

Vocabulary for Listening: Emerging Evidence for HIGH and mid-frequency Vocabulary Knowledge

M. Rahim Bohlooli Niri

Assistant professor, Department of English Language, Farhangian University, Zanjan, Iran.

Abstract: *This article presents empirical evidence aimed at informing approaches to vocabulary development for the purpose of supporting L2 listening comprehension. Inferential statistics were used to analyze the relationship between second language (L2) aural vocabulary knowledge (AVK), L2 listening comprehension and the overall L2 proficiency among 247 tertiary level L2 learners of English. Three frequency based levels of AVK were measured. Measures of level 1 (0e2000 frequency range) and level 2 words (2001e3000 frequency range) tapped AVK of high frequency words, and level 3 words (3001e5000 frequency range) tapped AVK of words just beyond the high frequency range. Listening comprehension was measured with a version of the International English Language Testing System (IELTS). Regression modeling showed that AVK at each of the three levels contributed uniquely to the prediction of L2 listening for the entire cohort. Only measures of level 2 and 3 AVK were uniquely predictive of L2 listening for a relatively high proficiency subgroup, whereas only level 1 AVK offered a unique contribution to the prediction of L2 listening scores for a relatively low proficiency subgroup. Results are interpreted in order to provide a range of pedagogical recommendations.*

Keywords: *Second language listening. Listening comprehension. Vocabulary. Aural vocabulary knowledge*

INTRODUCTION

Successful listening comprehension depends on a language user's knowledge of L2 words (Vandergrift and Baker, 2015). Words represent the lowest level of representation at which a stable connection between the form and the meaning of a word can be reliably established (Hulstijn, 2002); therefore, once words are recognized and associated with their literal meanings larger semantic units can be built in the mind of the listener. Key to the listening comprehension process is the rapid and appropriate association of such semantic units (linguistic knowledge) with the listener's pre-existing schemata (non-linguistic knowledge).

The ability to recognize, understand and interpret strings of words in connected speech typically poses little difficulty for native speakers, as throughout their lifetime native speakers have been exposed to an immense quantity of contextualized spoken input (Ellis, 2002). This is not the case for many L2 listeners (Field, 2008a). Without having encountered the equivalent degree of exposure to the target language characteristic of native speakers, L2 learners typically have sub-optimal aural vocabulary knowledge; that is learners have difficulty recognizing words in the spoken form. Learners may even have difficulty in recognizing words in speech which they know when presented in the written form (Goh, 2000).

Despite the importance of being able to recognize words from speech for L2 listening, and the difficulty word recognition causes L2 listeners (Field, 2008b; Goh, 2000), the relationship between aural vocabulary knowledge and L2 listening comprehension has received limited interest from researchers (Broersma and Cutler, 2008). Of particular interest to the current study are constructs of vocabulary knowledge shown to be

strongly associated with L2 listening comprehension success; for example, constructs of vocabulary knowledge which involve “the ability to recognize the phonological [aural] form of the word, access existing knowledge of that word and produce a representation of it under time constraints” (Matthews and Cheng, 2015: p. 4). The limited amount of research which has investigated the link between such constructs of L2 vocabulary knowledge and listening comprehension is of significance as the development of L2 vocabulary knowledge is likely to be an important component of L2 listening pedagogy (Vandergrift and Baker, 2015).

There is currently a relatively small body of empirical research which can guide evidence based approaches to the development of L2 vocabulary for listening (Stæhr, 2009; van Zeeland and Schmitt, 2013). However, there are key areas of research which have yet to be adequately investigated. In order to develop evidence based frameworks for L2 vocabulary for listening programs, a stronger understanding of the interaction between L2 aural vocabulary knowledge, L2 listen in comprehension and the overall L2 proficiency evident within cohorts of language learners is required. However, little previous research has investigated the relationship between these three constructs in real learning contexts. The current research seeks to begin filling this gap in the literature.

Literature Review

Which aspect of vocabulary knowledge is the most important for L2 listening comprehension?

When compared to L2 reading, far less research has focused on the relationship between vocabulary knowledge and L2 listening (Stæhr, 2008). Previous studies which have investigated the link between word knowledge and L2 listening demonstrate that vocabulary knowledge does have a strong link with L2 listening. However, several of these studies have not focused on the forms of vocabulary knowledge which are of strongest value to L2 listening comprehension (Bonk, 2000; Stæhr, 2008, 2009);

For example, Stæhr (2009) demonstrated that L2 vocabulary knowledge was strongly correlated with listening comprehension and could explain 51% of the variance observed in the listening test scores of a cohort of 115 Danish learners. However, each of the vocabulary tests used measured knowledge of the form-meaning connection in the written form, which did not tap the test takers ability to recognize words in their spoken form. Further, the tests used in the study did not measure the capability of the test taker to access word knowledge under time constraints. This limitation is also significant, as the ability to rapidly process language in real time is crucial for successful L2 listening comprehension (Goh, 2000; Hulstijn, 2003).;

As Stæhr (2009) acknowledged, any future study which examines the link between vocabulary knowledge and L2 listening “should ideally be based on a vocabulary test that involves hearing the target words rather than reading them” (p. 597)

Previous research findings have made clear that the relationship between written and aural word knowledge is not fixed (Milton and Hopkins, 2006) and that knowledge of a word in its orthographic form does not guarantee the capability to recognize that word when presented in spoken language (Goh, 2000). A growing appreciation of the specificity and importance of aural vocabulary knowledge is reflected by recent research on testing this construct (McLean et al., 2015), and by research which compares the relative contribution of written and aural forms of vocabulary knowledge on L2 listening comprehension. For example, Milton, Wade, and Hopkins (2010) administered two equivalent computerized tests which measured knowledge of English words in their spoken and written form among a group of 30 test subjects. Results indicated that participants’ aural vocabulary knowledge was more strongly correlated with, and predictive of, L2 listening comprehension ($r = 0.67$, $p < 0.05$) than were measures of written vocabulary knowledge ($r = 0.48$, $p < 0.01$). Although the specificity of aural vocabulary knowledge for L2 listening is strongly suggested in the results presented in Milton et al. (2010), the relatively small sample size limits the degree to which these results can be used to generalize beyond the study cohort. In a larger scale study, Cheng and Matthews (2018) investigated the relative strength of association between L2 listening comprehension and measures of both written and aural vocabulary knowledge among a cohort of 250 L2 learners. Results.

showed that aural vocabulary knowledge was more strongly associated with listening ($r = 0.71$, $p < 0.001$) than was written vocabulary knowledge ($r = 0.55$, $p < 0.001$). Factor analysis undertaken as part of the study showed that measures of aural and written vocabulary knowledge loaded onto discrete factors suggesting that these measures tapped distinct constructs.

