The impact of input format on written performance in a listening-into-writing assessment

Carolyn Westbrook

ARTICLE INFO

Abstract

Over the last five decades, research in teaching and testing (academic) listening has investigated different foci. Initially, teaching listening involved bottom-up approaches (Dirven and Oakeshott-Taylor, 1984) then both higher- and lower-level processes were integrated (Voss, 1984). In the early 2000s, different input formats (Read, 2002) and discourse features of lectures (Thompson, 2003) were the subjects of academic listening research. More recently, EAP tests have increasingly taken an integrated approach to reflect real-world tasks, yet few studies have looked at integrated listening-into-writing tasks (Cubilo and Winke, 2013).

This counter-balanced measures design study investigates how test taker performance differs on an integrated EAP listening-into-writing task when lecture input is presented as audio only in one half and video in the other half of the input. Two groups of test takers took part in the current study.

A Hotelling’s $T^2$ test revealed a statistically significant effect on scores when test takers were presented with the audio only input first but there was no significant effect on scores when the video input was presented first. Data on test taker preferences revealed that more people preferred the video input to audio only.

1. Introduction

Many university courses around the world are offered in English, which, among other skills, requires the ability to listen to lectures and write academic papers. Therefore, universities and testing organisations measure students’ English proficiency to ensure their suitability for study. This has typically been done using both discrete and integrated tasks.

Both the ELTS test and later, the IELTS test, were designed to reflect some ‘features of academic language’ (IELTS, 2019). The former included tasks with an integrated skills focus. Unfortunately, these were replaced by discrete skills sub-tests in the IELTS test. Yet, in recent decades, researchers have come to recognize that integrated tasks reflect the target language use domain (TLU) much more closely because ‘in an academic context there is necessarily some input for any writing task that has to be carried out’ (Weir, 1983, p. 376). Thus, many tests of English for Academic Purposes (EAP) now include integrated skills tasks (Plakans & Gebril, 2012, p. 217).

While integrated academic reading-into-writing tasks have received a good deal of attention in the literature, integrated EAP listening-into-writing tasks have received comparatively little. Given the huge strides made by developments in technology and the...
increased possibilities to teach and test academic listening-into-writing that these have led to, it is perhaps surprising that relatively little research has been done in this area, particularly with regard to the different options for teaching and testing integrated EAP listening.

Traditionally, lectures are delivered live by the lecturer and students’ assessments - whether formative or summative - focus on tasks which require them to demonstrate their understanding of the content, for example, as part of a group discussion, a presentation, a report or an essay. More recently, however, due to the influence of technology and changes in pedagogical practices, not to mention the impact of Coronavirus, ‘flipped’ and online learning have been widely adopted in the sector. Therefore, it is important to consider how best to present online input both for teaching and for assessment purposes. A good deal of research has investigated the impact of audio and video input in discrete listening tests but very little work has examined the impact of different input formats on integrated listening-into-writing tasks.

This paper will give a brief outline of the research and developments in the testing of listening and, specifically, academic listening practices over the last five decades since the emergence of BALEAP"[formerly SELMOUS"] and will then present the results of one part of a larger PhD study which focussed on how listening input provided in the form of a podcast (audio-only) versus a vodcast (video with PowerPoint) impacted on written performance in an integrated EAP listening-into-writing test.

2. Literature review

2.1. Overview of listening research

In JEAP specifically, with the exception of the 2011 Special Issue on academic listening, which includes four state-of-the-art articles, there appears to be relatively little research into academic listening assessment over the 20 years of the journal’s existence. Lynch (2011) attributed this to the ‘inherent complexity of listening and listening research’ due to the numerous internal and external factors that can impact on listening ability as well as the difficulty of researching listening effectively’ (p. 80). He states that this is not a criticism of the journal’s editorial policy but a reflection of ‘a wider neglect of listening’ (p. 80). Indeed, his review of the research published in JEAP up to that point reveals that, out of just nine listening-focussed articles, only one, a study by Read (2002), focussed on EAP listening assessment.

Nine years later, in their review of 416 JEAP articles between 2002 and 2019, Riazi et al. (2020) found that only six articles (just over 1%) focussed on academic listening whereas there were 276 (approximately 66%) articles on academic writing. Only 38 studies (9%) reported on more than one modality, e.g. reading and writing or listening and speaking.

Early pedagogy and research focussed largely on bottom-up approaches to listening comprehension with the emphasis on phoneme, syllable and word level processing (Dirven & Oakeshott-Taylor, 1984, p. 326). By the 1970s and 1980s, the emphasis was on the integration of both lower- and higher-level linguistic processes (Voss, 1984) and the difficulties that learners encounter with listening. These include discourse structure (Godfrey, 1979), internal and extraneous difficulties (Zimmerman, 1980, in Dirven & Oakeshott-Taylor, 1985, p. 7) and the effect of memory on listening comprehension (Richards, 1983). In the 1990s and 2000s, with the developments that new technology had brought, researchers started to investigate the impact of different input formats (Brett, 1995; Coniam, 2001; Gruba, 1997; Ockey, 2007) and were proposing ways of teaching (Field, 2008) and assessing listening comprehension (Coniam, 2001) based on the findings.

Around the same time, research in English for Academic Purposes was gaining momentum. Research into academic listening in the 1980s and beyond examined features of listening comprehension and lecture comprehension (Dunkel & Davis, 1994; Miller, 2009; Young, 1994). Richards (1983) presents a list of micro-skills required when listening to lectures, noting, among other aspects, the need to ‘identify relationships among units within discourse’, the ‘ability to identify the role of discourse markers in signaling structure of a lecture’, the ‘ability to follow different modes of lecturing: spoken, audio, audio-visual’, the ‘ability to follow lecture (sic) despite differences in accent and speed’ and ‘to recognize irrelevant matter’ (p. 229–230).

Although Richards’ (1983) list of micro-skills for academic English has been criticised for being limited to just academic lectures, research has investigated some of these skills over the last three decades. For example, areas of focus have included note-taking (Carrell, 2007; Chaudron et al., 1994; Dunkel, 1988; Siegel, 2018; 2020) and speech rate (Griffiths, 1990; Révész & Brunfaut, 1994). 1

1 In a ‘flipped classroom’, recorded lecture input is provided for students to access in their own time before a lesson, and contact time in the classroom is used to provide students with the opportunity to engage with the content, through discussion, analysis or other tasks which require an understanding of the principles or the theory provided in the input.

