitive load of the test taker, and that may interfere with the testing process" (p. 254). One can argue that had the visual information in this study increased the test takers' cognitive load or interfered with their testing process, they would have chosen to avoid making eye contact with the videos and instead look elsewhere, focus on note-taking, or simply close their eyes to concentrate on audio.

Responding to Ockey's (2007) and Wagner's (2006, 2008) calls for gathering new types of evidence to expand the construct of L2 listening by including the ability to understand visual information, this study provided eye-tracking evidence in support of such an expansion (namely, that visual information appears to play an important role in the L2 listening process in an academic setting). Meanwhile, considering a large variety of social contexts and settings in which listening may occur (e.g., an academic classroom, a phone conversation, news broadcasts, etc.), the definition of a universal construct of L2 listening seems unfeasible to me. Instead, I would argue that different constructs of L2 listening should be defined for different target language use (TLU) domains (Bachman & Palmer, 1996), and should include the ability to understand visual information if such information is part of the considered domain.

Next, although in this study I found differences between context videos and content videos in terms of their levels of informativeness and their semantic importance for L2 test takers, I also discovered that the context–content video distinction appears to be relatively crude and ambiguous. For example, content videos can provide information not only about the content, but also about the context of the verbal message, thus overlapping with the functions of context videos. In fact, in his analysis of visual languages Pettersson (2002) suggested that all visuals possess both content and context. Unlike images that lend themselves better to the context–content distinction (Bejar et al., 2000; Ginther, 2002), videos appear to be very complex multimodal texts, whose perception and interpretation by the viewer depends not only on the type of the recorded visual information, but also on constant dynamic changes of the scenes; peculiarities of the recorded objects, speakers, and their actions (e.g., gestures and lip movements); as well as the position of the camera, types of shots (e.g., close, medium, or close-up), and editing style – the characteristics that are not represented in the context–content classification. Consequently, future research on video-based L2 listening assessment should move beyond the limitations of the context–content distinction and explore other dimensions, such as the extent of semantic congruity between the visual and the auditory information, the degree of rhetorical effectiveness of visuals and their elements, and the types of discourse in videos.

One possible direction for such research is multimodal discourse analysis. From the perspective of multimodal film discourse analysis (Bateman & Schmidt, 2012; Wildfeuer, 2014), discourse in videos contains not only "clearly defined and predesigned meanings, but also the potential for new and non-specific meanings" (Wildfeuer, 2014, p. 174) that can be constructed by the viewers on the basis of their world knowledge. Based on this perspective, video-based listening would not be regarded simply as a receptive language skill that presupposes *comprehension of information* existing independently from the listener's conceptualization. Instead, it would be viewed as a multimodal skill that involves interaction with different modes (such as audio, text, and images) in a video resulting in the *construction of meaning* by the listener-reader-viewer. For such shift in the paradigm to occur, future research on L2 multimodal listening would need to be informed by the emerging field of multimodality and (systemic functional-)multimodal discourse analysis (see, e.g., Bateman, 2013; O'Halloran, 2008).

With regard to eye tracking, this study appears to be the first one in the field that capitalized on the potential of this technology and revealed its methodological value for research on video-based L2 listening assessment. By expanding the use of eye tracking into other areas of research on L2 listening assessment, future studies might be able to provide new types of evidence about L2 learners' test-taking behavior. While employing this technology, however, researchers must remember that eye tracking can yield data only about the participants' foveal vision (i.e., the area of the visual field that their eyes are directed at), not their peripheral vision, mental attention (i.e., the area of the visual field that they concentrate on cognitively), or perception (i.e., how they encode the information from the visual field). As such, eye-tracking data cannot provide information about the reasons why the participants watch a specific element of a visual. Thus, future studies should employ triangulation and complement eye-tracking data with verbal report data (also suggested by Godfroid & Schmidtke, 2013) to get information not only about L2 test takers' viewing patterns, but also about the reasons for those patterns.

There are many research areas in L2 listening assessment where eye tracking can be used. For instance, eye-tracking research can be done to determine which aspects of visual information L2 learners use to answer individual test items, as well as which aspects they find helpful and which ones they find distracting, and why. Eye tracking can also be employed to examine how L2 test takers interact with test items. Last, but not least, exploring the effect on note-taking on L2 test takers' viewing behavior during video-mediated L2 listening assessment is another interesting venue for this type of research.

Acknowledgement

I would like to thank Dr. Paula Winke from Michigan State University for her invaluable feedback on the earlier draft of this manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was part of a larger research project supported by the Small Grant for Doctoral Research in Second Language Assessment from Educational Testing Service (ETS) and the Assessment Research Award from the British Council.

References

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.

Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review*, *50*(3), 507–531.

Bateman, J. A. (2013). Hallidayan systemic-functional semiotics and the analysis of the moving audiovisual image. *Text & Talk*, *33*(4–5), 641–663.

Bateman, J. A., & Schmidt, K.-H. (2012). *Multimodal film analysis: How films mean*. New York and London: Routledge.

Batty, A. O. (2014). A comparison of video- and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing*. Prepublished May 14, 2014, DOI: 10.1177/0265532214531254.81

Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, *30*(4), 441–465.

Bax, S., & Weir, C. (2012). Investigating learners' cognitive processes during a computer-based CAE reading test. *University of Cambridge ESOL Examinations Research Notes*, *47*, 3–14.

Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL 2000 listening framework: A working paper*. Princeton, NJ: Educational Testing Service.

Bloomfield, A., Wayland, S., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult? Factors affecting second language listening comprehension* (Technical Report No. E.3.1 TTO 81434). College Park, MD: University of Maryland, Center for Advanced Study of Language.