Although few in number, previous studies which have investigated the strength of association between vocabulary knowledge and L2 listening suggest that aural vocabulary knowledge, that is the ability to recognize words as they are presented in spoken language, is more important to L2 listening comprehension than other forms of vocabulary knowledge.

In light of these findings, research aimed at investigating the relationship between word knowledge and L2 listening should emphasize aural vocabulary knowledge as the primary construct of relevance.

Which lexical items are most important for L2 listening comprehension?

In order to guide L2 vocabulary for listening programs, it is also necessary to identify the groups of words which are most strongly supportive of L2 listening comprehension. Knowing the lexical coverage needed for adequate levels of L2 listening comprehension provides important insight as it enables the calculation of the theoretical vocabulary size needed to comprehend a given text (van Zeeland and Schmitt, 2013).

Adolphs and Schmitt (2003) analyzed the Cambridge and Nottingham Corpus of Discourse in English (CANCODE), a corpus of spoken language containing approximately 5 million running words, and determined that knowledge of the most frequent 2000 word families would provide a lexical coverage of 94.76%, and knowledge of the most frequent 3000 words would provide 95.91% lexical coverage. Nation's (2006) smaller scale investigation concluded that for 98% lexical coverage of spoken texts, knowledge of the most frequent 6000-7000 word families was required. Webb and Rodgers (2009a, 2009b) analyzed a corpus of the lexical content of television programs and movies and determined that the most frequent 3000 word families provided over 95% coverage. However, a greater range of vocabulary knowledge was required to reach this level of coverage when analysis was undertaken for texts categorized within their respective genres. For example, for lexical coverage of 95% for spoken text analyzed within television and movie genre categories, knowledge of the most frequent 4000 word families would be required.

Other studies have adopted a more direct approach to investigate the lexical coverage necessary for L2 listeners to adequately understand spoken text. For example, van Zeeland and Schmitt (2013) asked forty participants to listen to four adapted spoken texts for which participants possessed 100%, 98%, 95% and 90% lexical coverage. Based on their results, for high levels of comprehension, 98% lexical coverage, or knowledge of the most frequent 6000-7000 word families is suggested, whereas if lesser levels of comprehension are adequate then 95% lexical coverage, or knowledge of the most frequent 2000-3000 word families, may be a more suitable target.

Matthews and Cheng (2015) sought to establish the relationship between aural word knowledge and L2 listening comprehension among 167 L2 tertiary level learners. Aural vocabulary knowledge was measured with a partial dictation test which targeted knowledge of words from the first, second and third thousand frequency levels and required test takers to produce a single target word after hearing a stimulus sentence once. Results indicated a strong and positive correlation between mean aural vocabulary knowledge and L2 listening comprehension ($r = 0.73$, $p < 0.01$). Additionally, regression modeling indicated that measures of vocabulary knowledge were predictive of L2 listening comprehension scores. Aural vocabulary knowledge for words in the 200-3000 frequency range alone was able to predict 52% of the variance observed in L2 listening scores. The study highlights the importance of knowledge of high frequency vocabulary in L2 listening; however, a significant limitation of the study is that it did not measure knowledge of words beyond the 3000 frequency range.

The contribution of vocabulary knowledge on L2 listening for low and high L2 proficiency

Although previous research has provided preliminary evidence for the importance of vocabulary knowledge in relation to L2 listening performance, few studies have addressed the question of how vocabulary knowledge

may impact on L2 listening comprehension for learners with relatively low and relatively high L2 proficiency levels. In a less recent yet important study, Tsui and Fullilove (1998) provide indirect evidence that knowledge of words in connected speech is an important factor that differentiates L2 listeners of low and high proficiency. This large study involved over 20,000 ESL learners answering listening comprehension test items, which were of two broad categories. One category of listening test items could be correctly answered through strategic application of appropriate contextual schema. The second category of test item could not be correctly answered solely through application of an appropriate schema, but instead demanded that test takers rapidly and accurately process specific information encoded at the word level within the spoken test stimulus. Results indicated that test items which demanded the ability to process the spoken language at the lower levels of linguistic processing (bottom up processing) were those which were most useful in differentiating test takers of relatively low and high L2 listening proficiency. Tsui and Fullilove's (1998) research is suggestive that lexical knowledge of words in the aural form is an important construct that may be indicative of L2 listening performance of those with higher levels of L2 proficiency. Bonk (2000) also provides evidence for the importance of word knowledge as an indicator of higher L2 listening proficiency.

Investigating the relationship between lexical familiarity and L2 listening comprehension among 50 Japanese ESL learners, it was determined that learners with measures of lexical familiarity lower than 80% were very unlikely to attain adequate levels on an L2 listening test. Although these results again suggest that word knowledge is an important attribute of L2 listeners that perform better on L2 listening comprehension tests, the results overall were not clear-cut. Among the group, Bonk (2000) reported that some learners with excellent measures of word familiarity achieved inadequate L2 listening comprehension scores. Conversely, results also indicated that some participants were able to achieve good levels of listening comprehension scores, despite depressed scores on the word familiarity test. Overall, Bonk (2000) provides another line of evidence which suggests L2 vocabulary knowledge is an important factor which contributes to the L2 listening success of those with relatively high L2 proficiency level. However, the ambiguities evident in the results encourage further research in this field.

The current study

The studies addressed in the literature review highlight a number of issues which motivate the current study. Firstly, it is important to note that aural vocabulary knowledge, that is knowledge of words mediated through the aural modality, is likely to be the most important construct of vocabulary knowledge in relation to L2 listening comprehension success (Milton et al., 2010).

Despite the established link between aural vocabulary knowledge and L2 listening, relatively few studies have directly measured this construct of vocabulary knowledge and related these measures to L2 listening comprehension in real language learning contexts. Secondly, the review encompasses studies which have put forward a broad range of suggested minimum word knowledge requirements for L2 listening. These suggestions range from knowledge of the most frequent 2000 word families (van Zeeland and Schmitt, 2013, p. 475) up to knowledge of words within the 7000 frequency range (Nation, 2006, p.79).