2 BALEAP, the global forum for EAP professionals (from 1989 to 2010 the British Association of Lecturers in English for Academic Purposes) is the professional body for EAP in the UK and internationally.

3 SELMOUS (Special English Language Materials for Overseas Students) was the predecessor to BALEAP, founded in 1972.
The importance of understanding discourse structure (Camiciottoli, 2004; Dudley-Evans, 1994; Zare & Keivanloo-Shahrestanaki, 2017) and the use of discourse markers (Chaudron & Richards, 1986; DeCarrico & Nattinger, 1988; Thompson, 2003) have also received a good deal of attention. Several studies have found that clearly signposted discourse can benefit both L1 and L2 learners (DeCarrico & Nattinger, 1988; Rickards, Fajen, Sullivan, & Gillespie, 1997). More recently, Zare and Keivanloo-Shahrestanaki (2017) found that an understanding of how importance is marked in academic lectures can improve comprehension of the main points. Other areas that have been investigated in relation to EAP listening include factors affecting performance including vocabulary acquisition (Paribakht & Webb, 2016; Vidal, 2003) and text length (Carrell et al., 2004; Locke, 1977).

The impact of lecture length in both live and asynchronous lectures has been investigated over the years. Locke (1977) found that there was a 17% drop on average between the quantity of lecture notes taken in the first 20 min of a lecture and those taken in the last 10–30 min of 50–70 min lectures. More recently, Inman and Myers (2018) cite several authors who recommend that lectures should be broken down into 10–15 min sections (p. 3). In an asynchronous environment, studies have found that students’ attention span may be even shorter. Guo et al. (2014) found a significant drop in engagement when students were presented with videos which were longer than 9–12 min and the median engagement time was 6 min (p. 44).

2.2. Input formats in listening tests

One aspect of listening which has received a good deal of attention in both general English and English for Academic Purposes is the impact of different input formats (audio, video and multimedia) (Batty, 2015; Coniam, 2001; Pardo-Ballester, 2016; Sueyoshi & Hardison, 2005; Suvorov, 2013; Wagner, 2007), as well as the use of context versus content stills (Ginther, 2002), and captions (Leveridge & Yang, 2013; Montero Perez et al., 2014; Sydorenko, 2010). However, these studies have provided mixed findings. Some researchers found no statistically significant difference across input formats (Coniam, 2001; Cubilo & Winke, 2013) while others found that test takers perform better on a listening test containing video input (Batty, 2015; Sueyoshi & Hardison, 2005; Wagner, 2010); in contrast, Suvorov’s (2008) study revealed that test takers performed significantly worse on a video-mediated lecture compared to an audio-only lecture or a listening text presented with a single photograph. However, his findings suggest that video might, in fact, be beneficial to test takers if the input is in the form of a dialogue.

It should be noted here that the visuals in Suvorov’s (2008) study were context visuals; if content visuals had been used, the results may have been different. Indeed, in her 2002 study, Ginther found that content visuals were more helpful than context visuals. Building on his earlier work, Suvorov (2013) investigated test takers’ interaction with context and content visuals in audio versus video-based input in multiple choice academic listening tests. Although there was no impact on scores relating to the type of visual, the use of eye-tracking software revealed differences in viewing behaviour with the mean fixation rate being statistically significantly higher for content videos than context videos. In addition, 97% of Suvorov’s participants reported that content visuals aided comprehension of the lecture.

With regard to multimedia input, Aldera (2015) investigated the impact over several classes of multimedia input (audio with visual animation) compared to audio only. The group exposed to multimedia input performed better in both a post-test and a delayed post-test. Although this suggests that the input helped the students’ listening skills, it should be noted that students may have performed better purely as a result of the visual input, and not necessarily because the multimedia input helped their listening skills. Nonetheless, these findings concur with Brett (1997) who also found that students presented with multimedia input outperformed those with audio and video input.

In terms of test taker preferences for the input formats, several studies have revealed that test takers prefer video input even if their scores are not always in line with this preference (Pardo-Ballester, 2016; Progosh, 1996). Other study findings have been mixed, with some test takers expressing a preference for video and others considering it a distraction (Chen et al., 2014; Coniam, 2001; Cubilo & Winke, 2013). This may be due to a number of factors including the type of visuals (whether these are context/content visuals) or the cognitive load imposed when the input is too fast.

Research examining the use of captions has also produced mixed findings. Montero Perez et al. (2014) found that learners who are exposed to fully-captioned input perform better on global comprehension questions than those with only keywords or no captions but not on detail questions; other studies have revealed differences based on proficiency level. Pujolà (2002) found that lower-proficiency learners may focus on reading the captions because they consider them a ‘necessary tool in their understanding of the authentic aural input’ (p. 254). Similarly, Leveridge and Yang (2013) found that lower-proficiency learners benefitted from the use of captions yet they caused interference for more proficient learners. In contrast, Aldukhayel’s (2021) study revealed that lower-proficiency students struggled to process all the inputs (audio, visual images and captions) simultaneously when watching a vlog.

2.3. Integrated tasks

What is striking about the studies above is that the vast majority are based on discrete listening tasks so, despite the fact that listening to lectures and taking notes is a frequently cited listening task, and responding to lecture input both in spoken and written

---

4 Context visuals are images which represent the context in which the audio exchange occurs, e.g. a photo showing the speaking in a dialogue (Ginther, 2002, p. 134).

5 Content visuals are images which are ‘related to the content of the verbal stimulus’ (Ginther, 2002, p. 134).
Table 1
Background information of the test takers.

<table>
<thead>
<tr>
<th>Total no. of test takers</th>
<th>Nationality</th>
<th>Gender</th>
<th>Age</th>
<th>Self-assessed English language level</th>
<th>Year of study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>F</td>
<td>No info</td>
<td>18-21</td>
</tr>
<tr>
<td>RF uni</td>
<td></td>
<td>22</td>
<td>50</td>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td>UA uni</td>
<td></td>
<td>4</td>
<td>42</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>26</td>
<td>88</td>
<td>2</td>
<td>102</td>
</tr>
</tbody>
</table>

1 RF uni = participants from Russian Federation university.
2 UA uni = participants from Ukrainian university.
3.1. Research questions

On the basis of the literature review, this paper will present the results of investigations designed to answer the following research questions:

1. How do input order and format (audio vs video) affect written performance on an integrated listening-into-writing EAP task?

