Reading in a Foreign Language April 2020, Volume 32, No. 1
ISSN 1539-0578 pp. 1–27

The role of vocabulary breadth and depth in IELTS academic reading tests

Chen Chen
Xi'an Jiaotong-Liverpool University
China

Yongcan Liu
University of Cambridge
United Kingdom

Abstract

This study explored the role of vocabulary breadth and depth in second language learners' performance in IELTS academic reading tests in China. Sixty-two Chinese learners of English as a foreign language completed a vocabulary size test, a vocabulary depth test, and an IELTS reading test. Results showed that vocabulary breadth and depth both correlated significantly with IELTS reading test scores. With regard to different IELTS question types, vocabulary breadth correlated significantly with True/False/Not Given questions, whereas vocabulary depth correlated more significantly with Multiple Choice, Matching Headings, and Sentence Completion tasks. Results of a multiple regression model indicated the increase of vocabulary size needed to improve certain IELTS band scores. This study has theoretical implications for broadening the conceptualisation of vocabulary depth, as well as pedagogical implications for supporting students' second language reading development.

Keywords: vocabulary breadth, vocabulary depth, reading comprehension, IELTS examination

In second language (L2) reading research, a significant body of work has been devoted to the relationship between vocabulary knowledge and reading comprehension, which have been consistently reported to significantly correlate with each other (Alavi & Akbarian, 2012; Chiang, 2018; Karakoç & Köse, 2017; Proctor, Silverman, Harring, & Montecillo, 2012; Qian, 1999, 2002). Anderson and Freebody (1983) first proposed two dimensions of vocabulary knowledge: vocabulary breadth, which refers to the number of words that the learner knows (at least some aspects of the meaning), and vocabulary depth, which refers to the quality of the meaning that the learner knows. In research on the role of vocabulary in reading, vocabulary knowledge has been mostly measured by vocabulary breadth, and it is only since the beginning of the 21st century that vocabulary depth has been taken into consideration (Li & Kirby, 2015; Proctor et al., 2012; Qian, 1999, 2002). However, debate still exists among researchers as to how the construct of vocabulary depth should be conceptualised and operationalised (Li & Kirby, 2015; Proctor et al., 2012). Moreover, the question of how vocabulary knowledge is related to performance in different reading comprehension tasks, as well as the question of how increases in vocabulary knowledge can predict performance in reading comprehension tasks, have been less extensively researched (Alavi & Akbarian, 2012; Laufer & Ravenhorst-Kalovski, 2010; Li & Kirby, 2015).

To address these issues, this study used correlation analysis to explore the relationship between vocabulary breadth and depth and International English Language Testing System (IELTS) academic reading examination results. In addition, the study employed multiple regression analysis to make predictions of how increases in vocabulary size can contribute to IELTS reading test scores. This research has pedagogical implications for improving students' vocabulary learning and L2 reading comprehension.

Literature review

Vocabulary breadth and vocabulary size tests

Vocabulary knowledge is complex and multifaceted. Recognising two primary dimensions of vocabulary knowledge—vocabulary breadth and depth—has become essential in understanding the relationship between vocabulary knowledge and reading comprehension (Qian, 2002). Anderson and Freebody (1983) defined vocabulary breadth as the number of words *known* by a learner. However, they did not clarify to what extent a word needs to be known to be counted as 'known'. The level of knowing a word can range from knowing only the form–meaning linkage of a word—which is considered as the most basic level of vocabulary knowledge (Schmitt, 2010a)—to knowing all components of a word, including its meaning, written and spoken forms, morphological knowledge, collocations, register, associations, and frequency (Nation, 2013). Most previous studies that focused on vocabulary breadth and reading comprehension defined vocabulary breadth as the form–meaning connection for a certain number of words (e.g., Li & Kirby, 2015; Qian, 2002); that is, learners are able to recall at least one aspect of the meaning of the word by its given form. This study will continue using this definition of vocabulary breadth.

Three types of vocabulary size tests have been widely used to measure learners' vocabulary breadth in English-as-a-second-language (ESL) research contexts, each having its advantages and disadvantages. The first test type is the checklist test where learners tick the words they know among a list of words. Meara and his colleagues have developed a number of checklist tests, for example, the Eurocentres Vocabulary Size Test (Meara & Jones, 1990). This test is efficient in measuring learners' vocabulary sizes within a short period of time. However, the test format depends on the test-takers' self-reports, which do not accurately measure their vocabulary breadth. The second type is the matching definitions test, among which the Vocabulary Levels Test (VLT) designed by Nation (1983) and a newer version later designed by Schmitt, Schmitt, and Clapham (2001) have been widely used to investigate the vocabulary–reading relationship (e.g., Karakoç & Köse, 2017; Qian, 1999, 2002; Stæhr, 2008). Later versions based on these have also been developed and validated, including the Listening Vocabulary Levels Test by McLean, Kramer, and Beglar (2015), and Webb, Sasao, and Ballance's (2017) two new forms of the VLT. The third test type is the multiple-choice format. The Vocabulary Size Test (VST) developed by Nation and Beglar (2007) has been the most widely used (Schmitt, 2010b, p. 198). While the VLT estimates learners' vocabulary breadth at different frequency levels, the VST can measure the overall vocabulary size of a learner. It was proved to be more representative of learners' vocabulary breadth than the definition-matching format (Kremmel, 2015).