Despite the broad range of recommendations put forward, there is a scarcity of studies which have measured both the aural vocabulary knowledge and the L2 listening comprehension ability among relatively large cohorts of L2 learners. Further, the developing body of empirical data which points to the critical importance of high frequency vocabulary (0-3000 frequency range) in L2 listening (Schmitt and Schmitt, 2014) is yet to be extensively investigated. Similarly, gaps in the literature also exist in relation to the relative importance of mid-frequency vocabulary knowledge for L2 listening comprehension success. Finally, there is a paucity of research which has sought to determine the relative importance of aural vocabulary knowledge on the L2 listening comprehension of learners of relatively low and high overall L2 proficiency. This gap in the literature is significant as learner proficiency level is a factor with measurable influence on the efficacy of L2 listening interventions (Matthews et al., 2017). Thus, data which provide insight into the way relative

proficiency levels affect the interaction between vocabulary knowledge and L2 listening comprehension are likely to be valuable in guiding the delivery of individualized L2 vocabulary for L2 listening programs.

Research Questions

This study aims to begin filling existing gaps in the literature by investigating the link between aural vocabulary knowledge, L2 listening comprehension and L2 proficiency among a relatively large cohort of L2 learners. In order to achieve these aims the following research questions will be addressed.

1. What is the contribution of aural vocabulary knowledge of high and mid-frequency words in the prediction of L2 listening comprehension for a single large cohort of learners?
2. What is the contribution of aural vocabulary knowledge of high and mid-frequency words in the prediction of L2 listening comprehension for learners of relatively low and relatively high L2 proficiency level?

The overarching methodological approach is similar to previous related work which has investigated the link between measures of word knowledge and L2 macro-skills (Li & Kirby, 2015; Milton et al., 2010; Qian, 2002; Stæhr, 2009). These studies have gathered measures of both vocabulary knowledge and performance measures of the macro-skill of interest, and applied multiple regression and other inferential statistics to explore the relationships between these constructs.

In an effort to advance the line of existing literature which has explored the relationship between word knowledge in the aural form and L2 listening comprehension, the current study will partially replicate and elaborate on Matthews and Cheng (2015). As with Matthews and Cheng (2015), the current study will use a partial dictation test format to measure knowledge of words in the aural form from various levels of frequency of occurrence as determined through comparison with word lists from the BNC-COCA corpus. Similarly, a standardized listening test (IELTS) will be applied to measure L2 listening comprehension. The current study will build on Matthews and Cheng (2015) by measuring knowledge of words in the aural form a broader range of frequency levels, and examine the relative importance of this knowledge for relatively low and high proficiency learners.

Materials and Methods

Participants

The participants for this study were a convenience sample recruited from a single university in Iran (Azad Islamic University of Ardebil Branch). There was a total of 247 participants, 89 females and 158 males. All participants were aged between 18 and 23, and all reported that their first language was Turkish. Although the researcher did not have access to the participants' previous scores on standardized proficiency tests, according to participants' self-report the group had received on average 10 years of English language study.

Instruments

- **Aural vocabulary knowledge test**

To test aural vocabulary knowledge (AVK) a partial dictation test was used (henceforth AVK test). The AVK test used was that developed for a previous larger study (Cheng & Matthews, 2016). The test targeted the test takers' ability to perceive and produce 63 different target words from three levels of word frequency: 23 words from level one (0e2000 frequency range),

27 words from level two (2001e3000 frequency range), and 13 words from level three (3001e5000 frequency range). Thus, the words from level one and words from level two were representative of high frequency words, and those of level three were representative of mid-frequency words (Schmitt and Schmitt, 2014). The frequency levels of these target words were identified via comparison with word frequency lists generated

with the British National Corpus and the Corpus of Contemporary American English (BNC-COCA) using an online vocabulary profiling tool (Vocabulary Profiler, BNC-COCA1 e25K version) (Cobb, n.d.)

Each item of the AVK test (See Appendix B) required participants to produce a single target word after listening to a stimulus sentence. Each test item's contextual sentence was provided in written form on the test paper, with the target word replaced with a blank. Test items were analyzed to ensure that the words which made up the contextual sentences were of a higher frequency than that of the target word within it. This step was taken to minimize the likelihood that unknown words in the sentence were a factor in limiting the ability of the test taker to correctly produce the target word. The spoken stimulus for the AVK test was recorded into digital form by a native speaker of English. The speaker who read the recorded sentences was asked to speak in a clear, but fluent manner. Stimulus sentences were on average 9.76 words in length, with an average duration of 3.03 s. The average rate of speech for the stimulus sentences was 3.25 words per second, which is indicated as an average rate of delivery of spoken conversation (Tauroza and Allison, 1990, p. 102). The stimulus sentences were edited to have approximately a four second pause between when one sentence finished and the next sentence began. Test takers only heard each sentence once. To provide a clear understanding of the attributes of the speech used in the AVK test stimulus, and to provide the opportunity to use the test in replication studies, the test sound file has been provided (see Appendix A).

The AVK test was validated in a number of steps. First, it was important to ensure that the test predominantly tapped the test taker's ability to perceive the phonological form of the word rather than the ability to strategically infer words from other cues, namely the written contextual sentence of the test items. In order to do this, a two phase validation process was applied.

In the first phase, the contextual sentences were presented to two native speakers, who were asked based solely on their reading of them, to infer the target word. The native speakers were unable to systematically guess the correct target words based on reading the contextual sentences alone. In the second phase, the test was then administered to two different native speakers, but this time in conjunction with the spoken stimulus. All target words were correctly produced by both native speakers. Additionally, after administration to the native speakers and before administration to the main study group, the test was also piloted among a group of L2 learners (N = 40) representative of, but not included within, the main target group. The test was shown to be appropriate for the pilot group and also produced test measures with high levels of reliability.

Another threat to the validity of the AVK test which demanded consideration was the requirement of test takers to produce the target word in the written form. The target construct of the test was a measure of aural vocabulary knowledge, and as such it was important that the test was not a test of spelling (Buck, 2001). Thus a rubric sourced from published peer-reviewed literature was used to standardize scoring protocols (Matthews et al., 2017, pp. 42-43). Target words which were written with minor spelling errors were given full credit, and those test responses which provided sufficiently compelling yet less than emphatic evidence of aural vocabulary knowledge of the target word were ascribed partial credit (0.5 points).

The rationale for using the AVK test as part of this research is as follows. Firstly, target words are presented in the aural modality and thus principally tap the test taker's knowledge of words as they are heard in speech. Secondly, as the test item are heard just once, the test taker is required to access and produce their vocabulary knowledge under time constraints. This aspect of the test is important as the ability to process linguistic information under time constraints is critical to successful listening comprehension (Hulstijn, 2003). Additionally, the test format enabled the measurement of knowledge of a pre-selected list of vocabulary items from a known range of word frequency levels.

- **Listening comprehension test**

To test L2 listening comprehension a version of the listening sub-section of the IELTS was used. This test was the same as that used in Matthews and Cheng (2015). This test required test takers to answer 40 questions including multiple choice, sentence completion, diagram interpretation and table completion while listening to 4 different spoken texts. The spoken texts were monologues and dialogues contextualized within hypothetical

vocational and educational scenarios. The stimulus material was heard once, and test takers had ten minutes after hearing the stimulus material to check their answers. None of the target words from the AVK test were present in the L2 listening comprehension test.