2. With regard to test takers’ perceptions of their performance, what reasons do test takers give for why they felt they performed better when presented with one input format rather than the other?

3. With regard to test takers’ perceptions of their performance, what reasons do test takers give for why they felt they performed worse when presented with one input format rather than the other?

3.2. Participants

The data for this study was collected from participants at two universities: one located in the Russian Federation (n = 74) and another in Ukraine (n = 42). In total, data was collected from 131 participants. However, five participants did not sign the consent form while three did not write anything in the main EAP task so these students were removed from the analysis. For a further seven students, the data collection was carried out incorrectly so they were also excluded thus resulting in a total of 116 test takers. Demographic data for the participants can be found in Table 1 below.

University students were selected because the main task was designed to investigate academic language use and required students to listen to a lecture as part of an integrated EAP listening-into-writing test. The two universities in the study were selected because they both have a post-Soviet educational culture and a similar linguistic background. The two universities also have a strong internationalisation agenda so students need to be able to follow lectures in English.

All participants were informed about the study by way of an information sheet and ethical consent was collected from all participants. A copy of the information sheet and ethical consent can be found in Appendix 1. Ethical clearance was obtained from the University of Bedfordshire and the two universities involved in the study.
3.3. Materials and methods

Further to an exploratory study during which the materials and methodology were refined, the main study comprised three main tasks:

1. Lexico-grammatical placement test: Quick Placement Test (QPT) (Oxford University Press, 2001)
2. Integrated EAP listening-into-writing task
   a. Listening to, and taking notes on, a lecture on the topic of ‘Culture shock’ which was split in half, with one half being provided as audio only and one half as a video including PowerPoint.
   b. Writing a summary of the lecture content
3. Post-task feedback questionnaire

3.3.1. Task 1 - Quick Placement Test (QPT)

In the first instance, test takers took a pen-and-paper version of the QPT (Oxford University Press, 2001). This 30-min test comprises 60 multiple-choice items. The rationale for selecting this test was fourfold. First, as a published test, it had already been trialled and validated externally. Second, it had been used for placement purposes at the author’s institution over many years. Third, lexis has been shown to play an important role in listening (Bonk, 2000; Stæhr, 2008, 2009; Van Zeeland & Schmitt, 2013) and writing (Laufer & Nation, 1995; Stæhr, 2008). Finally, the QPT provides a swift yet reliable estimate of students’ proficiency level so participants could be split into two groups of equal number and balanced levels of language proficiency. This was done to ensure that any group differences that were revealed in the EAP test performance were the result of the EAP test rather than the result of having two groups of very different ability ranges. An independent samples t-test revealed that there was no statistically significant difference between the two groups (p = .372) (see Results below).

3.3.2. Task 2 - integrated EAP listening-into-writing test

The main instrument for the study was the integrated EAP listening-into-writing test. The test comprised two tasks: a) watching and listening to a lecture and taking notes on the content b) writing a 350-word summary of the main and supporting points from the lecture.

The research methodology for the first task (the input part) employed an ‘AB-BA’ counterbalanced measures design (Mackey & Gass, 2005, p. 353) in which test takers listened to a lecture which was divided into two halves. The first group (‘audio first’) was presented with the first half of the lecture as audio input and the second half as video input while the second group (‘video first’) watched a video of the first part and listened to an audio recording of the second half (see Table 2 below).

In line with the findings from Guo et al. (2014) (see Literature Review above) and the MOOC platform provider, EdX, which recommends 6 minutes as the maximum length for a video lecture (Inman & Myers, 2018, p. 3), the whole lecture lasted 12:42 min. This was divided as follows:

- Introduction and outline: 0:33
- First half of the lecture: 6:14
- Second half of the lecture: 5:46
- Close of the lecture: 0:09

Following initial trials, the instrument was refined to ensure that both groups had a similar experience. Apart from the first slide at the beginning of the lecture, which showed the outline, each video included three slides: one with one sentence, one with five bullet points and one with a diagram which included five keywords. The reason for the three types of input on the slides was to ensure that the two halves were as comparable as possible while also providing variation in the way the visual information was presented.

The two halves of the input were designed to be of approximately equal playback length and were analysed using Text Inspector (Text Inspector, 2019) to ensure that they were approximately equal in terms of the linguistic features including word count and textual complexity. To measure this, the main body of each half of the transcript (having removed the introduction and the close of the lecture, and having cleaned the transcript to remove false starts and hesitations) was analysed using Text Inspector (Text Inspector, 2019). Text Inspector is an online, automated text analysis software which performs many of the same analyses as other similar products on the market. Like other software programmes, Text Inspector provides the usual descriptive statistics such as word and sentence count, and lexical diversity measures such as type/token ratio and MTLD; however, it also analyses input for occurrences of Academic Word List words (Coxhead, 2000) and provides CEFR levels for the words in the text based on the English Vocabulary Profile (Cambridge University Press, 2015). The contents of the slides in each half of the lecture were also analysed using Text Inspector. The results of both of these analyses can be found in Appendix 2 and demonstrate that the spoken input was very similar across the two halves of the input although there was more variation in the results for the language on the slides.

---

6 Lexical diversity measures the difficulty of a text based on ‘flexibility’ and ‘vocabulary richness’ (Read, 2000, in Durán et al., 2004, p. 221).
7 MTLD (Measure of textual lexical diversity) measures lexical diversity by calculating the mean length of word strings that maintain a criterion level of ‘lexical variation’ (McCarthy & Jarvis, 2010, p. 381).
In terms of speed of delivery of the lecture, Brindley and Slayter (2002) note that ‘normal’ speed texts are delivered at a speed of 180 words per minute (wpm) while Griffiths (1992) suggests that the ‘average’ speech rate is 188 wpm. In an earlier study, Tauraosa and Allison (1990) cited an average speech rate of 125–160 wpm. Camiciottoli (2005) measured the speech rate in a lecture delivered in an L1 environment and one delivered in an L2 environment. In line with Tauraosa and Allison (1990), the speech rate in the L2 lecture was 125 wpm while the lecture delivered in an L1 setting in the UK had a speech rate of 183 wpm, which is more in line with Brindley and Slayter (2002) and Griffiths (1992). Investigating speech rates in the Cambridge Suite of exams, Field (2013, p. 119) found that the speech rate on the PET exam - a CEFR Level B1 exam - was 167.4 wpm on average while the Cambridge FCE – a B2 level exam – had an average speech rate of 207.6 wpm. Therefore, the speed of delivery of the lecture was kept in line with the speed that would be expected for a B1/B2 exam. The speech rate for the introduction was slightly slower than the main content at 167.09 wpm to allow test takers to become accustomed to the speaker’s accent. The speech rate for the first half of the lecture was 177.75 wpm and for the second half of the lecture, it was 182.26.

