

RESEARCH

Open Access



Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension

Hung Tan Ha 

Correspondence: hatanhung1991@gmail.com

School of Foreign Languages,
University of Economics Ho Chi
Minh City, 279 Nguyen Tri Phuong
Street, District 10, Ho Chi Minh City,
Vietnam

Abstract

The article presents an empirical study that investigates the single- and cross-modality relationships between different dimensions of receptive vocabulary knowledge and language skills, as well as the importance of academic vocabulary knowledge in academic listening and reading comprehension. An Updated Vocabulary Levels Test (UFLT), a Vietnamese version of the Listening Vocabulary Levels Test (LFLT), an International English Language Testing System (IELTS) listening test and an academic IELTS reading test were administered to 234 tertiary level Vietnamese learners of English as a foreign language (EFL). Research findings showed that (1) orthographic and aural vocabulary knowledge were strongly correlated ($r = .88$) and of equal significance to L2 listening and reading comprehension, (2) receptive vocabulary knowledge was a very powerful and reliable predictor of learners' receptive language proficiency, (3) knowledge of academic vocabulary strongly correlated with academic listening ($r = .65$) and reading ($r = .60$) comprehension and the mastery of the Academic Word List (AWL) could suggest a band score 6.0 in both the IELTS listening and academic reading tests.

Keywords: Aural vocabulary knowledge, Academic listening, Academic reading, Academic vocabulary, IELTS, Listening vocabulary levels test, Orthographic vocabulary knowledge, Phonological vocabulary knowledge, Updated vocabulary levels test

Background of the study

Vocabulary is a fundamental aspect of language (Nation, 2006, 2013; Webb, 2020). The strong link between a learner's vocabulary knowledge and his or her reading and listening comprehension, quality of writing, and speaking fluency has been discussed at length (Bian, Cai, & Cai, 2021; Cheng & Matthews, 2018; Hilton, 2008; Matthews & Cheng, 2015; Qian & Lin, 2020; Schmitt, Jiang, & Grabe, 2011; Stæhr, 2008, 2009). Moreover, the knowledge of vocabulary has been found to be of greater importance than grammatical structures and the knowledge of subject matters (Guo & Roehrig, 2011; Zhang, 2012; Zhang & Koda, 2013). In fact, grammatical errors were proven to



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

have higher tendency to cause misunderstanding and disrupt communication than lexical errors (Lewis, 2002). It has even been argued that the application of grammar has to be based on a rich lexical resource, and the development of language lexicon is an essential prerequisite for the acquisition of grammar rules (Barcroft, 2007).

Most discussions in the field of vocabulary study concluded that in order to gain adequate comprehension in reading and listening, the learners should be familiar with at least 95% and preferably 98% of the running words in the text (Laufer, 2013; Nation, 2006; Schmitt et al., 2011). Regarding the relationship between lexical coverage and comprehension, two things have been repeatedly suggested, the first one is that 98% is the desirable threshold for adequate comprehension while 95% is only the acceptable threshold for minimal comprehension (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010). The second suggestion is that the 98% coverage can only be considered to be the requirement for optimal reading comprehension and that the same threshold should not be applied to listening comprehension (Laufer & Ravenhorst-Kalovski, 2010; van Zeeland & Schmitt, 2013).

“There is a presumption here that the foreign language mental lexicon has two halves; an orthographic half, where written representations of words are stored, and a phonological half, where the aural representations are stored” (Milton et al., 2010, p. 84). Nation’s (2001, p. 27) table of aspects of word knowledge gives a clear distinction between what a word looks like and what a word sounds like. Milton and Hopkins’s (2006) study showed that phonological and orthographic vocabulary knowledge are strongly correlated at .68 and suggested that orthographic vocabulary knowledge of a learner was broader than his or her aural vocabulary knowledge. Cheng and Matthews (2018) found those claims to be unsurprising due to the nature of written and spoken words. They believed that the orthographic form of a word was temporally permanent and thus could be revisited repeatedly by readers. On the other hand, spoken words were temporary and only available for processing for a very short period of time. As a result, learners always had to perceive and process spoken words in a highly time-constrained manner (Cheng & Matthews, 2018). Such differences are supposed to be the reasons why listening-based tests are generally more challenging than their reading counterparts, given that both input texts share similar lexical demands.

Most of the research that claimed to have investigated the relationship between vocabulary knowledge and language subskills did, in actual practice, only reflect the participants’ orthographic knowledge of vocabulary (Lange & Matthews, 2020). However, it is undeniable that the findings in those studies did record a remarkably strong link between orthographic knowledge of vocabulary and listening and reading comprehension (Stæhr, 2008, 2009). For example, orthographic knowledge of vocabulary was found to be correlated with L2 reading and listening at .77 (Qian, 2002) and .70 (Stæhr, 2009) respectively. Stæhr’s (2008) study also found positive, strong correlations between receptive orthographic vocabulary knowledge and listening ($r = .69$) and reading ($r = .83$).

In a recent study, Lange and Matthews (2020) used the LVL as a measure of aural vocabulary knowledge and explored the relationship between the capacity to process aural input at lexical level of 130 EFL students and their performance on two listening tests which included the TOEIC listening subtest and the Eiken Pre-2 listening test.

They reported surprisingly weak correlations between the participants' total score on the LVLT and the two listening tests ($r = .15$ and $r = .12$). However, when it came to the correlations between each 1000-word level and the two listening tests, the students' performance on the first and second 1000-word levels was found to have relatively stronger correlations with the TOEIC listening test ($r = .48$ and $r = .47$ respectively) and the Eiken Pre-2 listening test ($r = .42$ and $r = .44$ respectively) (Lange & Matthews, 2020). Additionally, Lange and Matthews (2020) reported a sharp drop in correlations between participant vocabulary knowledge and listening comprehension from the 3000 word level onwards.

Table 1 summarizes the reported correlations between the scores on different vocabulary and listening comprehension tests, as well as the levels of English proficiency of the participated learners in a number of studies. A close examination of these studies suggested that the correlations may vary greatly depending on the English proficiency levels of the participants and that the correlations between vocabulary knowledge and comprehension tended to be smaller when the participants took part in the study had low levels of English proficiency. As Lange and Matthews (2020) stated, "one possible limitation is that the relatively low proficiency level of the participants indicates they may have been unfamiliar with much of the low-frequency vocabulary from the 1,000 word-level and above. Possible floor effects for sections of the Listening Vocabulary Levels Test containing low-frequency vocabulary may have diluted the value of the aural vocabulary knowledge data" (p. 740).