Tests used to estimate relative L2 proficiency level

Two separate tests were used to establish subgroups of those with relatively low and relatively high L2 proficiency: a discrete-point vocabulary test and an integrative reading test. The discrete-point vocabulary test was of a format identical to the Controlled-Production Vocabulary Levels Test (Laufer and Nation, 1999). The test required test takers to read a contextual sentence within which a target word had been replaced with a blank space. The first letters of the target word were presented within the blank space to remove ambiguity in regards to the target word. The test contained 64 items and targeted words from the high and mid-frequency range. Care was taken to ensure that the target words of the discrete-point vocabulary test were not the same as those in the AVK test. The integrative test was a reading test which required test takers to read 8 texts (each approximately 500 word long) and demonstrate their comprehension of those texts by answering 40 multiple choice questions. The reading test tapped test takers' metacognitive and strategic competencies, as well as their linguistic knowledge.

The rationale for using these two tests to estimate relative proficiency level is based on the assumption that L2 proficiency consists of both linguistic and metacognitive competencies (Bachman and Palmer, 1996; Canale and Swain, 1980). Hulstijn (2011) builds on this two part categorization by proposing that measures of overall L2 proficiency can be effectively operationalized through a two-phase assessment approach. The first phase involves testing core proficiency through administration of a discrete-point test, which taps competencies most strongly dependent on linguistic knowledge (Hulstijn, 2011). The second phase involves administering an integrative-skills test which taps peripheral proficiency. This test aims to tap components of proficiency which are less exclusively linked with linguistic knowledge and which strongly depend on metacognitive or strategic competencies (Hulstijn, 2011, p. 239). Thus, as part of the current study, scores from the discrete-point vocabulary test which tapped core proficiency, and scores from the integrative reading test which tapped peripheral proficiency, were assumed to collectively provide sufficient data to categorize learners into relatively low and relatively high proficiency sub-groups.

Procedure

- **Test administration**

The four tests were delivered in two sessions, each of which lasted approximately one hour. The AVK test and the listening comprehension test were administered in the first session. The discrete-point vocabulary test and the reading test were delivered in the second session. Test takers were each given the same instructions and time frames for the completion of the tests. During the testing sessions participants were not allowed to use mobile devices and were instructed to complete the tests in silence.

- **Categorization of low and high proficiency subgroups**

The scores from the discrete-point vocabulary test and the reading test were used to categorize subgroups of relatively low and relatively high proficiency levels. This procedure involved first ranking all participants ($N = 247$) in ascending order according to their vocabulary test scores. Once established the participants in the top half of this ranked list were flagged as "high", and those in the bottom half of this list were flagged as "low". The entire group was then ranked in ascending order according to their reading test scores. Again, participants in the top and bottom half of this ranked list were flagged as "high" and "low" respectively. Participants that were in the top 50 percentile for both the vocabulary test and the reading test were categorized as possessing relatively high overall proficiency. Those participants that were in the bottom 50 percentile for both tests were categorized as relatively low proficiency participants. This resulted in 83 participants being categorized as possessing relatively low L2 proficiency, and 82 participants being categorized as possessing relatively high proficiency.

- **Assumptions**

The hierarchical multiple regression analyses undertaken in this study sought to determine the degree to which levels of AVK (predictor variables) could explain variance observed within L2 listening comprehension scores (outcome variable).

Preliminary analyses were undertaken to ensure the assumptions of collinearity, normality and homoscedasticity were satisfied, and that analyses were undertaken with an adequate sample size (Keith, 2015; Tabachnick and Fidell, 2007).

Results

Descriptive and reliability statistics

Table 1 presents descriptive and reliability statistics of the measures obtained from the three levels of AVK, listening comprehension test, discrete-point vocabulary test and the reading test. As can be seen in Table 1, mean scores for AVK for higher frequency levels were greater than for those of AVK of lower frequency levels. Test scores were of high reliability, with only the reading test having a slightly depressed Cronbach's alpha value. One reason for this lower Cronbach's alpha value may relate to the relative homogeneity of responses to the reading test (SD $\frac{1}{4}$ 11.07). However, as the reading test scores were used in conjunction with the discrete-point vocabulary test (a $\frac{1}{4}$ 0.88), the deflated measure of Cronbach's alpha of the reading test was not viewed as methodologically problematic.

Independent samples t-tests indicated that there was a statistically significant difference between the mean AVK scores attained by the relatively low and high proficiency participants (see Table 2). This was the case for each of the three frequency based categories of AVK: level one AVK, $t(153.81) - \frac{1}{4}$ 9.90, $p \frac{1}{4}$ 0.000, Cohen's $d \frac{1}{4}$ 1.54; level two AVK, $t(163) - \frac{1}{4}$ 11.43, $p \frac{1}{4}$ 0.000, Cohen's $d \frac{1}{4}$ 1.78; and level three AVK $t(163) - \frac{1}{4}$ 10.24, $p \frac{1}{4}$ 0.000, Cohen's $d \frac{1}{4}$ 1.60. Independent samples t-test also indicated a significant difference between the mean listening comprehension test scores of the lower (M $\frac{1}{4}$ 30.53%, SD $\frac{1}{4}$ 13.39) and higher (M $\frac{1}{4}$ 49.23%, SD $\frac{1}{4}$ 13.33) proficiency subgroups, $t(163) - \frac{1}{4}$ 9.00, $p \frac{1}{4}$ 0.000, Cohen's $d \frac{1}{4}$ 1.40, $p \frac{1}{4}$ 0.000.

Table 1: Descriptive statistics of the tests (N=247).

Tests	Items	Min (%)	Max (%)	M (%)	SD	Reliability (Cronbachs α)
AVK Level 1	23	2.17	97.83	55.06	17.49	0.78
AVK Level 2	27	0	88.90	44.33	16.10	0.81
AVK Level 3	13	0	92.31	38.35	20.36	0.75
Listening Comprehension	40	0	86.3	39.47	15.23	0.81
Discrete- point Vocabulary	64	4.7	90.6	55.02	16.57	0.88
Reading	40	32.5	97.5	67.03	11.07	0.62

Table 2 : Descriptive statistics of AVK levels tests for lower (n=83) and higher proficiency sub-groups (n=82).