Test takers heard the input twice and were allowed to take notes at any time. After the second playback, test takers moved on to the second task in the test (the output part). In this task, they had 45 min to write a 350-word summary.

### 3.3.2.2. Statistical analysis

#### 3.3.2.2.1. Rating of students’ written responses

To develop a model answer, an expert panel of three L1 English speaker EAP lecturers listened to the whole lecture as audio input and noted down the salient points. Then they agreed on a consensus version, which could be used while rating to evaluate the extent of task achievement. The audio only version was used for note-taking as they all had access to the PowerPoint slides while rating. Raters used the consensus version to assess the amount of content that test takers reproduced in relation to the CEFR descriptors in the Overall Written Production scale (Council of Europe, 2001) (see below). For example, if a test taker only reproduced a few isolated phrases, this would be an A1 performance but if they wrote a clear, well-structured text containing the vast majority of the points from the input, this would constitute a C1 performance.

Before being rated, the summaries were split into two halves, representing the two halves of the input. The three EAP lecturers underwent rater familiarisation training to develop a shared understanding of the criteria in line with (Trace et al., 2016, p. 41) then used the CEFR Overall Written Production descriptors (Council of Europe, 2001) to rate the papers. Each paper was rated by between 1 and 3 raters whereby rater 1 rated 96 papers, rater 2 rated 74 papers and rater 3 rated 96 papers. This allowed sufficient overlap for a Rasch analysis to be carried out.

The CEFR Overall Written Production scale (Council of Europe, 2001) was used as this scale provides a clear progression from ‘simple isolated phrases’ (A1) to ‘clear, detailed texts’ (B1) up to ‘clear, well-structured texts …’ (C1) (p. 23). This enabled raters to distinguish between those at the lower level who were only able to (re)produce odd words or phrases and the more detailed, well-written texts at higher levels.

### 3.3.2.2. Statistical analysis

The scores allocated to test takers were analysed quantitatively using Many Facet Rasch Analysis (Linacre, 1989) to calculate test takers’ Fair Average scores on each half of the summary writing task and to measure rater harshness/leniency and Infit. The Fair Average scores were then used as the basis for a Hotelling’s $T^2$ test, which was carried out using SPSS Version 22 (IBM Corp, 2013). In contrast to a $t$-test, which is used to test for differences between groups when there is only one dependent variable, Hotelling’s $T^2$ can be used when there are several dependent variables. This test is similar to a Multivariate Analysis of Variance (MANOVA) but a MANOVA is usually run when there are three or more groups in the independent variable whereas Hotelling’s $T^2$ can be used with two groups for the independent variable (Laerd, 2015), which was the case in this study.

### 3.3.3. Task 3 - feedback questionnaire

After participants had completed the integrated listening-into-writing task, they were asked to complete a pen-and-paper feedback questionnaire. This was designed to collect demographic data but also to investigate their perceptions of the task. All test takers were asked to complete the questionnaire but, in some cases, they did not answer all the questions.

### 4. Results

#### 4.1. Quick Placement Test

Table 3 below shows the QPT results. As can be seen, the mean score in the ‘audio first’ group was 1.98 points higher than that of the ‘video first’ group ($M = 36.38$, $SD = 10.349$ compared to $M = 34.40$, $SD = 10.719$).

For the independent samples $t$-test, tests of normality were carried out and the distributions for both groups were normal. The $t$-test results revealed that there was no statistically significant difference between the two groups ($p = .372$).
4.2. Integrated EAP listening-into-writing test

To answer RQ1, Fair Average scores and Rater Infit and Outfit were calculated using Facets. The Fair Average scores were used for the Hotelling’s $T^2$ test to investigate whether there were any statistically significant differences in performance across the two groups when each group was presented with the two types of input but in a different input order.

For scoring, the summaries were divided into two halves, representing the audio input in one half and the video input in the other half. Each half of the summary was allocated a CEFR level score and this was converted to a number format as follows:

- A0\(^8\): 1
- A1: 2
- A2: 3
- B1: 4
- B2: 5
- C1: 6

### 4.2.1. Facets analysis

The Facets analysis revealed that the three raters were consistent within themselves (see Table 4 below) with the Infit Mn Sq all falling within the 0.5 to 1.5 range that Linacre (2012) considers 'productive for measurement' (p. 11). Similarly, the Outfit MnSq values were also within an acceptable range.

Finally, the number of exact agreements was 147 (50%) compared to an expected number of agreements of 151.2 (51.4%), which suggests that raters were acting as independent experts, whereby rater 2 was the most lenient (Fair Average: 4.35) and rater 1 was the harshest (Fair Average: 4.04).

### 4.2.2. Hotelling’s $T^2$ analysis

Preliminary assumption checking revealed that the data for each group were not normally distributed as assessed by the Shapiro-Wilk’s test ($p < .05$). However, it should be noted that a MANOVA is ‘relatively robust to violations of the assumptions in many circumstances’ (Bray & Maxwell, 1985, p. 33); inspection of the boxplots showed that there were four univariate outliers in the scores for the first half of the lecture (audio input) for the ‘audio first’ group but there were no outliers in the ‘video first’ group; Mahalanobis distance revealed that there were no multivariate outliers in the data ($p > .001$); there were linear relationships, as assessed by scatterplot, and no multicollinearity ($|r|<0.9$); finally, there was homogeneity of variance-covariance matrices, as assessed by Box’s test of equality of covariance matrices ($p = .055$).

The results of the Hotelling’s $T^2$ analysis demonstrate that, despite the fact that there were no significant differences between the two groups on the QPT scores, the ‘audio first’ group scored more highly for both input formats (see Table 5 below).