In vocabulary size tests, it is also necessary to specify the unit of counting, as using different units to measure vocabulary size may lead to varying estimates in the vocabulary size of the

test taker. Nation (2007) suggested that lemmas, which consist of a base word and its inflected forms, serve as a valid unit to count learners' productive use of language because different grammatical constructions and collocations produced by a learner should be counted separately for different uses. However, word families, which include not only a base form and inflected forms but also derivative forms, are more suitable for measuring receptive understanding (Bertram, Baayen, & Schreuder, 2000). Moreover, using lemmas in receptive word knowledge may result in an overestimation of a learner's vocabulary breadth (Nation, 2007). Most previous studies on the relationship between vocabulary and reading have been based on the counting unit of word families (Alavi & Akbarian, 2012; Chiang, 2018; Laufer, 1992; Zhang & Anual, 2008). Thus, in order to allow for comparison with other studies and due to the reasons mentioned above, this study continued with this unit of counting.

Vocabulary depth and vocabulary depth tests

Vocabulary depth has been conceptualised and assessed by different researchers in various ways. In this paper, a framework of vocabulary depth was established, taking into account the merits of previous frameworks based on three approaches to conceptualising the construct—precision of meaning, comprehensive word knowledge, and network knowledge (Read, 2004).

Precision of meaning indicates the proximity of a learner's understanding of a word and the exact definition of the word (Read, 2004). It is also known as the developmental approach and reflects the learner's different stages of learning, from not knowing the word at all to full mastery of how to use the word with semantic appropriateness and grammatical accuracy (Paribakht & Wesche, 1997). This approach can indicate how well a learner understands a word.

Comprehensive word knowledge, also referred to as the componential approach, attempts to characterise word knowledge by listing different aspects or components of a word (Read, 2004). Since as early as the 1940s, researchers have attempted to outline the components of depth knowledge. They have included (a) understanding the different subtleties of the meaning of a word and being able to use the word appropriately (Cronbach, 1942); (b) a word's associations, derived forms, collocations and connotations (Richards, 1976); (c) form, meaning, and use, with each of the sub-categories covering both receptive and productive word knowledge (Nation, 2013); and (d) morphological, syntactic, and semantic word knowledge (Proctor et al., 2012). Taking into consideration the feasibility of operationalisation and the significant roles they play in reading comprehension, this study incorporated morphological, syntactic, and semantic word knowledge into the sub-construct of *comprehensive word knowledge*. This is because morphological awareness can help learners guess word meaning based on its root, prefix, and suffix; syntactic word knowledge facilitates reading fluency (Mokhtari & Thompson, 2006); and semantic word knowledge enables learners to determine the meaning of a word—particularly in the case of polysemy—in specific contexts (Proctor et al., 2012). While focusing on these three components, we also acknowledge other linguistic features which are equally important in understanding vocabulary depth knowledge. As a matter of fact, due to the overlap between the three dimensions of vocabulary depth, these linguistic features have been incorporated in the other two sub-constructs: precision of meaning and network knowledge.

Network knowledge was originally proposed by Meara (1992, 1996) who referred to this aspect of vocabulary depth as *organisation*, and it was expanded further by other researchers (Dóczi & Kormos, 2016; Meara & Fitzpatrick, 2000; Schmitt, 1998; Zhang & Koda, 2017).

In this approach, vocabulary depth knowledge is seen as the ability to link a word with other associated words and to incorporate a newly acquired word into a network of already known words, which is known as a mental lexicon (Schmitt, 2014). This approach is different from the other two as it focuses more on the connection between individual words and the mental lexicon (Read, 2004).

There is some overlap among the three components, but they complement each other to form a comprehensive framework of vocabulary depth knowledge. Choosing just one of these approaches would not be adequate to represent this complex and multifaceted construct. Based on vocabulary breadth and the three dimensions of vocabulary depth discussed above, therefore, a conceptual framework of vocabulary knowledge for this study is proposed in Figure 1.

In terms of the measurement of vocabulary depth, most vocabulary depth tests only assess part of the construct. Typical examples include the Vocabulary Knowledge Scale (Wesche & Paribakht, 1996) which is used to measure precision of meaning; the Test of English Derivatives (Schmitt, 2010b, p. 228) which is used to measure morphological knowledge; and the Word Associates Format (Read, 1993, 1998) and Collex and Collmatch (Gyllstad, 2007) which are used to measure network word knowledge. However, no single test covers all aspects of this complex construct.