Test Level	Lower				higher			
	M (%)	SD	Min (%)	Max (%)	M (%)	SD	Min (%)	Max (%)
AVK Level 1	44.32	15.90	2.17	76.09	66.14	12.23	34.78	97.83
AVK Level 2	34.27	12.84	1.70	74.07	56.39	12.00	27.78	88.89
AVK Level 3	26.18	15.98	0	76.92	53.05	17.69	15.38	92.31

Before undertaking multiple regression analysis, correlation coefficients between measures of AVK and L2 listening comprehension were calculated. Each pairing was strong, positive and statistically significant (Cohen, 1988) (See Table 3).

Research question 1: what is the contribution of aural vocabulary knowledge of high and mid-frequency words in the prediction of L2 listening comprehension for a single large cohort of learners?

The first phase of the analysis proper involved using hierarchical multiple regression to determine the degree to which the three measures of AVK were able to explain the variance within L2 listening comprehension scores for the entire cohort. Step one involved insertion of level one AVK and resulted in a model which could predict 40.7% of this variance. Adding level two AVK in step two provided an additional 7.1% of unique explanatory power, resulting in a model which could explain 47.8% of the variance in L2 listening scores. Adding level three AVK in step three provided an additional 5.0% of unique predictive power. Thus in its totality, the model within which all three levels of AVK had been entered, could explain 52.8% of the variance observed in L2 listening comprehension, $F(3, 243) = 90.51$, $p < 0.000$, $R^2 = 0.528$ (see Table 4).

The standardized beta weights associated with the AVK levels at step two and step three of the regression model provide empirical evidence of the relative strength of association between the three levels of AVK and increases within L2 listening comprehension scores. At step two, level one AVK ($b = 0.335$, $p < 0.000$) and level two AVK ($b = 0.403$, $p < 0.000$) were each significantly associated with increases in L2 listening scores. A comparison of the relative magnitude of the beta weights associated with each level of AVK suggests that within the high frequency range (0e3000), level two words, namely those representative of a frequency level ranging between 2001 and 3000, were of higher relative predictive value to the model (see Table 4).

At step three, each of the three levels of AVK contributed significantly to the predictive capacity of the model. Level three AVK contributed the greatest proportion of this predictive capacity ($b = 0.358$, $p < 0.000$), followed by level two AVK ($b = 0.224$, $p < 0.000$), and then level one AVK ($b = 0.217$, $p < 0.000$). Therefore, AVK at both the high frequency (0e3000) and mid-frequency range (3001e5000) contributed significantly to the prediction of L2 listening comprehension scores. This finding is suggestive that for the whole cohort, AVK for words throughout the high frequency range, as well as knowledge of mid-frequency words just beyond the high frequency range were significantly associated with L2 listening comprehension scores (see Table 4).

Table 3: Correlation among AVK levels and 12 listening test (N=247).

Measure	AVK level 1	AVK level 2	AVK level 3	Listening comprehension
AVK Level 1	---	0.75***	0.71***	0.64***
AVK Level 2		---	0.75***	0.66***
AVK Level 3			---	0.68***

*** $P < 0.001$

Table 4: Regression model including AVK levels for total group (N=247).

	R	R ²	R ² Δ	unstandardized		Standardized
				B	SEB	β
Step 1	0.638	0.407***				
Constant				8.865	2.475	
AVK level 1				0.556	0.043	0.638***
Step 2	0.691	0.478***	0.071***			
Constant				6.484	2.364	
AVK level 1				0.292	0.061	0.335***
AVK level 2				0.381	0.066	0.403***

Step 3	0.726	0.528***	0.050***			
Constant				9.425	2.327	
AVK level 1				0.189	0.062	0.217**
AVK level 2				0.212	0.072	0.224**
AVK level 3				0.268	0.053	0.358***

P<0.01, *P<0.001.

Research question 2: what is the contribution of aural vocabulary knowledge of high and mid-frequency words in the prediction of L2 listening comprehension for learners of relatively low and relatively high L2 proficiency levels?

In order to investigate the relative contribution of AVK of high and mid-frequency words in explaining variance within L2 listening comprehension scores for relatively low and relatively high proficiency subgroups, two more hierarchical multiple regression models were built. Each new model was built by including scores from either participants from the relatively low proficiency subgroup (n = 83) or from the relatively high proficiency subgroup (n = 82)

- **Results for the low proficiency subgroup**

From the model summarized in Table 5, it can be observed that only level one AVK provided a unique contribution to the prediction of L2 listening comprehension for the low proficiency group. Thus, step one which included just level one AVK could explain 27.6% of the variance in L2 listening comprehension scores for those with relatively low L2 proficiency ($F(1, 81) = 30.89$ (p = 0.000, $R^2 = 0.276$). Adding AVK levels 2 and 3 did not add any additional statistically significant predictive power to the regression model. Thus, for the low proficiency group only the AVK of words representative of the most frequent 2000 word families were predictive of L2 listening comprehension.

- **Results for the high proficiency subgroup**

The model summarized in Table 6 shows that entering only level one AVK (step one) could explain 16.7% of the variance in L2 listening comprehension scores ($F(1, 80) = 16.04$, p = 0.000, $R^2 = 0.167$). At step two, which included level one and level two AVK, 30.4% of the variance could be explained ($F(2, 79) = 17.26$, p = 0.000, $R^2 = 0.304$); however, only level two AVK contributed statistically significantly to the predictive capacity of this step of the model. At step three, which included all three levels of AVK, the model was able to predict 49.0% of the variance in the L2 listening comprehension scores of the relatively high proficiency subgroup ($F(3, 78) = 25.03$, p = 0.000, $R^2 = 0.490$).

Appraisal of the beta weights attributed to each of the AVK levels in the final step of the model shows that only level two AVK (b = 0.259, p = 0.019) and level three AVK (b = 0.560, p = 0.000) contributed significantly to the prediction of L2 listening comprehension scores. A comparison of the relative magnitude of the beta weights associated with step three, suggests that level three AVK, namely AVK of words representative of the mid-frequency level, are of higher relative predictive value to the model than those of level two AVK (see Table 6) A synthesis of the key information derived from the analyses undertaken in the present study is provided in Table 7.

Table 5: Regression model including AVK levels for the low proficiency group (n=83).

	R	R ²	R ² Δ	unstandardized		Standardized
				B	SEB	β
Step 1	0.525	0.276***				
Constant				10.923	3.745	
AVK level1				0.442	0.080	0.525***

Step 2	0.532	0.283	0.007			
Constant				9.724	3.976	
AVKlevel1				0.375	0.109	0.445**
AVKlevel2				0.123	0.135	0.118
Step 3	0.563	0.317	0.033			
Constant				10.683	3.938	
AVKlevel1				0.290	0.116	0.344*
AVKlevel 2				0.051	0.138	0.049
AVKlevel 3				0.201	0.103	0.240

*P<0.05, **P<0.01, ***P<0.001.