Only a few studies have investigated the cross-modality relationship between receptive phonological, orthographic vocabulary knowledge and listening and reading comprehension among a single cohort of L2 participants. Milton et al. (2010) utilized the X_Lex (Meara & Milton, 2003) and AuralLex (Milton & Hopkins, 2005), which used the yes/no format to investigate the relationship between the receptive orthographic

Table 1 A summary of the correlation coefficients reported in different studies using different vocabulary tests and listening tests

Participants' level of English proficiency	Vocabulary test used	Listening test used	Correlations		Study
			Total scores	The 1000- and 2000-word levels	
B2	VLT (Schmitt et al., 2001)	Cambridge certificate of proficiency in English (CPE)	.70	-	Stæhr (2009)
B1	VocabLab Test (Peters et al., 2015, 2019)	Preliminary English Test for schools (PET)	.63	-	Noreillie et al., (2018)
A2	LVLT (McLean et al., 2015)	TOEIC listening test	.15	.48 and .47 respectively	Lange and Matthews (2020)
		Eiken Pre-2 listening test	.12	.42 and .44 respectively	
CET Band 4	VST (Nation & Beglar, 2007)	IELTS listening test	.70	-	Feng (2016)
Mixed (87% low proficiency, 6.5% mid proficiency, 6.5% high proficiency)	LVLT (McLean et al., 2015)	Part 01 and 02 of the TOEIC listening test	.54	-	McLean et al. (2015)

and phonological vocabulary sizes of 30 ESL learners and their performance on the IELTS listening, reading, writing, and speaking subtests. Their results showed that learners' orthographic vocabulary knowledge significantly correlated with their scores on the reading and writing subtests (.70 and .76 respectively) but produced weaker correlations with the listening ($r = .48$) and speaking ($r = .35$) test scores. On the other hand, the aural vocabulary knowledge was reported to have low correlations with the reading ($r = .22$) and writing ($r = .44$) tests but strong correlations with the listening and speaking subtests at .67 and .71 correspondingly (Milton et al., 2010). Although their findings did make a significant contribution to the understanding of the relationship between vocabulary knowledge and language subskills, the study had two underlying limitations. First, the fact that only a very small number of participants took part in the study ($N = 30$) could have greatly weakened the arguments and even made the conclusions inconclusive. Second, the yes/no format of the test was believed to have only suggested the examinees' knowledge of the target words' existence, and it was uncertain to what extent the test takers know the meaning of the target words (McLean et al., 2015).

In a more recent attempt to address the cross-modality issue, Cheng and Matthews's (2018) study of 250 EFL students reported strong and moderate correlations between productive phonological vocabulary knowledge and listening ($r = .71$) and reading ($r = .46$) comprehension, and similarly strong correlations between productive orthographic word knowledge and listening ($r = .55$) and reading ($r = .57$) comprehension. Although Cheng and Matthews (2018) measured multiple aspects of vocabulary knowledge and L2 reading and listening among a single, large cohort of L2 participants, what the study really reflected was the relationship between productive phonological and orthographic knowledge of vocabulary and receptive L2 subskills. As a result, none of the reviewed studies has really investigated the relationship between receptive aural and orthographic vocabulary knowledge and listening and reading comprehension among a single large group of L2 learners.

Another under-researched area is the relationship between academic vocabulary and academic listening and reading (Matthews & Cheng, 2015). Research has proven the importance of academic vocabulary in the comprehension of academic texts. The AWL was found to account for approximately 10% of the running words in academic texts (Coxhead, 2000) and 4.41% in academic lectures and seminars (Dang & Webb, 2014). Dang and Webb's (2014) corpus-driven study that investigated the lexical profile of the British Academic Spoken English (BASE) corpus highlighted that the knowledge of the most frequent 3000 word families in the BNC word list plus the AWL could provide a 95% coverage of academic spoken English. In contrast, without the knowledge of the AWL, learners would need to be familiar with the most frequent 4000 word families in the BNC word list to obtain a 95% coverage (Dang & Webb, 2014). Their findings also suggested that the impact of the AWL on the coverage of academic spoken English decreased as vocabulary level went up and that such impact could only be deemed significant for the first two or three 1000-word levels.

While the previous studies gave a comprehensive view on the relationship between vocabulary levels, lexical coverage and listening and reading comprehension (Cheng & Matthews, 2018; Milton et al., 2010; Stæhr, 2009), the importance of the academic vocabulary in tests of English for academic purposes seemed to be left untouched. The

researchers' justification for excluding the AWL from data analysis was that the words in the AWL came from various word levels and thus should not be viewed as a distinct word level and an addition to the learners' vocabulary size (Laufer & Ravenhorst-Kalovski, 2010; Stæhr, 2009). While the argument is scientifically evident and acceptable to some extents, the impact of the AWL on the text coverage of first and second 1000-word levels is indeed significant and should not be overlooked. The present study involves the use of the LVLT, which also contains a 30-item level for the AWL. The researcher utilized the academic word level in the LVLT to examine the relationship between academic vocabulary knowledge and academic listening and reading.

Research questions

The present study's objective is to shed light upon the relationship between different dimensions of receptive vocabulary knowledge and receptive language skills. In particular, the research seeks to answer the following questions:

1. What is the relationship between aural and orthographic knowledge of vocabulary and academic listening and reading comprehension?
2. How much vocabulary is needed for adequate academic listening and reading comprehension?
3. What is the relationship between academic vocabulary and academic listening and reading comprehension?

Methodology

Participants

The participants were 234 Vietnamese EFL second year non-English majors at a highly-ranked public university in Ho Chi Minh City, Vietnam. Convenience sampling was applied. The participants were the students in 8 Business English classes which the researcher was the lecturer-in-charge. The participants' ages ranged from 20 to 23. None of the participants had lived in a country where English is the official language. At the time of data collection, all of the participants had studied English for at least 9 years and had completed three out of four compulsory business English courses at their university. The participants' English education background and their IELTS scores suggested an average English language proficiency level of B1.

Instruments

The participants were given four paper-and-pencil tests: A Vietnamese version of the Listening Vocabulary Levels Test (Ha, *in press*; McLean et al., 2015), an Updated Vocabulary Levels Test (Webb et al., 2017), an IELTS listening comprehension test, and an IELTS academic reading comprehension test.

The Vietnamese listening vocabulary levels test

In an attempt to develop an instrument for the measurement of phonological knowledge of vocabulary, McLean et al. (2015) created the Listening Vocabulary Levels Test (LVLT). The test assesses learners' knowledge of the 1000, 2000, 3000, 4000, and 5000 word frequency levels in Nation's (2012, cited in McLean et al., 2015) BNC/COCA

word list, the test also contains a 30-item academic vocabulary level from the AWL (Coxhead, 2000). The test is believed to outperform other tests of aural vocabulary knowledge which are Fountain and Nation's (2000) graded dictation test and the Aural-Lex created by Milton and Hopkins (2005) thanks to four reasons. First, each target word is followed by a context defining sentence that provides extra information on the word's part of speech and its contextualized meaning, which support examinees in accessing the meaning of the target word (Henning, 1991, cited in McLean et al., 2015). Second, test takers can listen to the target word and the context defining sentence only once, which is representative of the level of difficulty demanded in most listening situations (Buck, 2001, pp. 170-171, cited in McLean et al., 2015). Third, the LVLT inherits the 4-option multiple-choice format of the Vocabulary Size Test (Nation & Beglar, 2007), which investigates a deeper depth of vocabulary knowledge, as distinct from the shallow knowledge of the word's existence (McLean et al., 2015). Finally, the four options are given in the test taker's first language, which is believed to eliminate all the measurement errors caused by other constructs such as L2 reading ability and comprehension, which would lead to higher validity and reliability compared to monolingual vocabulary tests (McLean et al., 2015). The options in the LVLT were translated to develop a Vietnamese bilingual version of the test (Ha, *in press*). Both the Japanese and Vietnamese versions of the LVLT satisfied three major validating requirements. First, the test items displayed very good fit to the Rasch model and presented sufficient spread of difficulty. Second, the items showed very strong unidimensionality and were free of local dependency. Third, both versions of the LVLT presented a high degree of generalizability and was proven to strongly correlate with the TOEIC and IELTS listening tests (Ha, *in press*; McLean et al., 2015). The present study employed the Vietnamese LVLT to measure learners' receptive phonological knowledge of vocabulary.