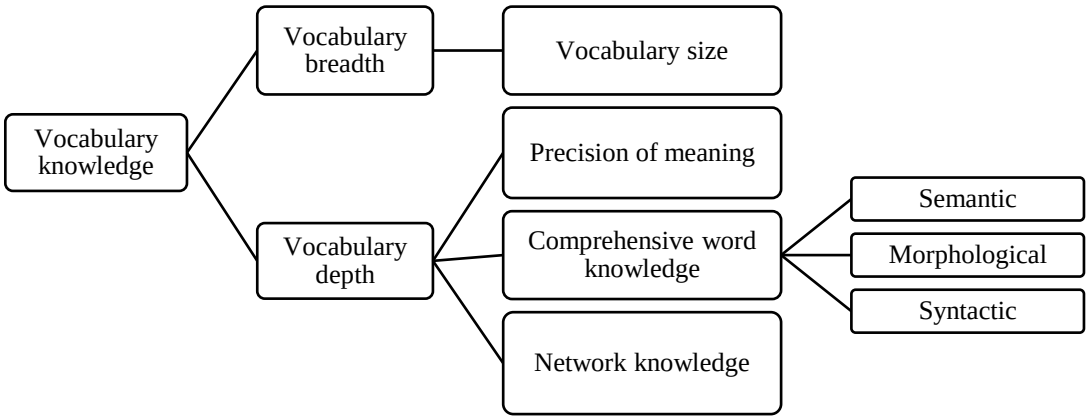


Figure 1. Conceptual framework of vocabulary knowledge

Vocabulary knowledge and reading comprehension

Relationship between vocabulary breadth and depth and reading comprehension. Various studies have consistently found significant correlations between vocabulary knowledge and L2 reading comprehension (Alavi & Akbarian, 2012; Chiang, 2018; Laufer, 1992; Li & Kirby, 2015; Qian, 1999, 2002). However, the majority of the studies investigated only the role of vocabulary breadth in reading comprehension (Alavi & Akbarian, 2012; Laufer, 1992; Zhang & Anual, 2008); only since the beginning of the 21st century has vocabulary depth gained research attention in relation to reading comprehension (Bardakci, 2016; Cakir, Unaldi, Arslan, & Kilic, 2016; Li & Kirby, 2015; Qian, 1999, 2002). Due to the fact that there are many different components of vocabulary knowledge, measuring vocabulary size does not give a comprehensive picture of a learner's vocabulary knowledge. The results of the recent studies that included vocabulary depth in their design indicate that vocabulary depth is another important factor which significantly influences reading comprehension.

However, there has been no consensus as to how vocabulary depth should be conceptualised. Different conceptual frameworks of vocabulary depth have been established and measured by different instruments. Qian (1999) gave a detailed definition of vocabulary depth including six components, but the assessment tool he used—a revised version of the Word Associates Format (Read, 1993)—only measured part of the construct he defined. Li and Kirby (2015) included morphological knowledge, multiplicity, and precision of meaning in their definition of the construct and employed three instruments to measure the three components. Therefore, a clear conceptualisation of vocabulary depth and appropriately-designed measurement tools are needed in order to explore its relationship with L2 reading comprehension.

As for the relationship between vocabulary knowledge and different international reading examinations, most previous literature has examined the role of vocabulary in Test of English as a Foreign Language (TOEFL) (e.g., Alavi & Akbarian, 2012; Qian, 2002; Rashidi & Khosravi, 2010), whereas few studies have investigated the role of vocabulary in IELTS tests (Akbariam & Alavi, 2013; Milton, Wade, & Hopkins, 2010). The reading comprehension tasks in the IELTS reading test—which include True/False/Not Given questions, Multiple Choice, Sentence Completion Questions, Short Answer Questions, and Matching Headings—differ greatly from those in the TOEFL test, which consists only of Multiple Choice. Different measures of reading comprehension require different cognitive processes (Pearson & Hamm, 2005). In addition, a large number of non-native speakers who seek to study in English-speaking countries are required to reach a certain IELTS band score to be eligible for enrolment into study programmes in institutions. The test is recognised by over 10,000 universities and other organisations, and the number of test-takers reached 3 million in 2017 ('IELTS numbers rise to three million a year', 2018). For these reasons, exploring the relationship between vocabulary knowledge and IELTS reading test results can provide pedagogical implications that can benefit the large number of IELTS exam takers. Among the very few studies that have examined the relationship between vocabulary knowledge and IELTS test scores, there is a lack of focus on the relationship between vocabulary depth and IELTS test scores (Akbariam & Alavi, 2013; Milton et al., 2010).

Relationship between vocabulary breadth and depth and different types of reading comprehension tasks. Only a limited number of studies have investigated the relationship between vocabulary knowledge and different types of reading comprehension tasks (Alavi & Akbarian, 2012; Li & Kirby, 2015; Zhang & Anual, 2008). With the exception of the study conducted by Li and Kirby (2015), who investigated the role of vocabulary *depth* across different reading question types, previous research has only examined the role of vocabulary *breadth* in different reading questions. Despite the varying results of correlations between vocabulary breadth or depth and different question types, the majority of the findings have reached similar conclusions: vocabulary breadth correlates more highly with questions that require only explicit information comprehension, such as understanding detailed information; for questions that require the students to process implicit information (for example, to infer or to summarise), vocabulary breadth does not have a significant impact. For instance, Zhang and Anual (2008) tested 37 Year 4 secondary students in Singapore and reported that vocabulary breadth correlated significantly with short-answer questions but not with summary questions. Alavi and Akbarian (2012) found that vocabulary size—indicated by the Vocabulary Levels Test (Schmitt et al., 2001)—has a higher Pearson's correlation coefficient (.443) with performance in Stated Detail questions than with that in other question types, including Main Idea (.208), Inference (.241), and Reference (.240). Li and Kirby (2015) took a step further to probe into the relationship between vocabulary breadth and depth and two reading comprehension question types: multiple-choice questions and summary writing. The

results indicated that vocabulary breadth significantly predicted multiple-choice performance, while vocabulary depth contributed to summary writing.