Table 6: Regression model including AVK levels for the high proficiency group (n=82).

	R	R ²	R ² Δ	unstandardized		Standardized
				B	SEB	β
Step 1	0.409	0.167***				
Constant				19.782	7.478	
AVK level1				0.445	0.111	0.409***
Step 2	0.551	0.304***	0.137***			
Constant				10.899	7.237	
AVKlevel1				0.140	0.128	0.128
AVKlevel2				0.516	0.131	0.465***
Step 3	0.700	0.490***	0.186***			
Constant				15.271	6.286	
AVKlevel1				-0.070	0.177	-0.065
AVKlevel 2				0.288	0.120	0.259*
AVKlevel 3				0.422	0.079	0.560***

*P<0.05, ***P<0.001.

Discussion

The current study was primarily focused on investigating the relative contribution of AVK of high and mid-frequency words on L2 listening comprehension. Research question one addressed this in relation to a single cohort of participants.

Results indicated that for the cohort involved in this study, AVK of words from across high and mid-frequency ranges was collectively able to predict over half of the variance observed within L2 listening scores. This finding, coupled with the strong and positive correlation observed between AVK and L2 listening comprehension, suggests that the AVK of words from the high and mid-frequency range was strongly associated with L2 listening comprehension success.

The strength of association between the measures of aural vocabulary knowledge and L2 listening comprehension scores observed ($r = 0.64$ to 0.68 , $p < 0.001$) were very similar to those put forward in previous related research. For example, Milto et al. (2010) reported a correlation of 0.67 between measures of aural

vocabulary knowledge and L2 listening comprehension test scores. Matthews and Cheng (2015) also reported a similar strength of correlation between measures of AVK and L2 listening comprehension scores ($r = 0.67$ to 0.73 , $p < 0.001$). Further, the capacity of the AVK test scores to predict variance observed in L2 listening comprehension reported in Matthews and Cheng (2015) ($R^2 = 0.54$) was remarkably similar to that reported in the current study ($R^2 = 0.53$). This was the case even though there were different target words drawn from different frequency levels in each of the AVK tests used for the two studies. The similarity of these results is emerging evidence that, at least for learners similar to those that were part of this study, AVK tests which measure knowledge of high frequency vocabulary items are likely to predict a significant proportion of variance in L2 listening test scores. The participants in both the current study and Matthews and Cheng (2015) had comparably low overall L2 listening comprehension scores ($M = 39.5\%$, and 36.7% respectively), but scores for AVK tests were lower for the participants of the current study. This is likely to relate to the increased speech rate used in the AVK stimulus sentences in the current study (approximately 3.25 words per second) as compared to that used in Matthews and Cheng (2015) (approximately 2.1 words per second).

Research question two sought to determine the degree to which high and mid-frequency AVK was differentially predictive of L2 listening comprehension scores for participants with relatively low and high overall proficiency. For the relatively low proficiency subgroup, AVK was only able to predict a little over a quarter of the variance observed in L2 listening scores. Only AVK of the very highest frequency range measured as part of this study (0-2000 frequency range) contributed (27.6%) significantly to the prediction of the low proficiency subgroup's L2 listening comprehension. This result suggests that the low proficiency subgroup's AVK for the upper limits of the high frequency range (2001-3000) and mid-frequency range measured (3001-5000) was not sufficient to impart a measurable influence on their L2 listening comprehension success.

For the relatively high proficiency subgroup, the final step of the regression model which included all three levels of AVK could explain close to half of the variance observed in L2 listening comprehension scores. Both high and mid-frequency AVK contributed to the prediction of L2 listening comprehension for this subgroup, but this was only the case for the words representative of the 2001-5000 frequency range. Apparently for this relatively high proficiency subgroup, the magnitude of AVK for words at the 2001-5000 frequency level subsumed any unique predictive capacity that AVK for the 0-2000 frequency range could provide. This finding suggests a strong association between AVK of words beyond the most frequent 2000 word families, and the L2 listening comprehension performance of those with relatively high overall L2 proficiency. Additionally for the high proficiency subgroup, the AVK of mid-frequency words was more than twice as strongly associated with changes in L2 listening comprehension than was AVK of high frequency words (see Table 6). This finding is suggestive of the benefit of learners possessing robust AVK of words beyond the high frequency range.

Pedagogical implications

This study presents empirical evidence which can inform pedagogical approaches to the development of L2 word knowledge for the purpose of supporting L2 listening comprehension. Firstly, the importance of AVK of the most frequent word families for the entire cohort encourages a prescription for learners to acquire a robust command of these words as

early as possible in the language learning process. Rather than being considered a body of vocabulary knowledge which facilitates adequate L2 listening success, AVK of the most frequent 2000 word families is considered here as a prerequisite foundation of knowledge from which, over time, superior levels of L2 listening comprehension may be built. To build AVK of high frequency words, computer applications which provide learners with increased engagement with aural texts predominantly composed of the most frequent 2000 words hold strong potential. Previously described computer applications (e.g., Hulstijn, 2003) can

provide learners with self-directed opportunities to listen to sections of connected speech, reconstruct the speech in stages and receive automated feedback on their accuracy of word recognition. Facilitating computer mediated, intensive listening has been associated with measurable improvements in AVK among L2 learners (Matthews et al., 2015), and the use of such applications may provide a practical way to help large cohorts of learners establish a solid foundation of AVK. Further, for learners with suboptimal AVK for words in the 0e2000 range, an ongoing pedagogical emphasis on the development of listening strategies is also warranted. As low proficiency learners are unlikely to possess sufficient AVK to deal with the processing demands of L2 listening, intensive instruction on the effective application of listening strategies also needs to be delivered while the AVK of words from the 0e2000 range is being developed. Without possessing the specific forms of word knowledge necessary to access the most frequent lexical items in spoken language, strategic competence may be the only cognitive resource low proficiency learners can usefully draw on in an attempt to comprehend L2 speech.

Table 7: Summary of results

	Total cohort (ß)	Low proficiency subgroup (ß)	High proficiency subgroup (ß)
Level 1 AVK	0.217**	0.344*	-0.065
Level 2 AVK	0.224**	0.049	0.259*
Level 3 AVK	0.358***	0.240	0.560***
Total variance explained	R ² =0.528***	R ² =0.276***	R ² =0.490***
By all three levels of AVK			

*P<0.05, **P<0.01, ***P<0.001.