The updated vocabulary levels test

The Updated Vocabulary Levels Test (UVLT) created by Webb et al. (2017) was used as an instrument for the measurement of learners' receptive orthographic vocabulary knowledge. The UVLT inherits the VLT's matching format with 10 3-item clusters per vocabulary level and the proportion of 15 nouns, 9 verbs, and 6 adjectives per word level (Webb et al., 2017). Three improvements were made to cover the limitations of Schmitt et al.'s (2001) VLT. Firstly, Webb et al. (2017) used Nation's (2012, cited in Webb et al., 2017) BNC/COCA word list as the source of test items for the UVLT in replacement of the outdated items in the VLT which were derived from texts from the 1930s to the 1940s. Secondly, Webb et al. (2017) omitted the 10,000 and academic word levels and added the 1000- and 4000-word levels, introducing a new format of five word levels of 1000, 2000, 3000, 4000, and 5000. Webb et al. (2017) believed that the first five 1000-word levels were the most useful to provide a lexical profile of the learners and that knowledge of the 10,000-word level could be better depicted by a vocabulary size test. They also argued that the value of the words in the AWL varied greatly depending on the sublists that they belonged and knowledge of the words in each sublist should be measured separately. Lastly, the presentation of the test items was also changed for better test experience. Webb et al. (2017) organized the items in a table-like grid in which the test items were presented in bold horizontally and the

definitions were presented vertically, in this way, test takers could simply check the correct item box for each definition, instead of having to write down the number of the item.

Listening comprehension test

The International English Language Testing System (IELTS) is a standardized English test globally used for assessing English language proficiency of the test takers in various contexts such as education, employment, and immigration (Fernandez, 2018). The IELTS was jointly developed by the British Council, The University of Cambridge Local Examination Syndicate (UCLES) and IDP Education Australia (Pearson, 2019; Quaid, 2018). The IELTS listening test is made up of 40 questions divided into four major sections. Sections 1 and 2 include a conversation and a monolog with transactional purposes; sections 3 and 4 comprise a discussion dialog and a monolog in academic contexts. The examinees hear the audio recording only once for each section. A test section may consist of different question types ranging from multiple-choice questions, sentence completion tasks or information transfer tasks, etc., which requires different cognitive processes and assesses various aspects of listening skills, from listening fluency to listening comprehension.

Reading comprehension test

The IELTS in its current state offers two distinct reading tests, the general training module and the academic module. While the former is designed for employment and immigration purposes and assesses “basic survival skills in a broad social and educational context,” the latter is for higher education and professional purposes and assesses “the English language skills required for academic study or professional recognition” (IELTS, 2007, p. iii, cited in Moore et al., 2011). Similar to the IELTS listening test, the IELTS academic reading test is composed of three major sections, with a total of 40 questions. Each section also consists of more than one test format, assessing different aspects of reading skills and mental processing.

Procedure

Data collection

The four tests were administered to the participants in two 90-min sessions. The students completed the Vietnamese version of the LVLT and the UVLT in the first week of the class. In the following week, the IELTS listening and reading comprehension tests were administered. In the first week, 311 students completed and satisfied the data collection requirements for the LVLT and UVLT. In fact, the research was originally expected to report data from 311 participants. However, in the second week, only 234 students satisfied the data collection requirements for the IELTS listening and reading tests. All the students were well informed of the relevance and objectives of the study as well as the confidentiality, anonymity, and security of the collected data. The Vietnamese version of the LVLT and the IELTS listening comprehension tests were administered through speakers. Sound checks confirmed that all the instructions and test items were clearly heard, and at no time did the researcher and the participants encountered any difficulties hearing the recordings.

Data analysis

Firstly, a method for identifying learners' vocabulary size must be decided. The present study applied Laufer and Ravenhorst-Kalovski's (2010) method for examining learners' vocabulary size, with a small modification. Items in the five 1000-word levels represented the knowledge of 5000 word families. Therefore, each point scored on the UVLT and LVLT represented knowledge of $5000/150 = 33.3$ words and $5000/120 = 41.6$ words, respectively, I call it the words per score (WpS) value. Students' vocabulary size was calculated based on their scores on the different levels of the vocabulary tests using the following formula:

$$\text{Vocabulary size} = ((1k \times 0.8) + (2k \times 0.9) + 3k + (4k \times 1.1) + (5k \times 1.2)) \times \text{WpS value}$$

For example, based on the formula, if a student scored 23-18-16-10-17 on the five word frequency levels of the LVLT, he or she would have a knowledge of $((23 \times 0.8) + (18 \times 0.9) + 16 + (10 \times 1.1) + (17 \times 1.2)) \times 41.6 = 3416.6$ word families. Then, students were divided into vocabulary level groups based on their vocabulary size as follows: anyone with vocabulary knowledge from 500 to 1500 word families was placed at 1000 level, those with a score representing 1500-2500 word families were placed at 2000 level, anyone who received vocabulary scores between 2500 and 3500 was placed at 3000 level, those with knowledge of 3500-4500 word families were at 4000 level and participants with knowledge of above 4500 word families were placed at 5000 level. Knowledge of the AWL was analyzed separately.

Secondly, the lexical profile of the input texts from the listening and reading comprehension tests were analyzed by the *Vocabprofile* program on the Compleat Lexical Tutor website (Cobb, 2000). This program currently contains 14 versions including BNC-COCA 1-25k, CLASSIC (GSL/AWL), BNC-COCA Core-4, CEFR-English, BNL, and FRENCH v.5, 1-25k. The present study utilized the BNC-COCA 1-25k version of the program which consists of 25 frequency lists developed from the British National Corpus (BNC) and Corpus of Contemporary American English (COCA) (Nation, 2017); each list is comprised of 1000 word families. All proper nouns were excluded from the texts before the analysis was conducted. The lexical frequency analysis, therefore, assumed that the participants understood the proper nouns in the texts. The revised input texts from the listening comprehension test consisted of 3727 tokens, 986 different word types, and 778 different word families. The passages in the IELTS reading test contained 3181 tokens, 1198 different word types, and 973 different word families. In an influential paper, Chujo and Utiyama (2005) suggested that to obtain reliable text coverage information for reading materials, the minimum text length must be at least 1750 words. The texts used in the present study, therefore, satisfied the requirement.

Finally, students' raw scores were imported into SPSS. Data from participants who did not satisfy the requirements for data collection were excluded. Correlation and simple linear regression statistical techniques were applied in the present study. The Durbin Watson statistics were about 1.8 for all the analyses. The maximum Cook's distances were within the acceptable range suggested by Stevens (2002) and Tabachnick and Fidell (2001, 2007). The scatterplot of standardized residuals versus standardized

predicted values showed that the data met the assumptions of homoscedasticity and linearity.