Predicting reading comprehension using vocabulary breadth and depth. Most previous studies in this area have investigated how vocabulary breadth and depth predicted learners' reading comprehension based on multiple regression analysis and looked at the percentage of variance that vocabulary breadth and depth accounted for in reading comprehension (Alavi & Akbarian, 2012; Li & Kirby, 2015; Qian, 2002). Qian (2002) found that vocabulary breadth and depth accounted for 59% and 54% of the variance in TOEFL reading scores, respectively. Alavi and Akbarian (2012) reported that vocabulary size represented 33% of the variance in Guessing Vocabulary, one of the question types in the TOEFL reading test.

Among the few studies that have examined how the increase in vocabulary breadth can predict the improvement in reading comprehension, studies that investigated lexical coverage (Hsueh-Chao & Nation, 2000; Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010) found certain vocabulary size thresholds. Lexical coverage refers to how much vocabulary breadth is needed for reading comprehension (Nation, 2013). An increase in vocabulary size below or above the threshold would result in different levels of increase in reading comprehension. However, these studies only looked at learners' vocabulary size and their reading comprehension scores; the variable of vocabulary depth was not considered. Another limitation in these studies is that linear multiple regressions were conducted to predict how an increase in vocabulary size levels can result in an increase in reading comprehension scores (Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010). However, due to the non-linearity of the relationship between vocabulary size and reading comprehension in these two studies, the linear multiple regression analysis would not give an accurate prediction of the reading comprehension scores. Therefore, to address this issue, this study conducted a different type of multiple regression by coding different vocabulary size levels as different dummy variables to capture the non-linearity. Additionally, vocabulary depth was also entered into the multiple regression model as a control variable to ensure that the prediction of reading comprehension by vocabulary size would not be affected by the variable of vocabulary depth.

Research Questions

This study aims to answer the following research questions based on the literature discussed above:

1. Are vocabulary breadth and vocabulary depth related to performance on the IELTS academic reading test?
2. Are vocabulary breadth and vocabulary depth related to performance on different question types in the IELTS academic reading test?
3. How does an increase in vocabulary breadth and depth affect IELTS academic reading test scores?

Method

Participants

This study adopted convenience sampling and the participants were 62 Chinese English-as-a-foreign-language (EFL) students aged between 16 and 20. At the time when the study was conducted, they were attending IELTS preparation courses in a language training school in a metropolitan city in China. They had learnt English for 8–10 years and followed a national English syllabus enforced by the Ministry of Education in China. The participants consisted of 24 students from the preliminary class, 25 students from the intermediate class, and 13 VIP students[1]. All the participants had taken IELTS courses for 80 hours, so they were reasonably familiar with the test format. Based on their IELTS mock test results, the students' English proficiency varied from A2 to C1 with reference to the Common European Framework of Reference for Languages (CEFR). These three groups of students with different proficiency levels were chosen to ensure normal distribution of the data collected.

Testing instruments

Based on the conceptual framework discussed above, this study used three vocabulary tests to measure participants' vocabulary breadth and depth: the Vocabulary Size Test, the Vocabulary Knowledge Scale, and the Word Associates Format. The first of these tests measures vocabulary breadth, and the other two measure vocabulary depth (see Figure 2).