**Table 3**

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Min</th>
<th>Max (out of 60)</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio first</td>
<td>36.38</td>
<td>10.349</td>
<td>18</td>
<td>58</td>
<td>37</td>
</tr>
<tr>
<td>Video first</td>
<td>34.40</td>
<td>10.719</td>
<td>16</td>
<td>55</td>
<td>32.50</td>
</tr>
</tbody>
</table>

$^8$ There is no A0 level in the CEFR but this grade was assigned to summaries which were deemed to be below A1.
Table 4
Rater fit statistics.

<table>
<thead>
<tr>
<th>Rater</th>
<th>Infit MnSq</th>
<th>Outfit MnSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td>.85</td>
<td>.86</td>
</tr>
<tr>
<td>Rater 2</td>
<td>1.08</td>
<td>.96</td>
</tr>
<tr>
<td>Rater 3</td>
<td>1.05</td>
<td>.78</td>
</tr>
</tbody>
</table>

Table 5
Mean scores for each group and input format.

<table>
<thead>
<tr>
<th>Group</th>
<th>First half of the lecture</th>
<th>Second half of the lecture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Audio input: $M = 4.006$, $SD = 1.246$</td>
<td>Video input: $M = 3.571$, $SD = 1.340$</td>
</tr>
<tr>
<td>'Audio first' group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>'Video first' group</td>
<td>Video input: $M = 3.452$, $SD = 1.212$</td>
<td>Audio input: $M = 2.886$, $SD = 1.444$</td>
</tr>
</tbody>
</table>

Fig. 1. Performance by group.

Fig. 2. Performance by input format.

Table 6
Performance breakdown by group and summary half.

<table>
<thead>
<tr>
<th>Group</th>
<th>Number performing equally well on both halves of the lecture</th>
<th>No. of test takers performing better on the first half of the lecture</th>
<th>No. of test takers performing better on the second half of the lecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both groups combined</td>
<td>46</td>
<td>57</td>
<td>13</td>
</tr>
<tr>
<td>Audio first</td>
<td>25</td>
<td>25</td>
<td>8</td>
</tr>
<tr>
<td>Video first</td>
<td>21</td>
<td>32</td>
<td>5</td>
</tr>
</tbody>
</table>
half of the input (video input). In the ‘video first’ group, only approximately 36% (n = 21) scored equally well on both halves of the summary whereas 55% (n = 32) performed better on the summary relating to the first half of the input (video input). Only approximately 9% (n = 5) performed better on the summary relating to the second half of the input (audio input).

4.3. Feedback questionnaire

Participants were asked to give reasons for why they thought they had performed better on the summary relating to one half of the lecture input than the other. Table 7 below shows the responses for this question.

As can be seen, 32.8% (n = 38) of respondents felt that they had been assisted by having access to the text/visual. An additional 7.8% (n = 9) felt that the video was more understandable. Conversely, 15.5% (n = 18) felt that listening only was easier. 26 test takers did not respond to this question.

Some test takers felt the video helped them because they could:
- ‘see slides in the presentation’ (2SA3)
- ‘both listen and read information’ (SV10)

Test taker, 2SV8, explained the usefulness of the visuals clearly:
- ‘there were slides that improved the understanding and because visual perception helped to focus’.

On the other hand, one of the test takers who expressed a preference for audio only input stated:
- ‘I was focused only on listening’ (KA6)

while another one felt they performed better when presented with the audio input:
- ‘because there was not any distraction’ (SSA2).

When asked for the reasons why they felt they performed worse on one part of the task than the other (Q.8) (see Table 8 below), again 26 test takers did not respond. However, 20.7% (n = 24) felt that listening only was more difficult. Conversely, 13.8% (n = 16) felt that both watching and listening were difficult and/or they felt distracted by the video while 17.2% (n = 20) said they did not understand the content.

In contrast to the comment from test taker KA6 above, one test taker felt they performed worse on the audio half of the input because:
- ‘it don’t concentration (sic) my attention’ (2SA4).

Despite the fact that the input was almost exactly the same in terms of difficulty, one test taker who was presented with the video first felt that, when it came to the audio:
- ‘the audio information is much more difficult’ (SV9).

4.4. Summary of results

This study attempted to shed light on three research questions:

1. How do input order and format (audio vs video) affect written performance on an integrated listening-into-writing EAP task?

Table 7

<table>
<thead>
<tr>
<th>Perceived reasons for better performance</th>
<th>Percentage of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text/visual helped me</td>
<td>33</td>
</tr>
<tr>
<td>No response</td>
<td>22</td>
</tr>
<tr>
<td>Listening only was easier</td>
<td>15</td>
</tr>
<tr>
<td>Video was more understandable</td>
<td>8</td>
</tr>
<tr>
<td>I understood the content better (second half of the lecture)</td>
<td>6</td>
</tr>
<tr>
<td>Incomprehensible/miscellaneous response</td>
<td>6</td>
</tr>
<tr>
<td>I don’t know</td>
<td>4</td>
</tr>
<tr>
<td>I understood the content better (first half of the lecture)</td>
<td>3</td>
</tr>
<tr>
<td>Interesting topic</td>
<td>2</td>
</tr>
<tr>
<td>Clear presentation structure</td>
<td>1</td>
</tr>
</tbody>
</table>
With regard to test takers’ perceptions of their performance, what reasons do test takers give for why they felt they performed better when presented with one input format rather than the other?

With regard to test takers’ perceptions of their performance, what reasons do test takers give for why they felt they performed worse when presented with one input format rather than the other?

The Hotelling’s $T^2$ analysis revealed that the mean scores in both groups were lower for the second half of the input than for the first half of the input and that the mean scores for ‘video first’ group were lower than for the ‘audio first’ group on both halves of the input. This resulted in a statistically significant difference between the two groups on the combined variables. Similarly, there was a statistically significant difference between the two groups in performance by input format when test takers were presented with the audio input, yet there was no statistical difference when test takers were presented with the video input.

In terms of test taker preferences, 40.6% of test takers in total (n = 49) felt that they had been assisted by access to the text/visual and that the video was more understandable whereas only 15% (n = 18) felt that the audio only input was better.

### 5. Discussion

This study has built on previous studies relating to input formats on discrete listening tests by investigating the impact of input formats and input order on an integrated listening-into-writing task.