Results

Descriptive and reliability statistics

Table 2 reports the means, standard deviations of the test results, and Cronbach’s Alpha reliability coefficients for four tests as a measure of their internal consistency.

As Table 2 shows, none of the mean scores recorded exceeded 60% of the maximum possible score. The large standard deviation suggested a reasonable spread in the scores. Reliability statistics of the two vocabulary levels tests are also high (0.92 and 0.91). The Shapiro-Wilk test of normality showed that the test data are normally distributed. These indicated that a potential ceiling effect was unlikely to be the cause for concern.

The listening and reading comprehension tests presented an appropriate level of difficulty for the test takers and displayed relatively high reliability coefficients of 0.8 and 0.73 respectively. The listening and reading comprehension tests were actual IELTS tests administered in accordance with Cambridge ESOL examination’s guidelines. Although the statistics were indeed lower than the average Cronbach’ Alpha of 0.88 recorded from the performance of more than 90,000 IELTS examinees by The University of Cambridge Local Examination Syndicate (UCLES) (2007, cited in Hashemi & Daneshfar, 2018), the reliability coefficients were within an acceptable range (0.7 or above) (Alavi et al., 2018; Pallant, 2010; Phakiti, 2016). Moreover, researchers in the field reported even lower reliability coefficients of 0.6 (Stæhr, 2009) and 0.7 (Feng, 2016) for standardized international tests provided by Cambridge ESOL examination. Taken together with the acceptable standard deviation of 6.8 and 8.4 for the listening and reading comprehension tests accordingly, these statistics may be viewed as normal and do not necessarily compromise the quality of the tests.

Research question 1: What is the relationship between aural and orthographic knowledge of vocabulary and academic listening and reading comprehension?

A Pearson product-moment correlation was run to determine the relationship between the two dimensions of vocabulary knowledge and listening and reading comprehension. Then, a Z test was performed based on Meng et al.’s (1992) method to test if there were statistically significant differences in the strength of the correlations between phonological and orthographic knowledge of vocabulary and listening and reading comprehension test scores. The results of the analyses are illustrated in Tables 3 and 4. Four simple linear regression analyses were also conducted to examine the extent to which the independent

Table 2 Means, standard deviations, and reliability coefficients (N = 234)

Test	MPS	Mean	SD	Reliability
LC	40	16.02	6.774	.805
RC	40	15.86	8.377	.735
LVLT	150	89.40	23.381	.919
UVLT	150	86.22	26.657	.910

Note. MPS maximum possible score, LC listening comprehension, RC reading comprehension, LVLT listening vocabulary levels test, UVLT updated vocabulary levels test

Table 3 Pearson correlations among scores on listening and reading comprehension and orthographic and lexical knowledge of vocabulary ($N = 234$)

Test	LC	RC	LVLT	UVLT
LC	1	-	-	-
RC	.689**	1	-	-
LVLT	.652**	.609**	1	-
UVLT	.648**	.624**	.886**	1

Note. “***” indicates that correlation is significant at 0.01 level (2-tailed). LC listening comprehension, RC reading comprehension, LVLT listening vocabulary levels test, UVLT updated vocabulary levels test

variables of aural and lexical vocabulary knowledge can explain the variance in the dependent variables of listening and reading comprehension. Overall, there were positive correlations between vocabulary test scores and listening and reading comprehension, phonological, and orthographic knowledge of vocabulary were also found to be strongly correlated at nearly .90. The correlations were statistically significant at $p < .01$ level. Relatively high correlations (approx. .65) were found between aural and orthographic vocabulary knowledge and listening comprehension. The two vocabulary level tests produced slightly lower correlations (.61 and .62) with reading comprehension. The Z test showed no statistically significant difference between the correlations ($p > .05$). Therefore, according to the results of this analysis, the two dimensions of vocabulary knowledge must be regarded as being equally correlated to listening and reading comprehension.

The simple linear regression analyses indicated that the students’ scores on the LVLT could explain up to 43% and 37% of the variance in the listening and reading test respectively. Students’ scores on the UVLT were also found to predict up to 42% and 39% of the variance in the IELTS listening and reading test correspondingly. Results from the analyses suggested significant relationships between phonological and orthographic vocabulary knowledge and academic listening and reading comprehension ($p < .001$).

Research question 2: How much vocabulary is needed for adequate academic listening and reading comprehension?

To examine the relationship between vocabulary knowledge and listening and reading comprehension, the lexical profile of the input texts was compared against students’ test scores on the two vocabulary tests and the IELTS listening and reading comprehension tests.

The results of the lexical frequency analyses of the input texts from the listening and reading tests are presented in both Tables 5 and 6. According to the analysis, the first 1000-word level covered approximately 86% and 75% of the running words in the texts from the listening and reading tests respectively. It is clear that the passages in the IELTS reading comprehension test were more lexically demanding than the input texts from the listening test. While it only took the first three 1000-word levels in the BNC/

Table 4 Difference between correlations

	LC	RC	Z	<i>p</i> (1-tailed)	<i>p</i> (2-tailed)
LVLT	.652**	.609**	1.13	.13	.26
UVLT	.648**	.624**	.633	.26	.53

Note. “***” indicates that correlation is significant at 0.01 level (2-tailed). LC listening comprehension, RC reading comprehension, LVLT listening vocabulary levels test, UVLT updated vocabulary levels test

Table 5 Relationship between the LVLT score, lexical coverage, listening and reading comprehension

Vocabulary level	No. of students	IELTS listening		IELTS reading	
		Lexical coverage	Mean score	Lexical coverage	Mean score
1000	9	86.4%	8.2 (20.5%)	75.2%	5.3 (13.3%)
2000	48	92.8%	11.5 (28.8%)	85.1%	11 (27.5%)
3000	121	95.9%	15.8 (39.5%)	91%	15.3 (38.3%)
4000	51	97.6%	21.2 (53%)	94%	22.4 (56%)
5000	5	98.1%	27.2 (68%)	95.3%	29 (72.5%)
6000	-	98.7%	-	97.5%	-
7000	-	98.9%	-	98.1%	-
8000	-	99.1%	-	98.6%	-
9000	-	99.3%	-	98.8%	-
10,000-20,000	-	99.5-99.7%	-	99-99.4%	-
Off list	-	0.3%	-	0.6%	-

Note. LVLT Listening Vocabulary Test

COCA word list (Nation, 2017) to cover 95% of the running words in the listening texts, knowledge of the most frequent 5000 word families was required to provide 95% coverage of the words in the reading passages. Coincidentally, the most frequent 5000 word families in the BNC/COCA word list were also found to make up 98% of the words in the listening texts. Students who wished to be familiar with 98% of the words in the reading passages would need to have a knowledge of the most frequent 8000 word families in the BNC/COCA word list.