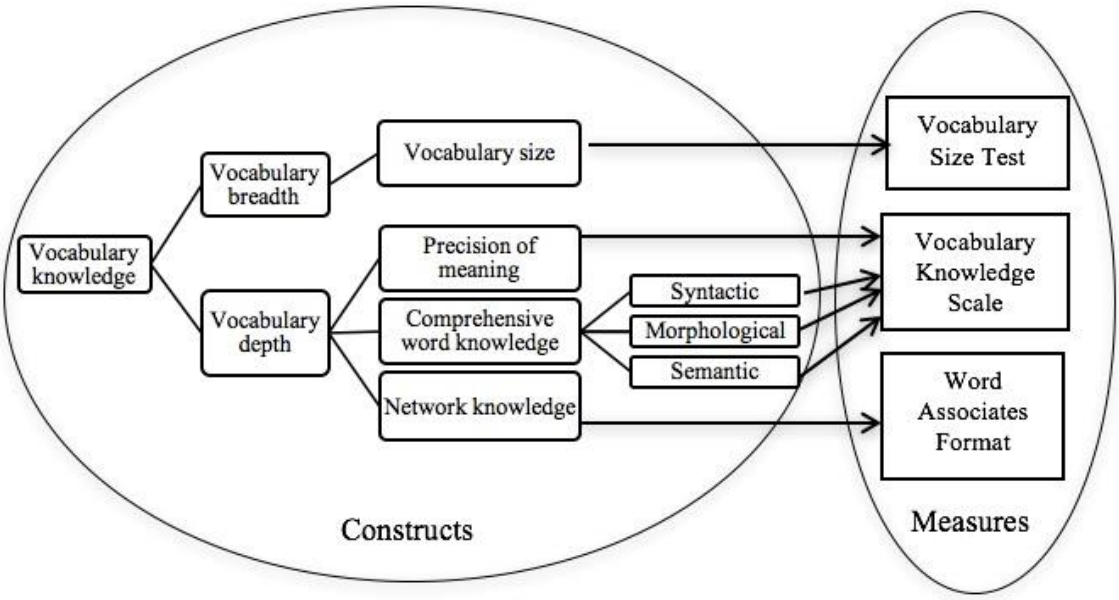


Figure 2. Vocabulary knowledge tests with reference to the conceptual framework

Vocabulary Size Test. The vocabulary size test used in this study is a bilingual version of Nation's Vocabulary Size Test (VST) (Nation & Beglar, 2007). The monolingual version of the VST has been empirically validated by Beglar (2010). Previous studies that explored bilingual versions of the VST—for example, the Persian bilingual version developed by Karami (2012) and the Vietnamese bilingual version by Nguyen and Nation (2011)—found

[1] VIP students are students who pay higher tuition fees to study in small classes (normally 4-5 students) and to be provided with extra support from teachers.

that bilingual vocabulary size tests were more efficient. Read (2000) also suggested bilingual VST versions for learners with a common first language or learners with lower proficiency levels. This study used only the first 10 of the total 14 frequency levels in the original VST, because based on the participants' learning background, they had not advanced beyond a vocabulary size of 10,000 words, while beyond this level guesswork will affect the validity of the result. The rationale for doing this is also supported by Nation (2007), who stated that it is not necessary for elementary or intermediate learners to answer questions for all fourteen frequency levels. The following is an extract from the VST:

see: They saw it.
a. 切 *qie* (cut)
b. 等待 *dengdai* (waited)
c. 看 *kan* (saw)
d. 开始 *kaishi* (started)

Vocabulary Knowledge Scale. In the Vocabulary Knowledge Scale (VKS), participants were required to choose one category from a list of five to measure how well they knew the word. An example of the VKS elicitation scale for the word *edit* can be seen below (Paribakht & Wesche, 1997, p. 180).

Edit	
I	I don't remember having seen this word before.
II	I have seen this word before, but I don't know what it means.
III	I have seen this word before, and I think it means _____. (synonym or translation)
IV	I know this word. It means _____. (synonym or translation)
V	I can use this word in a sentence: _____. (Write a sentence.)
	(If you do this section, please also do Section IV.)

As can be seen from this example, the VKS measures *precision of meaning* in the conceptual framework, because learners will receive different scores for different stages of learning a word. The VKS can also capture some aspects of *comprehensive word knowledge*. Category III taps learners' morphological knowledge, as they can guess the meaning of the word based on roots, prefixes, and suffixes. Category V involves learners' syntactic word knowledge because it is necessary to make a sentence. Semantic knowledge is assessed by all the categories as it overlaps with all the other components—precision of meaning, morphological, syntactic, and network word knowledge. The VKS gives an overall integrated score of different components of *comprehensive word knowledge*, although it does not measure each individual component directly. Stewart, Batty, and Bovee (2012) conducted an empirical study based on this measure and obtained a reliability of .90 with person separation at 2.92, and a reliability of .99 with item separation at 10.67, by using the Rasch Partial Credit Model (Masters, 1982).

As the VKS does not provide specific word items, researchers need to choose target words in accordance with their research needs. In this study, 25 target words from the IELTS reading passages were selected (see Appendix A) based on frequency levels by using the online software *VocabProfile* (Cobb, 2015; see Appendix B for the scoring system of the VKS).

Word Associates Format. In the Word Associates Format (WAF; Read, 1998), learners were required to select 4 appropriate word associates among 8 different words. There are 40

sample words in the test, all of which are high-frequency words selected from the University Word List (Zhang & Koda, 2017). Two examples of the test format are given below:

sudden

beautiful	quick	surprising	thirsty	change	doctor	noise	school

common

complete	light	ordinary	share	boundary	circle	name	party

The words in the left box are synonyms of the target word, and those on the right can form collocations with the stimulus word. This test measures the construct of network knowledge in the conceptual framework of this study. We shortened the original version of the test from 40 questions to 25 questions (see Appendix C) and investigated the reliability of the shortened test by calculating the Cronbach's Alpha coefficient. The result (.921) shows that our revised version of the WAF is a reliable measure.

IELTS reading test. Two passages from authentic IELTS reading tests were chosen: one from Passage 1, Test 1 in Cambridge IELTS 9 (Cambridge ESOL, 2013) and the other from Passage 3, Test 2 in Cambridge IELTS 8 (Cambridge ESOL, 2011). Each passage has 2,150–2,750 words, and there are 27 questions in total. These two passages were chosen because (a) the level of difficulty matched the authentic IELTS reading test, and (b) the two passages both included the question types that this study aimed to investigate, namely True/False/Not Given, Short Answer Questions, Matching Headings, Multiple Choice, and Sentence Completion. The lexical profile of the two passages, created by using the online concordance software *VocabProfile* (Cobb, 2015), is presented in Figure 3.