First of all, with regard to the validity of the scoring, the results of the Facets analysis demonstrate that all three raters were within 0.5 logits of each other according to the Fair Average scores. This was most likely due to the fact that the three raters involved in the study had worked together at the same institution over many years and had used a rating scale based on the CEFR in their daily work. The fact that they underwent specific rater training for this particular task would have also contributed to their shared understanding of the performance levels required for a given CEFR level and thus the close agreement achieved. Nonetheless, it is important to note that, although they were very close, they were still different enough to be acting as independent raters.

The Hotelling’s $T^2$ analysis revealed that, when presented with input in two different formats (audio vs video), it appears that test takers perform better on a follow-up writing task when the first input format that they are exposed to is audio rather than video. As pointed out by some test takers, access to the video may have caused them to watch the video rather than focussing on taking notes so this may have been a distraction, thus resulting in fewer notes and less to write about.

When the ‘audio first’ group was exposed to the second half of the input (video), the mean score was half a CEFR level lower than the score for the first half of the input (audio) ($M = 4.006$ in the summary for the first half compared to $M = 3.571$ in the summary for the second half). This could have been due to test takers also being distracted by the video or due to fatigue among some test takers. However, what is interesting is that the mean scores on the video input were similar across both groups ($M = 4.452$ for the ‘video first’ group and $M = 3.571$ for the ‘audio first’ group) such that there was no statistically significant difference between the groups. This suggests that test takers perform similarly when exposed to video input irrespective of whether this is the first type of input they are exposed to or not.

On the other hand, the ‘video first’ group performed worse on the audio only input - their second input format - with the mean score falling by .7 of a CEFR level (from $M = 3.452$ for the summary relating to the first half of the input to $M = 2.866$ for the summary relating to the second half). This was 1.14 CEFR levels lower than the mean score for the ‘audio first’ group ($M = 4.006$) and was statistically significant. One possible explanation for this could be that the language in the two halves of the lecture varied in difficulty; however, care was taken to ensure that both halves of the lecture were as close as possible in terms of the word length, text complexity and duration. In fact, the second half of the input was slightly shorter (see Appendix 2) but was of almost exactly equal difficulty: the Flesch-Kincaid Grade Level, as assessed by Text Inspector, for the first half of the input was 59.61 compared to 59.43 for the second half of the input while lexical diversity (MTLD) for the first half was 39.76 compared to 40.25 for the second half. As such, it is unlikely that the text itself was the cause of the differing results for the audio-only input.

Another possible reason could be that test takers who were exposed to the video input first relied on the visual support (text,
research was generally based on listening tests rather than integrated listening-into-writing tests, and in some cases, the video input while the lower level learners perform better with audio (Pardo-Ballester, 2016; Chen et al., 2014). It should be noted that previous Other studies have found that higher proficiency learners perform better on discrete listening tests when they have access to video (2015) found no significant interaction between delivery format and text type, nor between proficiency level and delivery format.

A small number of test takers (n = 9) did comment that they would have liked more time, yet feedback did not suggest that this was a major issue in this study.

The findings of the current study are, to some extent, in line with Suvorov’s (2008) findings because his test takers performed significantly worse with video compared to audio only and audio with a single photograph. However, other research into input formats has yielded mixed findings. Gruba (1993) did not find any significant difference between audio and video in his study. Similarly, Batty (2015) found no significant interaction between delivery format and text type, nor between proficiency level and delivery format. Other studies have found that higher proficiency learners perform better on discrete listening tests when they have access to video while the lower level learners perform better with audio (Pardo-Ballester, 2016; Chen et al., 2014). It should be noted that previous research was generally based on listening tests rather than integrated listening-into-writing tests, and in some cases, the video input included only context visuals rather than content visuals so this may account for the differences in the findings to some extent.

Irrespective of performance, there appears to be a greater preference among test takers for the use of video than for audio-only input with over 40% stating a preference for video since they could use the visual stimulus to aid comprehension and concentration. The greater preference for video concurs with other studies (Pardo-Ballester, 2016; Progosh, 1996; Suvorov, 2008; Wagner, 2010) and the comments in support of the use of textual input are in line with findings from Chang et al. (2011) and Montero Perez et al. (2014). In contrast, only around 15% preferred audio, with some test takers stating that the video was distracting. This appears to be in line with other research which has compared the two formats (Coniam, 2001; Gruba, 1994; Suvorov, 2013).

As educators, we need to consider what the implications of these findings are for teaching and testing academic listening-into-writing. Vandergrift (2004) argues that ‘students need to learn to rely on the acoustic signal and relevant contextual factors to develop listening strategies’ (pp. 10–11). While this may be true, many lectures include written support in the form of a PowerPoint presentation. Moreover, many online materials include captioned video (Winke et al., 2010). Similarly, many online lectures (whether pre-recorded or recorded live then uploaded) will also often have access to an accompanying transcript in addition to any visual support included in the lecture.

Referring to Mayer’s (2001) Cognitive Theory of Multimedia Learning, Leveridge and Yang (2013) state that, as audio input becomes too difficult, the L2 listener turns to the visual input mode, as this may be more easily understood. Thus, L2 listeners who do not have the listening ability required for the task may resort to reading rather than listening, as may have been the case in the current study. There are, of course, construct implications to this – the listening construct may change to a reading construct (at least, partially)
– but this reflects the TLU domain, in which learners do have access to textual and/or visual support as well as the transcript. Therefore, this additional support can help lower-level learners to understand the input and, consequently, one hopes, perform better on their written assessments. Batty (2015) recognises not only the authenticity but also the increased face validity that audiovisual input affords and therefore argues in favour of the use of video.

In addition to helping linguistically weaker L2 learners, there are other equality and accessibility advantages of using video. Learners who may be hard-of-hearing can also use the textual input and read the speaker’s lips to aid comprehension.

6. Conclusions and implications

This study has implications for both teaching and testing integrated listening-into-writing in an EAP environment which we need to consider when deciding whether to make content available for remote, blended or flipped learning.

As EAP lecturers, I would argue, our aims are twofold: one is to develop students’ listening skills so that they are able to integrate into the academic environment, both in and outside of class; the other is to prepare them for the real-world academic environment, where they may often have access to audiovisual input. Therefore, the way we present information to our students and assess them should serve both of these aims.