Tables 5 and 6 also display the number of students who managed to be classified into different vocabulary level groups and their mean IELTS listening and reading scores. As mentioned earlier, the participants were divided into groups based on their vocabulary size by intervals of 1000 word families. Therefore, if, for instance, 4000 word families covered 94% of a text, then learners with the knowledge of 4000 word families were supposed to understand a corresponding percentage of this text (Laufer & Ravenhorst-Kalovski, 2010). Overall, it could be seen that the number of students in the vocabulary level groups formed a pyramid-like

Table 6 Relationship between the UVLT score, lexical coverage, listening and reading comprehension

Vocabulary level	No. of students	IELTS listening		IELTS reading	
		Lexical coverage	Mean score	Lexical coverage	Mean score
1000	25	86.4%	8.7 (21.8%)	75.2%	8.5 (21.3%)
2000	68	92.8%	12.1 (30.3%)	85.1%	10.3 (25.8%)
3000	92	95.9%	17.8 (44.5%)	91%	17.9 (44.8%)
4000	40	97.6%	21.1 (52.8%)	94%	22.4 (56%)
5000	9	98.1%	24.7 (61.8%)	95.3%	28.2 (70.5%)
6000	-	98.7%	-	97.5%	-
7000	-	98.9%	-	98.1%	-
8000	-	99.1%	-	98.6%	-
9000	-	99.3%	-	98.8%	-
10,000-20,000	-	99.5-99.7%	-	99-99.4%	-
Off list	-	0.3%	-	0.6%	-

Note. UVLT Updated Vocabulary Test

distribution where the proportion of participants was largest at the center and gradually decreased as the levels moved to the sides. This shape of distribution was in line with the participants' average levels of English proficiency.

The relationship between vocabulary knowledge and listening and reading comprehension are also illustrated in Tables 5 and 6. Interestingly, despite the great difference in the input texts' lexical demands, students' scores on the IELTS reading and listening tests did not seem to significantly differ. This is especially observable for the 2000, 3000, and 4000 vocabulary levels where differences greater than 5% were not spotted. Opposite to the pyramid-like distribution of participants across the vocabulary levels, we could clearly see an upward tendency of students' mean IELTS listening and reading scores, which were directly proportional to the increase of vocabulary levels. On average, an increase of 1000 word families raised the IELTS scores by 10%. However, this did not hold true for the lexical coverage of the input texts as less coverage was increased by each additional 1000 word families. One way to look at this is that each increase of 1000 word families took the learners one step closer to the optimal coverage of 95% or 98%.

It is interesting to see that both the tests of phonological and orthographic vocabulary indicated similar degree of listening and reading comprehension, especially when it is examined through the lens of IELTS band scores. For instance, both the LVLT and UVLT suggested that the 1000 level groups would score less than 10 on the listening and reading tests, which were equivalent to IELTS band scores of 3.5 or less. According to the analyses, students who knew the most frequent 2000 word families would be able to answer correctly 10-11 items in the reading tests and 11-12 items in the listening tests, which suggested an IELTS band score of 4.0. Students in 3000 level groups were shown to be likely to obtain 15-17 in both tests, which indicated an IELTS band score of 5.0. Knowledge of the most frequent 4000 word families strongly suggested an IELTS band score of 5.5, which was reflected in the consistent 21-22 scores of both the vocabulary level groups measured by different vocabulary tests. The 5000 level vocabulary group showed a range of 24-27 scores for the IELTS listening test and 28-29 for the academic reading test. While 24-27 was a wide range and could signal two possible band scores of 6.0 and 6.5, answering correctly 28-29 items on the IELTS listening and academic reading tests reliably highlighted the IELTS band score of 6.5.

Research question 3: What is the relationship between academic vocabulary and academic listening and reading comprehension?

To answer the research question, a Pearson product-moment correlation was run to determine the relationship between the knowledge of different word levels in the LVLT and academic listening and reading comprehension. The results of the analysis are presented in Table 7. Two simple linear regression analyses were also conducted to examine the extent to which the independent variable of academic vocabulary knowledge can predict the variance in the dependent variables of academic listening and reading comprehension. In addition, students were divided into different groups based on their scores on the academic word level in the LVLT by intervals of 5 points. Then, their mean scores on the IELTS listening and reading tests were examined to see if there

Table 7 Pearson correlations among scores on listening and reading comprehension and different word levels in the LVLTL ($N = 234$)

	IELTS listening test	IELTS reading test
1000 level	.455**	.395**
2000 level	.593**	.530**
3000 level	.571**	.543**
4000 level	.582**	.551**
5000 level	.472**	.485**
AWL level	.648**	.595**

Note. “***” indicates that correlation is significant at 0.01 level (2-tailed) ($N = 234$), AWL Academic Word List

were any changes in academic listening and reading comprehension as knowledge of academic vocabulary went up. The findings are illustrated in Table 8.

Statistically significant correlations of .65 and .60 were found between academic vocabulary knowledge and academic listening and reading comprehension respectively. Moreover, it can be seen that knowledge of academic vocabulary had stronger correlations with reading and listening comprehension than knowledge of any other word level. The simple linear regression analyses showed significant relationships between academic word knowledge and academic listening and reading comprehension ($p < .001$). Results from the analyses also indicated that students’ scores on the academic level in the LVLTL could predict up to 42% and 35% of the variance in the IELTS listening and reading test scores correspondingly.

Overall, Table 8 shows a strong relationship between knowledge of academic vocabulary and academic listening and reading. On average, an increase of 5 points in the academic word level would raise the IELTS band scores on the listening and academic reading tests by 0.5. More notably, the mastery of the academic vocabulary level, which was indicated by the threshold of 26 correct answers out of 30 items (Schmitt et al., 2001), would suggest an IELTS band score of 6.0 for both the listening and reading tests.

Discussion

The present study confirmed the strong correlation between orthographic and phonological knowledge of vocabulary ($r = .88$) and highlighted the significant relationships between phonological and orthographic knowledge of vocabulary and reading and listening comprehension. In general, the results indicated that both orthographic and phonological knowledge of vocabulary were strongly correlated with academic listening

Table 8 Relationship between academic vocabulary knowledge and academic listening and reading comprehension

AWL score	No. of students	IELTS listening		IELTS reading	
		IELTS band score	Mean score	IELTS band score	Mean score
26-30	12	6.0	24	6.0	25.9
21-25	44	5.5	21.3	5.5	22.4
16-20	62	5.0	17.9	5.0	17.9
11-15	58	4.5	14.2	4.0	12.8
6-10	41	3.5	10.6	3.5	9.7

Note. AWL Academic Word List

and reading comprehension. In addition, the two dimensions of vocabulary knowledge could explain approximately 40% of the variance in the IELTS listening and reading tests. When compared with reading comprehension, the findings suggested that listening comprehension had stronger correlations with vocabulary knowledge. Moreover, scores on vocabulary tests were also found to predict more of the variance in a listening test (about 5%) than in a reading test.

Although the research findings were consistent with those of Noreillie et al. (2018), Feng (2016) and Stæhr (2009) who examined the relationship between receptive vocabulary knowledge measured by written vocabulary tests and listening comprehension, they were, to some extent, contradictory to Milton et al.'s (2010) study that addressed the same cross-modality issue. In actual practice, the current study did record positive, strong single-modality correlations between orthographic word knowledge and reading comprehension as well as between aural vocabulary knowledge and listening comprehension, which were in line with Milton et al.'s (2010) and Cheng and Matthews's (2018) studies. However, the results did not show any significant cross-modality differences, in fact, the analyses showed similarities between single-modality and cross-modality correlations, which hypothesized that phonological and orthographic word knowledge were of equal value to listening and reading comprehension, at least in EFL contexts, where learners' exposures to English was limited and most of the input came from the classrooms. The hypothesis was also partially supported by Cheng and Matthews's (2018) findings.