Figure 3. Lexical profile of the reading passages in the IELTS reading test

Procedure

Before the main study, five students from a language training school were involved in a pilot study. Based on their feedback, the WAF was shortened as mentioned above, and the order of the four tests were adjusted to ensure that the students would continue answering the questions and finish the complete set. In the main study, the test was administered during the

students' practice sessions in their IELTS course. They were given 90 minutes to answer all the questions. The sequence of the whole test was as follows: VKS (20 mins), IELTS reading test (40 mins), WAF and VST (30 mins).

Data analysis

The data collected were a set of quantitative statistics including the scores of the Vocabulary Size Test (VS), the Vocabulary Depth Score (VD), the Word Associates Format (WAF), and the Vocabulary Knowledge Scale (VKS). In addition, the data also consist of the overall scores of the IELTS reading test and scores from the five different types of IELTS questions including True/False/Not Given, Short Answer Questions, Matching Headings, Multiple Choice, and Sentence Completion. Since WAF and VKS were used to assess different aspects of vocabulary depth, and the scores for both were interval data, the two scores were added up jointly and adjusted to 100 points. All these 10 test scores were interval data and were entered into IBM SPSS (version 23) for descriptive and inferential analysis.

The normality of all the variables was checked by the Kolmogorov-Smirnov test and the Shapiro-Wilk test in conjunction with visual inspections of the histograms (see Appendix D). It was found that the scores of the Vocabulary Size Test were normally distributed, while the distributions of the other variables were not. Therefore, a Spearman's correlation coefficient was used as the normality assumption was violated. The variables and the correlations that this study sought to investigate can be seen in Appendix E.

Regarding the sample size of the multiple regression, many 'rules of thumb' seem to suggest different minimum sample sizes. Field (2009) summarized the required sample size to achieve high levels of predicting power based on the findings by Miles and Shevlin (2001) (See Appendix F). Since the correlation coefficients between VS, VD and IELTS are above .50 (see Table 2), the expected effect falls within the 'large effect' category according to Field (2009, p.57). As the number of predictors (independent variables) is seven in total (see Figure 5), a minimum sample size of less than 60 is required according to the figure in Appendix F. Thus, the sample size of 62 in this study has met the minimum sample size requirement for multiple regression.

Lastly, in order to check whether a linear multiple regression analysis could be conducted, the linearity of the relationship between VS and IELTS was investigated by visually examining the scatterplot of the two variables (Figure 4). It can be seen in the figure that the relationship was not linear. Therefore, it was decided that a multiple regression with dummy variables would be conducted.

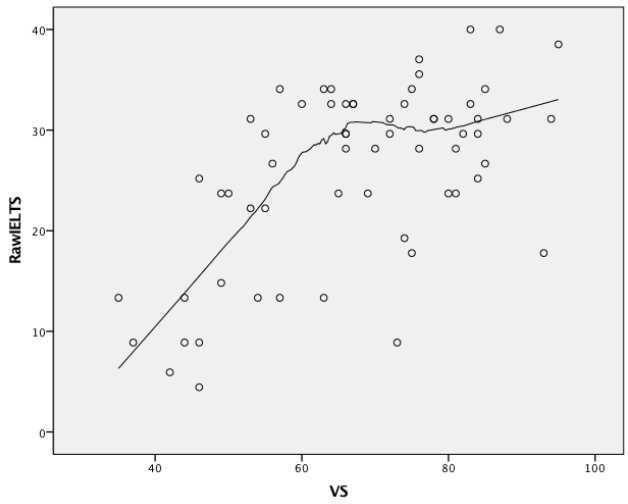


Figure 4. Scatterplot of VS and IELTS

Results

Descriptive statistics

Table 1 shows the means and standard deviations of participants' results in the Vocabulary Size Test (VS), Vocabulary Depth Score (VD), Word Associates Format (WAF), Vocabulary Knowledge Scale (VKS), IELTS reading test scores (IELTS), and different question types in the IELTS reading test including True/False/Not Given (TF), Short Answer Questions (SAQ), Matching Headings (MH), Multiple Choice (MC), and Sentence Completion (SC).

Table 1**:** *Descriptive statistics for all measures*

	Minimum score	Maximum score	Maximum possible score	Mean	Std. Deviation
VS	35	95	100	67.61	15.12
VD	45.33	92.44	100	68.02	13.20
WAF	43	92	100	73.32	12.37
VKS	48	123	125	79.73	20.49
IELTS	3	27	27	17.44	6.05
TF	0	7	7	4.56	1.68
SAQ	0	6	6	4.55	1.68
MH	0	6	6	3.98	1.94
MC	0	4	4	2.56	1.42
SC	0	4	4	1.74	1.37

Table 2: *Correlations between vocabulary knowledge and IELTS test scores*

Spearman's Correlations				
	VS	VD	WAF	VKS
IELTS	.51***	.61***	.61***	.57***

***$p < .001$ (two-tailed).

Correlations between vocabulary knowledge and overall IELTS scores

Table 2 shows the correlations between the scores of VS, VD, WAF, VKS, and IELTS.

The IELTS reading test scores were found to be significantly correlated with VS, r_{xy} = .51 (p < .001), and with VD, r_{yz} = .61 (p < .001). The scores of WAF and VKS also correlated significantly with IELTS reading test scores, r = .61 (p < .001) and r = .57 respectively (p < .001).