As the AVK of words representative of the 2001e3000 frequency range was significantly predictive of greater L2 listening comprehension test scores, it is suggested that L2 vocabulary for listening programs do not stop once the learner possesses an adequate command of AVK for words from the 0e2000 range. However, it is suggested words in the 2001e3000 range only be explicitly addressed once learners have an adequate AVK of words from the 0e2000 range. A solid foundation of AVK of individual word forms from the 0e2000 frequency range, which provides a lexical coverage of approximately 92%, is likely to greatly assist learners to successfully infer the aural form and meaning of unknown words from the 2001e3000 frequency range, which represents just an additional 2% lexical coverage of spoken texts (Adolphs and Schmitt, 2003, p. 431). Adequate knowledge of AVK of the most frequent 3000 word families is a valuable basic goal for most listening programs. Knowledge of the 2001e3000 word frequency range was predictive of L2 listening comprehension for both the entire group and the relatively high proficiency subgroup (see Table 7). Accordingly, pedagogical approaches which succeed in pushing learners' AVK right up to and beyond the high frequency level are assumed to be of strong value to learners of a range of proficiency levels. Listening texts, both adapted and authentic, which possess lexical content within the 3000 frequency range, are likely to be of high value in enhancing listeners' knowledge of words from the 2001e3000 range.

In terms of the value of developing AVK beyond the high frequency range, the capacity of mid-frequency words to predict listening comprehension for the entire cohort and the relatively high proficiency subgroup must be acknowledged. Although AVK of high frequency words may provide listeners with knowledge of the vast majority of running words they are likely to encounter while listening to authentic texts, expansion of a learner's AVK into the mid-frequency range is likely to have a positive impact on the speed by which vocabulary knowledge can be accessed for both high and mid-frequency words (Laufer and Nation, 2001; Schmitt and Schmitt, 2014). It is possible that knowledge of mid-frequency AVK is associated with improved L2 listening comprehension due to the direct positive effect of having knowledge of the individual mid-frequency words encountered while listening. It may also be the case that depth of AVK of words in this

frequency range is an indication of a more generalized construct of proficiency which is a by-product of greater levels of exposure to, and meaningful engagement with, spoken L2 input. In either case, it remains that AVK of mid-frequency words is likely to be a robust indicator of both the fluency with which words from both the high and mid-frequency range can be processed, as well as L2 listening comprehension capabilities more generally. Learners possessing adequate levels of AVK of high frequency words (0e3000) are likely to be capable of negotiating selected authentic texts relatively well (Webb and Rodgers, 2009b), and thus should be encouraged to listen extensively to target language podcasts, television shows, movies, radio and so forth. In this regard, it is the view of the author that it is especially important that these texts be chosen by the learners themselves such that learners are engaged with spoken texts which they are intrinsically motivated to understand.

Lastly, the implications of the modality specific nature of word knowledge needs to be emphasized. Programs aimed at supporting L2 listening through vocabulary development should place a pedagogical focus on increasing learners' engagement with words in the aural modality. Further, regular diagnostic and formative testing of the AVK of high and mid-frequency words, with tests similar to those implemented as part of this study, is likely to provide important information to language teachers about what level of AVK learners possess. Such information will enable language teachers to make informed pedagogical decisions about which level of AVK should be focused on in an effort to support L2 listening. Additionally, systematic use of such tests may give rise to positive wash back in the form of learners actively seeking to monitor and refine their AVK throughout the duration of their L2 study.

Limitations

There are a number of limitations to the current study, the identification of which may assist future research efforts to build on this work. Firstly, due to the feasibility of administering multiple tests to a relatively large cohort, only one measure of listening comprehension was used. Using a variety of standardized listening tests would have provided more robust evidence of the strength of association between AVK and a broader range of L2 listening comprehension constructs. Secondly, the study was undertaken among a relatively homogenous participant group in regards to background demographics. Exploring the relationships between AVK and L2 listening comprehension among a broader cross-section of language learners from a range of proficiency levels, linguistic backgrounds, ages and educational levels would provide a more generalizable picture of the relationship between AVK and L2 listening comprehension.

In addition, due to reasons of feasibility there was a relatively limited range of mid-frequency vocabulary items tested as part of this study. Expanding the range and number of words tested from beyond the 3000 frequency range would provide additional information about the relative predictive capacity of words from the mid-frequency. Another limitation in relation to the testing procedures of the current study was that a broadly comparable measure of proficiency (e.g., IELTS test scores for all sub-skills) was not available for the group due to issues of feasibility and privacy of information. Establishing estimates of L2 proficiency through use of such measures is recommended for future studies.

Finally, the recommendations put forward here are drawn from inferential statistics which have examined the interrelationships between AVK, L2 listening and relative learner proficiency level. Although based on empirical evidence, the effectiveness of these recommendations in real education contexts is yet to be validated. Of particular interest in this regard will be the implementation of longitudinal studies which follow cohorts of learners of different proficiency levels, and examine the impact that improvements of AVK have on L2 listening comprehension success.

Conclusions

The current paper has presented evidence of the importance of aural vocabulary knowledge in L2 listeners' capacity to comprehend L2 spoken language. It seems likely that knowledge of words in the aural form from high to mid-frequency, and beyond, will provide learners with an important foundation of linguistic

knowledge supportive of L2 listening capabilities. Based on this evidence, it is reasonable to suggest that the development of L2 aural vocabulary knowledge represents a useful piece of the puzzle in efforts to systematically enhance learners' L2 listening capacities.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.system.2017.10.005>.

Appendix B

Aural vocabulary test

Instructions: Listen to the sentences and fill in the blanks according to what you hear. Each sentence will be heard only once.

Each sentence has only one word missing. Look at the example in the box below

If you hear: "*He lives in Europe*" – You write "*Europe*" in the space provided.

Example: He lives in.....

1. Her two favourite subjects at university were_____ and computer studies
2. The worker had a lot of _____ in the field.
3. She has worked as a _____ for most of her life.
4. I called his _____ this morning but he wasn't there.
5. The police officer made sure the _____ was secure.
6. The teacher _____ for the children every day.
7. This _____ has very good food.
8. Two things which I love the most are _____ and painting.
9. The poor _____ made it difficult to enjoy travelling.
10. A major_____ of the machine is its cost.
11. The _____ between the two groups went for over an hour.
12. The woman wanted to _____ a few issues to the student.
13. I will try to _____ the office tomorrow morning.
14. The man found it difficult to _____ in the hot weather.
15. The student was_____ she would be able to complete the work on time.
16. The young man was more _____ than the older man.
17. The government tried to increase _____ within the country.
18. The student had seen a similar _____ in his book last year.
19. He explained the _____ to his friend.
20. The student couldn't decide which _____ would be best for him.
21. The best_____ was near the post office.
22. Living in a small town can sometimes be a _____ for students.
23. Which _____ of the play did you like the most?
24. The door to the _____ was very difficult to open.
25. He had disliked_____ his entire life.
26. Some children will choose to eat a _____ rather than a piece of fruit.
27. The woman wanted to _____ the book as soon as possible.
28. The man wanted to _____ some of his money.
29. It was important to _____ the information before the end of the day.
30. They wanted to _____ the worker the following day.
31. The _____ team was the best in the state.
32. The boy surprised the man by speaking in a very _____ way.