The present study added knowledge to the relationship between correlations and participants' levels of English proficiency. The present study collected data from second-year university students with the average level of English proficiency at B1 and reported a strong correlation ($r = .65$) between participants' vocabulary knowledge and listening comprehension. The analyses also found that the correlations between knowledge of the 1000- and 5000-word levels and academic listening comprehension were lower than the correlation between the IELTS listening test scores and students' scores on the whole LVLT. Results from the Z test showed that the differences were statistically significant ($p = .000$, 2-tailed). The 1000-word level was the easiest level in the LVLT and most students at the B1 level are expected to have considerable knowledge of the 1000 band. On the contrary, the 5000-word level was believed to be the most challenging level where students were least likely to achieve high scores. Moreover, it can be seen that similarly strong correlations were reported by studies that included students with English proficiency levels from B1 and above. All of these lead to two assumptions: the first one is that learners need to be familiar with the words in a specific vocabulary level to a certain extent before the correlation between the knowledge of that particular word level and comprehension could be appropriately recorded. This means that vocabulary levels that are too easy or too difficult would be likely to yield biased results. Another assumption is that the B1 (or intermediate) level is the threshold at which the participants' test scores could provide sufficient data for the investigation of the relationship between vocabulary knowledge and comprehension. Obviously, more research using multiple measures of vocabulary knowledge and comprehension which include participants of different language proficiency cohorts is needed to uncover these hypotheses.

The present study did not use a universal cut-off score as a tool for estimating learners' vocabulary size. Setting a "general" threshold for mastery, whether stringent or lenient, was supported by researchers (Stæhr, 2009; Webb & Chang, 2012) because it allowed the vocabulary levels to classify or rank the examinees in accordance with a hypothesized order of difficulty, and test takers would not be necessarily "excluded" but "moved" up and down the vocabulary levels (they were only excluded in case they failed to master the 1000 level). While the rationale for using the cutting points as thresholds for the mastery of a particular vocabulary level may sound convincing to a certain extent, this kind of analysis could falsify the concept of "mastering" a vocabulary level and give a blurry image of learners' vocabulary size. The reason lies with the process of excluding under- and overqualified candidates, which I personally consider to be too strict and may result in inappropriate ranking. For example, if I decided to set a score of 23/24 as the universal cutting point for the 1000, 2000, 3000, 4000, and 5000 word levels in the LVLTL and I would like to know how many students have mastered the fourth 1000-word level, then I would have to exclude students who scored 22/24 and below for the 1000, 2000, and 3000 word levels as they were underqualified, I would also have to exclude those who answered correctly more than 22/24 items in the 5000 word level since they were overqualified. In this way, even if a student scored 24-24-22-22-20 on the 1000, 2000, 3000, 4000, and 5000 word levels respectively, he or she would still be excluded from the 3000 and 4000 word levels and ranked as "only mastered the 1000 and 2000 word levels" for being two-point different from someone who scored 24-24-23-23-20 in the order given.

Webb et al.'s (2017) flexible method of setting cutting point for the UVLT, which gives a cutting score of 29/30 for the 1000, 2000, and 3000 word levels and a cutoff point of 24/30 for the 4000 and 5000 vocabulary levels, could lead to even greater constraints. First, if the creator of a vocabulary levels test hypothesized that higher frequency levels would be easier than lower frequency levels, then it is natural to expect most of the learners who mastered the 4000 and 5000 levels to satisfy the requirement for the mastery of 1000, 2000, and 3000 word levels. However, things would get complicated if a substantial number of examinees who were considered to have "mastered" the 4000 and 5000 word levels failed to obtain the necessary score for the mastery of the 3000, 2000, or even 1000 levels. Those students would not be able to stay in the lower frequency levels (e.g., 4000 and 5000 levels) since they were "disqualified" for not mastering the preceding levels, but they could not be "pushed" down to the higher frequency levels (e.g., 1000, 2000, or 3000 levels) either as they were "accidentally qualified" for the 4000 and 5000 levels. This conflict of qualification clearly goes against the hypothesized order of difficulty. The scenario could also lead to a considerable number of test takers at a certain word level being unnecessarily excluded. When the thresholds for mastery proposed by Webb et al. (2017) for the UVLT were trialed on the 311 participants who completed the UVLT, 23 out of the 28 students who mastered the 4000-word level were excluded for not satisfying the requirements for the mastery of the higher frequency levels. Similarly, 16 out of 20 participants who scored 24 and above on the 5000 level were disqualified for not being able to achieve the score of 29/30 on either the 3000, 2000, or 1000 word level. The proportion of participants who mastered the 3000 level was also found to be the smallest, with only 7 students, but more importantly, all these 7 students achieved the mastery requirements for either the 4000 or

5000 level. It can be seen that the 3000-word level was made the most challenging word level and was “squeezed” by the stringent 29/30 cutting point applied to the higher frequency levels and the lenient 24/30 cutoff point suggested for the lower frequency levels.

While using cut-off scores could severely limit the sample and the potential of a study’s data analysis, giving each participant a vocabulary score based on his or her overall performance on the word levels in a vocabulary test could give a better view of the situation. Although students could still be placed at distinct word levels for small differences, they would only be moved one level up or down. And the possibility that a 2- or 3-level distance might be created between two students with only one- or two-point difference would never happen.

While Laufer and Ravenhorst-Kalovski (2010) believed that points scored on different word levels hold equal values, I would argue that points scored on the lower frequency levels (4000 and 5000) tell us more about a students’ vocabulary than those scored on the higher ones (1000 and 2000), and therefore, should be of greater weight. With the formula used in the study, the scores on lower frequency levels would have 10% greater value than the higher frequency level next to it, and scores on the 5000-word level would be of 40% greater value than scores on the 1000 level. This difference is not only significant enough to give scores on the low frequency levels some influence over the estimation of students’ vocabulary size but also small enough not to give them the power to overwhelm higher frequency levels.

The concept of “adequate comprehension” is another matter of judgment and each researcher had his or her justification when setting a threshold for reasonable comprehension, which, in most cases, was the minimum passing grade in the testing system of the institution or country where they worked or conducted the study. While the rationale for Laufer’s (1989) 55% threshold for reasonable reading comprehension was that this cutting point represented the lowest passing grade in the Haifa University system, Hazenberg and Hulstijn’s (1996) study used the minimum passing score in a reading test from a Dutch language university entry examination as the threshold for adequate comprehension, which was 70%. Laufer and Ravenhorst-Kalovski (2010) used the strict score of 134/150, which was nearly 90%, on the Psychometric Entrance Test created by the National Institute for Testing and Evaluation (NITE) in Israel as the threshold that would ensure adequate reading comprehension. Their justification was that the 134/150 score would exempt students from studying English as a foreign language. Similarly, in his study, Stæhr (2009) utilized the listening comprehension test from the Cambridge Certificate of Proficiency in English and also used a score of 70%, which was equivalent to a grade C, to represent reasonable comprehension.