In order to statistically compare whether the relationship between VD (z) and IELTS reading test score (y) is stronger than that between VS (x) and IELTS test score, t-statistics were calculated based on the following equation (Chen & Popovich, 2002):

$$t_{\text{Difference}} = (r_{xy} - r_{zy})\sqrt{\frac{(n-3)(1+r_{xz})}{2(1-r_{xy}^2-r_{xz}^2-r_{zy}^2+2r_{xy}r_{xz}r_{zy})}}$$

$$= (0.51 - 0.61)\sqrt{\frac{(62-3)(1+0.41)}{2(1-0.51^2-0.41^2-0.61^2+2\times0.51\times0.41\times0.61)}} = -9.5633$$

After checking the result against the appropriate critical value in the table (Field, 2009, p. 803), it indicated that, compared with vocabulary breadth, vocabulary depth had a significantly higher correlation with the IELTS reading test scores.

Correlations between vocabulary knowledge and different question types

Table 3 presents correlations between vocabulary knowledge and different question types. Results showed that vocabulary size correlated significantly with True/False/Not Given at .42 (p < 0.001), while vocabulary depth did not have any correlation with this type of question. The scores for Short Answer Questions, Matching Headings, and Multiple Choice correlated more closely with vocabulary depth scores at .50, .62, .53 (all ps < 0.001), respectively, than with vocabulary size scores at .41 (p = 0.01), .49 (p < 0.001), .35 (p = 0.006), respectively. The scores for Sentence Completion had similar correlations with vocabulary size and vocabulary depth at .44 (p < 0.001) and .46 (p < 0.001), respectively. It seems clear that vocabulary breadth correlated with question types that required explicit information processing such as True/False/Not Given, while vocabulary depth correlated with questions that required implicit information processing such as Matching Headings (where the learners must summarize main ideas) and Multiple Choice (where the learners had to infer meaning from the passage or 'read between the lines'). Another interesting finding was that vocabulary depth correlated more highly with those questions that required the learners to provide answers with words or phrases such as Short Answer Questions and Sentence Completion.

Table 3: *Spearman's correlations between vocabulary knowledge and different IELTS question types*

	VS	VD
True/False/Not Given	.42***	.15
Short Answer Questions	.41**	.50***
Matching Headings	.49***	.62***
Multiple Choice	.35**	.53***
Sentence Completion	.44***	.46***

*p < .05, **p < .01, ***p < .001 (two-tailed).

Predicting IELTS reading test scores using vocabulary breadth and depth

To answer the third research question, multiple regression with dummy variables was performed after having checked the assumptions of the sample size requirement and linearity in the *data analysis* section.

First, the samples were divided into groups with different vocabulary-size levels: the 3,000–3,999 vocabulary-size group, the 4,000–4,999 vocabulary-size group, and so forth, up to the 9,000–10,000 vocabulary-size group. In this way, 7 dummy variables, namely VS30, VS40, VS50, VS60, VS70, VS80, and VS90, were created.

Second, for the first dummy variable, VS30, the value 1 was assigned to the 3,000–3,999 group, and the value 0 was assigned to all other groups. For the second dummy variable, VS40, the value 1 was assigned to the 4,000–4,999 group and, for all other groups, 0 was assigned. The same process was repeated for the rest of the dummy variables (see Table 4).

Table 4: *Dummy coding*

Dummy variable / Vocabulary Size	VS30	VS40	VS50	VS60	VS70	VS80	VS90
3,000–3,999	1	0	0	0	0	0	0
4,000–4,999	0	1	0	0	0	0	0
5,000–5,999	0	0	1	0	0	0	0
6,000–6,999	0	0	0	1	0	0	0
7,000–7,999	0	0	0	0	1	0	0
8,000–8,999	0	0	0	0	0	1	0
9,000–10,000	0	0	0	0	0	0	1

Finally, multiple regression analysis was conducted with IELTS scores as the dependent variable, and VD and VS30, VS50, VS60, VS70, VS80 and VS90 as independent variables (see Figure 5). VS40 was not added because it was left out as a default group that the others could be compared against. VS40 was chosen to be the default group because the lowest vocabulary level group, VS30, contained only 2 samples, which was too small to make any valid comparison.