Video-mediated input can provide valuable scaffolding in lectures to assist learners in understanding the content of the input; however, for some, this may also lead to cognitive overload so we need to bear this in mind. To aid automaticity of processing and reduce the cognitive load, we can raise awareness of the discourse features used to structure different types of input. We can also help students to relate the spoken input to the visual input by encouraging them to compare and contrast the content of the audio and visual input. This can be done by encouraging students to ‘notice’ how what is written on the PowerPoint slides differs from what the speaker actually says. To assist with decoding, students can consider how the graphological form of a word which may appear on screen compares to its pronunciation.

However, we must also be aware of the potential negative impact that audiovisual input may have on listening skills over the longer term if learners become accustomed to being able to focus solely, or primarily, on the written word. Test taker questionnaire data revealed that some test takers felt that the audio only input was too fast, when, in actual fact, the speech rate was almost exactly the same in both input formats. This is possibly due to the additional effort of having to process the input in real-time without the support of the visual. The findings of the current study also suggest that losing the visual support has a greater impact on performance than when test takers are ‘accustomed’ to audio only input and are then provided with the additional support offered by the visual stimuli. Therefore, we should also provide targeted exercises to develop students’ listening/decoding skills as this will also aid automaticity of processing and decrease cognitive load. This can be done by raising awareness of how pronunciation changes in connected speech and when vowels are stressed or unstressed. Thus, providing audio-only input, which can be used explicitly for such purposes and to teach general listening comprehension skills that learners can use in their interactions outside of a lecture situation, is also vitally important.

The impact of fatigue should not be underestimated. It is recognised that a video or audio text which serves as content input for a flipped or blended learning lesson can be watched as often as necessary and at different times so learners are not limited to listening/watching everything in one sitting. Nonetheless, teaching input should not be too long – 6 min is recommended by EdX for an online lesson (Inman & Myers, 2018). Guo et al. (2014) observed a drop-off in engagement between 9 and 12 min and this study appears to support those findings as attention appears to have dropped off when exposed to the second half of the lecture, that is, between 6 and 12 min in both groups. In a testing environment, of course, the issue of fatigue is an important consideration as test takers may or may not have control over the input. In this case, fatigue appears to have been greater when test takers had become ‘accustomed’ to the visual input in the first half of the lecture and were then exposed to audio only in the second half. Of course, this may have been exacerbated by the change of format and, clearly, one would not change the format of the input mid-test; however, the effect may well be the same if we use audiovisual input in our teaching and audio only in a test. Therefore, our students need to be exposed to and comfortable with both input formats.

Although several researchers have claimed that 10–20 min is a rule of thumb for lecture input (Bradbury, 2016), it would be useful to know whether age and L2 proficiency levels affect concentration levels, particularly in an EAP environment. Similarly, further research could investigate whether there is a difference in concentration levels between recorded and live lectures since individual lecturer traits and personalities are also likely to affect engagement levels, perhaps more than lecture length (Bradbury, 2016). This would be particularly timely given the changes brought about by the global pandemic and the move to more online or blended teaching. One important point to consider here is that recorded input enables listeners/viewers to replay or rewind as often as they wish. They can also stop the recording at any time to take a break. Therefore, research into the impact of individual control over pre-recorded lecture input and how this might affect performance in a time-constrained assessment would also be valuable. However, we should bear in mind that, if input is provided as a video, this provides scaffolding and may help to maintain concentration thereby possibly reducing the need for repetition or pauses.

To conclude then, this study has attempted to address a gap in the research by bringing together research on input formats in (academic) listening and integrated listening-into-writing assessment. The findings revealed that there is less of a drop-off in performance when a video is used and there seems to be more of a preference for video-based input than audio only input. On the basis of these findings, it seems wise to suggest that video-mediated input is an appropriate way to provide input both for teaching and testing. This also reflects the TLU domain and is therefore more authentic. Conversely, since some test takers may be distracted by the video and bearing in mind the impact of visual input on the construct of academic listening, it could be argued that audio only should be used. I would argue that a combination of input types should be used since the video can provide support for important content while audio can be useful in training students’ listening skills. However, I should stress the need for further research to help us fully

C. Westbrook
understand the role of input formats and the extent to which they can benefit students in understanding input which they need for a follow-up integrated task.

Funding

The author acknowledges the role of the British Council in making this study possible: The British Council provided the research grant which enabled me to conduct the study as part of the ARAGs 2017 programme. I would like to express my deep gratitude for this support.

Any opinions, findings, conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the British Council, its related bodies or its partners.

Author statement


Declaration of competing interest

None.

Acknowledgement

I would like to thank the editors, Sarah Brewer and Olwyn Alexander, the anonymous reviewers, and Dr Karen Dunn and Professor Barry O’Sullivan OBE for their advice in preparing this paper and for their invaluable feedback on earlier drafts.

Appendix 1. Main study - Information sheet and consent form

Introduction to the research

This doctoral research will investigate the relationship between input task characteristics (video vs audio input) of listening comprehension texts, and performance on written output tasks.

The objectives are:

- to investigate how test takers perform when input is presented in audio format only compared to video format;
- to investigate the extent to which test takers are influenced by the written word in an EAP lecture compared to the spoken word.

Methodology

The research questions are as follows:

- Do test takers perform better when presented with video input for a lecture?
- To what extent do test takers rely on the written word compared to the spoken word?

The aim of the study is to find out how test takers’ performance on an integrated EAP test varies when the input (i.e. the lecture content) is presented in audio and video formats.

To assess the differences in performance, the data will be assessed qualitatively and quantitatively if numbers allow.

All data will be anonymised and informed consent will be sought from all participants.

The study comprises three steps:

- quick lexico-grammatical placement test (approx. 30 mins)
- language test comprising an academic lecture and a follow-up writing task (approx. 80 mins);
- brief post-task questionnaire (approx. 5 mins);

Test format

A specifically developed integrated EAP listening and writing task. Test-takers will watch/listen to a short EAP lecture (on a general academic topic) presented in audio only and video formats and produce a thematically-linked written summary. The same auditory input will be provided in each case (90 min).

The research will be conducted as follows: one x 30 min followed by one x 1.5 h session.
Information sheet for participants – main study

First of all, thank you for showing an interest in participating in this study, which will be the basis of the researcher’s PhD thesis. Participation is completely voluntary and you may withdraw at any time.

Please see the information sheet provided for information about the specific tasks.

The data you provide (your test results, post-test questionnaires and, where applicable, the interview data) will be analysed to see if the type of input has an effect on your performance.

However, please note that the performance on these tasks does not affect your academic course in any way.