In actual practice, a universal vocabulary threshold for comprehension does not exist and different learning goals or objectives may require the mastery of different vocabulary levels. The present study used the IELTS listening and academic reading tests as the instruments for the measurement of learners’ listening and reading proficiency, and these abovementioned percentages can lead to different interpretations of the score. For example, the 55% (22/40) cutoff point represents the band score of 5.5 in both the IELTS listening and academic reading tests, which has been widely used in Vietnamese universities as a graduation requirement for non-English majors. This threshold (IELTS 5.5) has also been applied as a minimum English requirement for officials in

universities or government-related sectors in Vietnam. However, the acceptance of this band score is relatively regional and could not be applied to an international context. On the other hand, the cutting points of 23/40 (57.5%) or 28/40 (70%) of the maximum possible score, which indicate the band scores of 6.0-6.5 for the IELTS listening and academic reading tests (UCLES, 2019), have been globally used and accepted as the minimum language requirement for international students at undergraduate and post-graduate levels by most universities.

Instead of headbutting the thorny question of “how much lexical coverage and/or vocabulary knowledge is needed for “adequate” listening and reading comprehension?,” the present study’s focus is on the linear relationship between lexical coverage, vocabulary knowledge, and listening and reading comprehension. For people who need a large lexical resource for entering universities in English speaking countries or other academic purposes, knowledge of the most frequent 5000 word families is generally recommended. For professionals who would like to apply for office jobs that demand a certain degree of English proficiency in Vietnam, knowing the most frequent 3000 or 4000 word families in the BNC/COCA word list (Nation, 2017) may be the requirement. It is worth noting that the mere knowledge at 2000 word families level is not likely to result in acceptable listening or reading comprehension in any situation.

The study also highlighted the importance of academic vocabulary to academic listening and reading comprehension. Academic vocabulary knowledge was found to be strongly correlated with academic listening and reading at .65 and .60 correspondingly. Furthermore, knowledge of the AWL alone could predict up to 42% and 35% of the variance in the IELTS listening and reading tests respectively. Besides, the mastery of the academic word level in the LVLT can reliably suggest an IELTS band score of 6.0 in both the listening and reading tests. Most importantly, the strong relationship between the scores on the academic level in the LVLT and the IELTS listening and reading tests strongly confirmed the role of academic vocabulary as a reliable predictor of successful academic listening and reading comprehension.

Conclusion

The present study has provided empirical evidence for the strong relationship between receptive vocabulary knowledge and receptive language skills, confirming the major contribution of vocabulary to successful listening and reading comprehension. The study also shed light upon how universities and organizations may use the vocabulary level tests as instruments for measuring vocabulary knowledge. In fact, with the empirical evidence for the strong link between receptive vocabulary knowledge and learners’ language ability, which have been built for decades (Nation, 2013; Webb, 2020), vocabulary tests have been proven to be very valid, reliable, and powerful tools for the estimation of learners’ language proficiency. And on top of that, all of them are convenient, free (at least for the ones used in this study) and have the potential to be administered in both paper- and computer-based, online format. Take the LVLT for an example, it only takes approximately 30 min to administer (McLean et al., 2015), can be easily developed to be delivered in computer-based, online format, and gives information on learners’ knowledge of the AWL (Coxhead, 2000) and the most frequent 5000 word families in the BNC/COCA word list (Nation, 2017). Scores on vocabulary tests can be interpreted in different ways, either using a cut-off score for the word

levels (McLean & Krammer, 2015) or calculated as a whole using the formula suggested in this study. Institutions can use a test of receptive vocabulary knowledge in combination with other tests of English to obtain a broad picture of learners' English proficiency from different aspects. Vocabulary level tests could also be administered in isolation and can still be a really powerful predictor of students' listening and reading comprehension, as shown in this study.

Despite the helpful findings, the study itself has certain limitations. Firstly, results from the regression analyses indicated that vocabulary knowledge can only predict about 40% of the variance in the scores on the listening and reading test. This points out that approximately 60% of the variance in the tests are explained by other factors. Moreover, despite being less lexically demanding, listening test scores were found to be lower than reading scores among certain vocabulary cohorts. This may indicate that different use of compensation strategies can facilitate learners' comprehension to a considerable extent (Stæhr, 2009). The contribution of those strategies is even more noticeable when it comes to reading comprehension as only 37-39% of the variance in the reading test can be explained by vocabulary knowledge. A more comprehensive study that compares and sheds light upon the effects of receptive compensation strategies on listening and reading comprehension is needed to uncover this myth.

Secondly, in the present study, the knowledge of academic words was merely measured by a test of aural vocabulary. This may be the reason why students' scores on the academic word level tended to have better correlations with and could explain more variance in the IELTS listening test. Further comparison using different tests of academic vocabulary and academic listening and reading comprehension is demanded to confirm the assumption. Finally, the relatively small sample in certain vocabulary groups (1000, 5000 level groups and the 26-30 AWL group) may have limited the generalizability of the findings. Future studies should re-investigate the issue with larger cohort of participants.

Abbreviations

AWL: Academic Word List; BASE: British Academic Spoken English; BNC: British National Corpus; COCA: Corpus of Contemporary American English; EFL: English as a foreign language; ESL: English as a second language; ESOL: English for Speakers of Other Languages; IELTS: International English Language Testing System; LVL: Listening Vocabulary Levels Test; TOEIC: Test of English for International Communication; UCLES: University of Cambridge Local Examination Syndicate; UVLT: Updated Vocabulary Levels Test; VST: Vocabulary Size Test

Acknowledgements

Not applicable.

Author's contributions

The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation. The author read and approved the final manuscript.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The author declares that he has no competing interests.