Boraston, Z., & Blakemore, S.-J. (2007). The application of eye-tracking technology in the study of autism. *Journal of Physiology*, *581*(3), 893–898.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2002). Different kinds of DIF: A distinction between absolute and relative forms of measurement invariance and bias. *Applied Psychological Measurement*, *26*, 433–450.

Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. *System*, *25*(1), 39–53.

Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, *18*, 171–191.

Buck, G. (2001). *Assessing listening*. Cambridge, UK: Cambridge University Press.

Carr, N. (2011). *Designing and analyzing language tests*. Oxford, UK: Oxford University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Collewijn, H. (1999). Eye movement recording. In R. H. S. Carpenter & J. G. Robson (Eds.), *Vision research: A practical guide to laboratory methods* (pp. 245–285). Oxford, UK: Oxford University Press.

Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. *System*, *29*, 1–14.

Cubilo, J., & Winke, P. (2013). Redefining the L2 listening construct within an integrated writing task: Considering the impacts of visual-cue interpretation and note-taking. *Language Assessment Quarterly*, *10*(4), 371–397.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (2nd ed.). London, UK: Springer-Verlag.

Flowerdew, J., & Miller, L. (2010). *Listening in a second language*. In A. D. Wolvin (Ed.), *Listening and human communication in the 21st century* (pp. 158–177). Oxford, UK: Wiley-Blackwell.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.

Friedman, A., & Liebelt, L. S. (1981). On the time course of viewing pictures with a view towards remembering. In D. F. Fisher, R. A. Monty & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 137–155). Hillsdale, NJ: Lawrence Erlbaum.

Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, *19*(2), 133–167.

Glenn, E. C. (1989). A content analysis of fifty definitions of listening. *Journal of the International Listening Association*, *3*, 21–31.

Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports, and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183–205). Honolulu, HI: University of Hawai'i at Manoa, National Foreign Language Resource Center.

Gruba, P. (1993). A comparison study of audio and video in language testing. *JALT Journal*, *15*(1), 85–88.

Gruba, P. (1999). The role of digital video media in second language listening comprehension (Unpublished doctoral thesis). *Department of Linguistics and Applied Linguistics, University of Melbourne, Melbourne, Australia*. Retrieved from http://eprints.unimelb.edu.au/archive/00000244.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.

iMotions. (2013a). Attention Tool (Version 4.8) [software]. Available from http://imotionsglobal.com/software/attention-tool-core/.

iMotions. (2013b). Dynamic Media Module [software]. Available from http://imotionsglobal.com/software/add-on-modules/attention-tool-dynamic-media-module/.

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human–computer interaction and usability research: Ready to deliver the promises. In J. Hyönä, R. Radach & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–605). Amsterdam, Netherlands: Elsevier Science.

Londe, Z. C. (2009). The effects of video media in English as a second language listening comprehension tests. *Issues in Applied Linguistics*, *17*(1), 41–50.

Nagle, S. J., & Sanders, S. L. (1986). Comprehension theory and second language pedagogy. *TESOL Quarterly*, *20*(1), 9–26.

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

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.

Pettersson, R. (2002). *Information design: An introduction*. Amsterdam, Netherlands: John Benjamins.

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. Y. (2011). Visual cues and listening effort: Individual variability. *Journal of Speech, Language, and Hearing Research*, *54*, 1416–1430.

Pieters, R., Rosbergen, E., & Hartog, M. (1996). Visual attention to advertising: The impact of motivation and repetition. *Advances in Consumer Research*, *23*, 242–248.

Poole, A., & Ball, L. J. (2006). Eye tracking in human–computer interaction and usability research: Current status and future prospects. In C. Ghaoui (Ed.). *Encyclopedia of human–computer interaction* (pp. 211–219). Hershey, PA: Idea Group.

Progosh, D. (1996). Using video for listening assessment: Opinions of test-takers. *TESL Canada Journal*, *14*(1), 34–44.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Harlow, UK: Pearson.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, *78*(2), 199–221.

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

Sueyoshi, A., & Hardison, D. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661–699.

Suvorov, R. (2009). Context visuals in L2 listening tests: The effects of photographs and video vs. audio-only format. In C. A. Chapelle, H. G. Jun & I. Katz (Eds.), *Developing and evaluating language learning materials* (pp. 53–68). Ames, IA: Iowa State University.

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

Taylor, L. (2013). Introduction. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening*. Studies in Language Testing 35 (pp. 1–35). Cambridge, UK: UCLES/Cambridge University Press.

Wagner, E. (2002). Video listening tests: A pilot study. *Working Papers in TESOL & Applied Linguistics, Teachers College, Columbia University*, *2*(1). Retrieved from http://journals.tc-library.org/index.php/tesol/issue/view/2

Wagner, E. (2006). Utilizing the visual channel: An investigation of the use of video texts on tests of second language listening ability (Unpublished doctoral dissertation). New York: Teachers College, Columbia University.

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

Wagner, E. (2008). Video listening tests: What are they measuring? *Language Assessment Quarterly*, *5*(3), 218–243.

Wagner, E. (2010a). Test-takers' interaction with an L2 video listening test. *System*, *38*, 280–291.

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

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

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke, UK: Palgrave Macmillan.

Wildfeuer, J. (2014). *Film discourse interpretation: Towards a new paradigm for multimodal film analysis*. New York and London: Routledge.

Wolvin, A., & Coakley, C. G. (1996). *Listening* (5th ed.). Dubuque, IA: Brown & Benchmark.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.