You are more than welcome to find out your test result. If you wish to do that, please send me an e-mail: XXXXXXX.

The initial findings from this research were presented at the XXXXX. It is intended that the finished research will be published in papers and journal articles as well as through the researcher’s PhD thesis. If you wish to have a copy of the results of the study, please send me an e-mail at the above address, too.

If you are willing to participate, please complete and sign the consent form overleaf and return it to the tutor/researcher.

Consent form for participants

First of all, thank you for agreeing to participate in this study, which will be the basis of the researcher’s PhD thesis.

Please read the consent information below and tick the boxes to confirm your agreement. Then please sign your name below and write your name clearly.

I confirm that:

I have been given clear and detailed information about the study I will be involved in.
I understand that participation is completely voluntary and I may withdraw at any time.
I agree to my data being used for the purposes of this study.
I am aware that all data will be anonymised and any personal data will be treated as confidential.
I am aware of how the results of this study will be disseminated (= passed on to other people) and I agree to this.
I understand that I may request a copy of my results and of the results of the study by sending an e-mail to the researcher and that I have been given the researcher’s e-mail address.

Signed: ______________________________________________
Name: (please print) ______________________________________________

THANK YOU FOR YOUR PARTICIPATION AND YOUR HELP. THEY ARE GREATLY APPRECIATED!

Appendix 2. Text Inspector analysis of lecture language and slides

<table>
<thead>
<tr>
<th>Operation</th>
<th>Data</th>
<th>Lecture First half</th>
<th>Lecture Second half</th>
<th>Lecture Slides First half</th>
<th>Lecture Slides Second half</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics</td>
<td>Sentence count</td>
<td>34</td>
<td>31</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Statistics</td>
<td>Token count</td>
<td>1108</td>
<td>1048</td>
<td>54</td>
<td>61</td>
</tr>
<tr>
<td>Statistics</td>
<td>Type count</td>
<td>290</td>
<td>277</td>
<td>42</td>
<td>44</td>
</tr>
<tr>
<td>Statistics</td>
<td>Type/token ratio</td>
<td>0.26</td>
<td>0.26</td>
<td>0.78</td>
<td>0.72</td>
</tr>
<tr>
<td>Statistics</td>
<td>Avge syllables per sentence</td>
<td>43.97</td>
<td>45.19</td>
<td>54.5</td>
<td>36.67</td>
</tr>
<tr>
<td>Statistics</td>
<td>Avge syllables per word</td>
<td>1.35</td>
<td>1.34</td>
<td>2.02</td>
<td>1.8</td>
</tr>
<tr>
<td>Statistics</td>
<td>Flesch Reading Ease</td>
<td>59.61</td>
<td>59.43</td>
<td>8.66</td>
<td>33.64</td>
</tr>
<tr>
<td>Statistics</td>
<td>Flesch-Kincaid Grade</td>
<td>13.04</td>
<td>13.37</td>
<td>18.76</td>
<td>13.62</td>
</tr>
<tr>
<td>Statistics</td>
<td>Average Sentence Length</td>
<td>32.59</td>
<td>33.81</td>
<td>27</td>
<td>20.33</td>
</tr>
<tr>
<td>Lexical Diversity</td>
<td>Lexical diversity (VOCD)</td>
<td>71.86</td>
<td>62.8</td>
<td>71.16</td>
<td>54.36</td>
</tr>
<tr>
<td>Lexical Diversity</td>
<td>Lexical diversity (MTLD)</td>
<td>39.76</td>
<td>40.25</td>
<td>44.27</td>
<td>59.14</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>A1 type %</td>
<td>45.97</td>
<td>52.45</td>
<td>28.57</td>
<td>51.16</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>A2 type %</td>
<td>20.47</td>
<td>16.43</td>
<td>9.52</td>
<td>2.33</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>B1 type %</td>
<td>10.07</td>
<td>13.99</td>
<td>19.05</td>
<td>25.58</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>B2 type %</td>
<td>8.72</td>
<td>8.74</td>
<td>16.67</td>
<td>0</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>C1 type %</td>
<td>2.01</td>
<td>1.4</td>
<td>4.76</td>
<td>11.63</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>C2 type %</td>
<td>0.67</td>
<td>0</td>
<td>2.38</td>
<td>0</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>Known Words type %</td>
<td>0.67</td>
<td>0.7</td>
<td>4.76</td>
<td>4.65</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>Unlisted type %</td>
<td>11.41</td>
<td>6.29</td>
<td>14.29</td>
<td>4.65</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>A1 token %</td>
<td>64.09</td>
<td>67.71</td>
<td>28.3</td>
<td>48.28</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>A2 token %</td>
<td>13.76</td>
<td>11.54</td>
<td>15.09</td>
<td>1.72</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>B1 token %</td>
<td>4.71</td>
<td>5.06</td>
<td>15.09</td>
<td>18.97</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>B2 token %</td>
<td>3.68</td>
<td>3.44</td>
<td>13.21</td>
<td>0</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>C1 token %</td>
<td>0.85</td>
<td>0.91</td>
<td>3.77</td>
<td>8.62</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>C2 token %</td>
<td>0.19</td>
<td>0</td>
<td>1.89</td>
<td>0</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>Known Words token %</td>
<td>3.3</td>
<td>6.07</td>
<td>11.32</td>
<td>18.97</td>
</tr>
<tr>
<td>Lexis: EVP</td>
<td>Unlisted token %</td>
<td>9.43</td>
<td>5.26</td>
<td>11.32</td>
<td>3.45</td>
</tr>
</tbody>
</table>

(continued on next page)
References


---

### Table: Notetaking Strategies and Their Relationship to Performance on Listening Comprehension and Communicative Assessment Tasks

<table>
<thead>
<tr>
<th>Operation</th>
<th>Data</th>
<th>Lecture First half</th>
<th>Lecture Second half</th>
<th>Lecture Slides First half</th>
<th>Lecture Slides Second half</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexis: AWL</td>
<td>AWL All Types %</td>
<td>5.83</td>
<td>7.48</td>
<td>11.9</td>
<td>15.56</td>
</tr>
<tr>
<td>Lexis: AWL</td>
<td>AWL All Tokens %</td>
<td>4.98</td>
<td>8.2</td>
<td>12.96</td>
<td>18.03</td>
</tr>
</tbody>
</table>

---

16