Received: 20 April 2021 Accepted: 21 June 2021

Published online: 14 September 2021

References

- Alavi, S. M., Kaivanpanah, S., & Masjedlou, A. P. (2018). Validity of the listening module of international English language testing system: Multiple sources of evidence. *Language Testing in Asia*, 8(8) <https://doi.org/10.1186/s40468-018-0057-4>.
- Barcroft, J. (2007). When knowing grammar depends on knowing vocabulary: Native speaker grammaticality judgements of sentences with real and unreal words. *The Canadian Modern Language Review / La revue canadienne des langues vivantes*, 63(3), 313–343 <https://doi.org/10.3138/R601-H212-5582-0737>.
- Bian, X., Cai, X., & Cai, D. (2021). The contributions of listening and reading vocabularies to listening comprehension of Chinese EFL students. *International Journal of Listening*, 35(2), 110–122 <https://doi.org/10.1080/10904018.2019.1623678>.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- 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>.
- Chujo, K., & Utiyama, M. (2005). Understanding the role of text length, sample size and vocabulary size in determining text coverage. *Reading in a Foreign Language*, 17, 1–22.
- Cobb, T. (2000). The Compleat lexical tutor [Website]. Available at: <https://lexutor.ca/> (accessed 22 Feb 2021).
- Coxhead, A. (2000). A new Academic Word List. *TESOL Quarterly*, 34(2), 213–238 <https://doi.org/10.2307/3587951>.
- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33(1), 66–76 <https://doi.org/10.1016/j.esp.2013.08.001>.
- Feng, T. (2016). An in-depth investigation into the relationship between vocabulary knowledge and academic listening comprehension. *The Electronic Journal for English as a Second Language*, 20(2).
- Fernandez, C. J. (2018). Behind a spoken performance: Test takers' strategic reactions in a simulated part 3 of the IELTS speaking test. *Language Testing in Asia*, 8(18) <https://doi.org/10.1186/s40468-018-0073-4>.
- Fountain, R. L., & Nation, I. S. P. (2000). A vocabulary-based graded dictation test. *RELC Journal*, 31(2), 29–44 <https://doi.org/10.1177/003368820003100202>.
- Guo, Y., & Roehrig, A. D. (2011). Roles of general versus second language (L2) knowledge in L2 reading comprehension. *Reading in a Foreign Language*, 23(1), 42–64.
- Ha, H. T. (in press). A Rasch-based validation of the Vietnamese version of the listening vocabulary levels test. *Language Testing in Asia*.
- Hashemi, A., & Daneshfar, S. (2018). A review of the IELTS test: Focus on validity, reliability, and washback. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, 3(1), 39–52 <https://doi.org/10.21093/ijeltal.v3i1.123>.
- Hazenbergh, S., & Hulstijn, J. H. (1996). Defining a minimal second language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics*, 17(2), 145–163 <https://doi.org/10.1093/applin/17.2.145>.
- Henning, G. (1991). *A study of the effects of contextualization and familiarization on responses to the TOEFL vocabulary test items*. Princeton: Educational Testing Service.
- Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *The Language Learning Journal*, 36(2), 153–166 <https://doi.org/10.1080/09571730802389983>.
- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430.
- IELTS (2007). *IELTS official practice materials*. Cambridge: Cambridge University Press.
- Lange, K., & Matthews, J. (2020). Exploring the relationships between L2 vocabulary knowledge, lexical segmentation, and L2 listening comprehension. *Studies in Second Language Learning and Teaching*, 10(4), 723–749 <https://doi.org/10.14746/ssl.2020.10.4.4>.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren, & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines*, (pp. 316–223).
- Laufer, B. (2013). Lexical thresholds for reading comprehension: What they are and how they can be used for teaching purposes. *TESOL Quarterly*, 47(4), 867–872 <https://doi.org/10.1002/tesq.140>.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- Lewis, M. (2002). *Implementing the Lexical approach: Putting theory into practice*. Boston, MA: Thomson Heinle.
- Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, 52, 1–13 <https://doi.org/10.1016/j.system.2015.04.015>.
- McLean, S., & Kramer, B. (2015). The creation of a new vocabulary levels test. *Shiken*, 19(1), 1–11.
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19(6), 741–760 <https://doi.org/10.1177/1362168814567889>.
- Meara, P., & Milton, J. (2003). *X_Lex, The Swansea levels test*. Newbury: Express.
- Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111(1), 172–175 <https://doi.org/10.1037/0033-2909.111.1.172>.
- Milton, J., & Hopkins, N. (2005). *Aural Lex*. Swansea: Swansea University.
- Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some teachers. *The Canadian Modern Language Review*, 63, 127–147.
- Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, M. Torreblanca-López, & M. López-Jiménez (Eds.), *Further insights into non-native vocabulary teaching and learning*, (pp. 83–97). Bristol, UK: Multilingual Matters <https://doi.org/10.21832/9781847692900-007>.
- Moore, T., Morton, J., & Price, S. (2011). Construct validity in the IELTS Academic reading test: A comparison of reading requirements in IELTS test items and in university study. *IELTS collected papers*, 2, 120–211.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press <https://doi.org/10.1017/CBO9781139524759>.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59–82 <https://doi.org/10.3138/cmlr.63.1.59>.

- Nation, I. S. P. (2012). The BNC/COCA word family lists. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>.
- Nation, I. S. P. (2013). *Learning vocabulary in another language*, (2nd ed.,). Cambridge: Cambridge University Press <https://doi.org/10.1017/CBO9781139858656>.
- Nation, I. S. P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. Available at: <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists> (accessed 22 Feb 2021).
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31, 9–13.
- Noreillie, A. S., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages: An approximate replication study of Stæhr (2009). *ITL - International Journal of Applied Linguistics*, 169(1), 212–231 <https://doi.org/10.1075/itl.00013.nor>.
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Buckingham: Open University Press.
- Pearson, W. S. (2019). 'Remark or retake'? A study of candidate performance in IELTS and perceptions towards test failure. *Language Testing in Asia*, 9(17) <https://doi.org/10.1186/s40468-019-0093-8>.
- Peters, E., Velghe, T., & Van Rompaey, T. (2015). *A post-entry English and French vocabulary size for Flemish learners*. Copenhagen, Denmark: Paper presented at EALTA.
- Peters, E., Velghe, T., & Van Rompaey, T. (2019). The VocabLab tests: The development of an English and French vocabulary test. *ITL - International Journal of Applied Linguistics*, 170(1), 53–78 <https://doi.org/10.1075/itl.17029.pet>.
- Phakiti, A. (2016). Test-takers' performance appraisals, appraisal calibration, state-trait strategy use, and state-trait IELTS listening difficulty in a simulated IELTS Listening test. *IELTS Research Reports Series*, 6, 1–140.
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3), 513–536 <https://doi.org/10.1111/1467-9922.00193>.
- Qian, D. D., & Lin, L. H. F. (2020). The relationship between vocabulary knowledge and language proficiency. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*, (pp. 66–80).
- Quaid, E. D. (2018). Reviewing the IELTS speaking test in East Asia: Theoretical and practice-based insights. *Language Testing in Asia*, 8(2) <https://doi.org/10.1186/s40468-018-0056-5>.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26–43 <https://doi.org/10.1111/j.1540-4781.2011.01146.x>.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–89 <https://doi.org/10.1177/026553220101800103>.
- Stæhr, L. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, 36(2), 139–152 <https://doi.org/10.1080/09571730802389975>.
- 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 <https://doi.org/10.1017/s0272263109990039>.
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*, (4th ed.,). Mahwah, NJ: LEA.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*, (4th ed.,). Boston, MA: Allyn and Bacon.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*, (5th ed.,). Boston: Allyn & Bacon/Pearson Education.
- UCLES. (2007). IELTS handbook 2007. Retrieved from the IELTS Website: http://www.ielts.org/pdf/IELTS_Handbook_2007.pdf.
- UCLES. (2019). *IELTS guide for teachers*. Available at: <https://www.ielts.org/-/media/publications/guide-for-teachers/ielts-guide-for-teachers-ukashx> ().
- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457e479 <https://doi.org/10.1093/applin/ams074>.
- Webb, S. (Ed.) (2020). *The Routledge handbook of vocabulary studies*. London: Routledge.
- Webb, S., Sasao, Y., & Balance, O. (2017). The updated vocabulary levels test. *ITL - International Journal of Applied Linguistics*, 168(1), 33–69 <https://doi.org/10.1075/itl.168.1.02web>.
- Webb, S. A., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, 43(1), 113–126 <https://doi.org/10.1177/0033688212439367>.
- Zhang, D. (2012). Vocabulary and grammar knowledge in second language reading comprehension: A structural equation modeling study. *The Modern Language Journal*, 96(4), 558–575 <https://doi.org/10.1111/j.1540-4781.2012.01398.x>.
- Zhang, D., & Koda, K. (2013). Morphological awareness and reading comprehension in a foreign language: A study of young Chinese EFL learners. *System*, 41(4), 901–913 <https://doi.org/10.1016/j.system.2013.09.009>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)