33. she planned to contact the _____ the following day
34. the children rode their bikes along the _____ path
35. The cleaner needed to buy a new _____ before the end of the week.
36. It was necessary for the _____ to take place as soon as possible.
37. The woman wasn't sure what the _____ meant.
38. He was surprised to see a large _____ on his kitchen table.
39. The child's _____ was very important to his grandparents.
40. The police found the _____ in the park yesterday.
41. The boy's _____ was to assist his friends.
42. The family was hoping to reach the _____ before evening.
43. The man wanted to _____ the appearance of the room.
44. The politician wanted to get a _____ as soon as possible.
45. It is often important to _____ your plans to friends.
46. The man was hoping to _____ the book within a month.
47. The _____ dancing went on for hours.
48. The children had a _____ experience at the park.
49. She didn't have any information about the _____ of the meeting.
50. She had been a _____ teacher for about three years.
51. The _____ had involved many years of planning.
52. The politician's _____ was popular with his supporters.
53. The woman bought _____ and bread from the supermarket.
54. The _____ of the project would involve higher costs.
55. The man had lived in _____ with his family for many years.
56. The team's _____ was totally unexpected.
57. His _____ was unhappy with the meal.
58. She could _____ the colours very clearly.
59. The student wasn't able to _____ the poem very clearly.
60. Some things are very difficult to. _____ .
61. The boy found it hard to _____ the hot weather.
62. The _____ number of problems created stress in the office.
63. The man was quite _____ towards the young boy.

Target words (not provided to test takers during the test)

- .Finance (level 1)
- .Experience (level 1)
- .Researcher (level 1)
- .Office (level 1)
- .Evidence (level 1)
- .Cares (level 1)
- .Region (level 1)
- .Nature (level 1)
- .Transportation (level 2)
- .Limitation (level 1)
- .Debate (level 2)
- .Explain (level 1)
- .Contact (level 1)
- .Operate (level 1)

.Positive (level 1)
.Successful (level 1)
.Stability (level 1)
.Diagram (level 3)
.Method (level 2)
.Career (level 1)
.Accommodation (level 2)
.Disadvantage (level 3)
.Aspect (level 2)
.Garage (level 1)
.Conflict (level 2)
.Biscuit (level 1)
.Publish (level 2)
.Withdraw (level 2)
.Analyze (level 2)
.Interview (level 1)
.Legal (level 1)
.Formal (level 2)
.Author (level 2)
.Cement (level 3)
.Uniform (level 2)
.Surgery (level 2)
.Symbol (level 2)
.Parcel (level 3)
.Welfare (level 2)
.Weapon (level 1)
.Motive (level 2)
.Border (level 2)
.Transform (level 2)
.Divorce (level 1)
.Reveal (level 2)
.Translate (level 2)
.Classical (level 1)
.Terrific (level 3)
.Duration (level 3)
.Chemistry (level 3)
.Construction (level 2)
.Objective (level 2)
.Vitamins (level 3)
.Expansion (level 2)
.Harmony (level 3)
.Elimination (level 2)
.Acquaintance (level 3)
.Perceive (level 2)
.Recite (level 3)
.Attain (level 3)
.Endure (level 2)

.Infinite (level 3)
 .Hostile (level 2).

References

1. Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied linguistics*, 24(4), 425-438.
2. Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice* Oxford: Oxford University Press.
3. Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International journal of listening*, 14(1), 14-31.
4. Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, 36(1), 22-34.
5. Buck, G. (2001). *Assessing listening*. Cambridge University Press.
6. Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, 1(1), 1-47.
7. Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, 35(1), 3-25. <https://doi.org/10.1177/0265532216676851>.
8. Cobb, T. (n.d.). Web Vocabprofile. Retrieved from <http://www.lexutor.ca/>
9. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
10. Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, 24(2), 143-188.
11. Field, J. (2008a). *Listening in the language classroom*. Cambridge, UK: Cambridge University Press.
12. Field, J. (2008b). Revising segmentation hypotheses in first and second language listening. *System*, 36(1), 35-51.
13. Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55-75.
14. Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, 18(3), 193-223.
15. Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning*, 16(5), 413-425.
16. Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, 8(3), 229-249.
17. Keith, T. Z. (2015). *Multiple regression and beyond* (2nd ed.). New York, NY: Routledge.
18. Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language testing*, 16(1), 33-51.
19. Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related?. *Eurosla Yearbook*, 1(1), 7-28.
20. Li, M., & Kirby, J. R. (2014). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics*, 36(5), 611-634.
21. Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, 52, 1-13.
22. Matthews, J., Cheng, J., & O'Toole, J. M. (2015). Computer-mediated input, output and feedback in the development of L2 word recognition from speech. *ReCALL*, 27(3), 321-339.
23. Matthews, J., O'Toole, J. M., & Chen, S. (2017). The impact of word recognition from speech (WRS) proficiency level on interaction, task success and word learning: design implications for CALL to develop L2 WRS. *Computer Assisted Language Learning*, 30(1-2), 22-43.
24. McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19(6), 741-760.

25. Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners?. *Canadian Modern Language Review*, 63(1), 127-147.
26. Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacon-Beltr an, C. Abello-Contesse, M .Torreblanca-Lopez, & M. L opez-Jim enez (Eds.), Further insights into non-native vocabulary teaching and learning (pp. 83e97). Bristol, UK: Multilingual Matters.
27. Nation, I. (2006). How large a vocabulary is needed for reading and listening?. *Canadian modern language review*, 63(1), 59-82.
28. Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language learning*, 52(3), 513-536.
29. Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484-503.
30. Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139-152.
31. Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in second language acquisition*, 31(4), 577-607.
32. Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education.
33. Tauroza, S., & Allison, D. (1990). Speech rates in british english. *Applied linguistics*, 11(1), 90-105.
34. Tsui, A. B., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied linguistics*, 19(4), 432-451.
35. van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension?. *Applied Linguistics*, 34(4), 457-479.
36. Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, 65(2), 390-416.
37. Webb, S., & Rodgers, M. P. (2009). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407-427.
38. Webb, S., & Rodgers, M. P. (2009). Vocabulary demands of television programs. *Language Learning*, 59(2), 335-366.