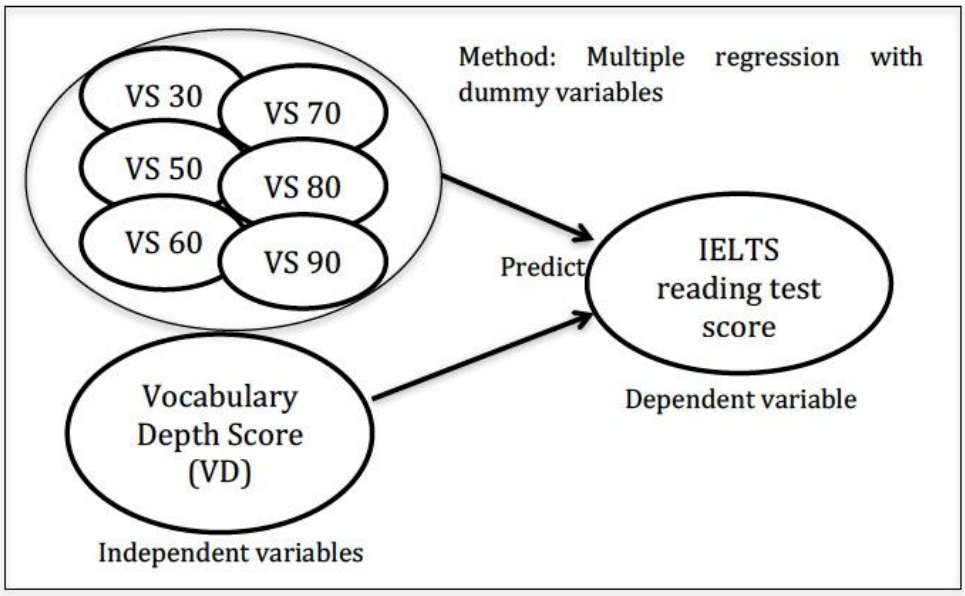


Figure 5. Multiple regression with dummy variables

Table 5: *Multiple regression coefficients*[a]

Model		Unstandardised Coefficients		Standardised Coefficients		
		B	Std. Error	Beta	*t*	Sig.
1	(Constant)	-3.181	4.615		-.689	.494
	VD	.297	.074	.437	4.004	.000
	VS30	-1.492	4.831	-.030	-.309	.759
	VS50	9.045	3.004	.358	3.011	.004
	VS60	11.302	2.988	.517	3.783	.000
	VS70	7.937	3.248	.363	2.444	.018
	VS80	12.665	2.947	.595	4.297	.000
	VS90	13.164	4.195	.318	3.138	.003

a. Dependent variable: IELTS scores

The results of the multiple regression model showed that the adjusted *R* square was .536, which indicated that the variables accounted for 53.6% of IELTS reading test scores (F = 11.067, $p < .001$). Table 5 displays the coefficients of the multiple regression model. The unstandardised coefficient of VS50, $B = 9.045$, indicates that if candidates increase their vocabulary size from the 4,000–4,999 level to the 5,000–5,999 level, they can increase their IELTS raw scores by approximately 9 points, holding the VD constant. Similarly, if a learner increases his or her vocabulary size from the 4,000–4,999 level to the 6,000–6,999 level, the IELTS raw scores will rise by about 11.3 points. This means that with a vocabulary size increase from the 5,000–5,999 level to the 6,000–6,999 level, a test-taker can only increase the raw scores by 2.3 marks (11.3 - 9 = 2.3). It is also interesting to note that when the participants reached the vocabulary-size level of 7,000–7,999, their IELTS reading scores decreased in comparison with the 6,000–6,999 level. This phenomenon may imply a bottleneck stage where learners stop improving their IELTS reading scores, even though they are acquiring more new words. When they reached the vocabulary size of 8,000 words, their IELTS reading scores started to increase again with the development of vocabulary size, but it grew more slightly than the 4,000–6,000 level. This suggests that, after the bottleneck stage,

the increase in vocabulary size can continue helping learners improve their IELTS reading test scores, yet the improvement is relatively slow compared with that at the early stages.

Discussion

Correlations between vocabulary breadth and depth and overall IELTS reading test scores

As regards the first research question, this study found that both vocabulary breadth and vocabulary depth were significantly correlated with IELTS reading performance. This finding is consistent with previous studies that investigated the role of vocabulary breadth in reading comprehension (Akbariam & Alavi, 2013; Alavi & Akbarian, 2012; Chiang, 2018; Laufer, 1992; Milton et al., 2010; Zhang & Anual, 2008) as well as the relationship between vocabulary depth and reading comprehension (Li & Kirby, 2015; Qian, 1999, 2002).

As for the relationship between vocabulary depth and reading comprehension, although the large[2] correlation of .61 in this study was similar to the few studies that took vocabulary depth into consideration, the construct of vocabulary depth was conceptualised differently and was assessed by diverse instruments. Among the three aspects of vocabulary depth in the conceptual framework suggested in this study—including precision of meaning, word comprehensive knowledge, and network knowledge—Qian (1999, 2000) only looked at the third aspect using a revised version of the Word Associates Format and found it correlated with TOEFL test scores at .82 and .77 in his two studies, respectively. Li and Kirby (2015) measured all three constructs with word definition, morphological awareness, and multiple-meaning tests and reported correlations with reading comprehension at .18, .22, and .45, respectively. The present study has broadened the scope of the construct of vocabulary depth and measured all three aspects—precision of meaning, comprehensive word knowledge (including semantic, morphological, syntactic word knowledge), and network knowledge—of vocabulary depth by using well-established vocabulary tests; it found that the IELTS reading test scores correlated at .57 with precision of meaning and comprehensive word knowledge and at .61 with network knowledge. This significant correlation provides strong evidence showing that in addition to network knowledge explored in Qian's (1999, 2000) studies and precision of meaning, morphological awareness, and multiplicity explored in Li and Kirby's (2015) study, all three aspects of vocabulary depth knowledge assessed in this study play important roles in reading comprehension test performance.

Correlations between vocabulary breadth and depth and different question types

The result for the second research question showed that for question types that required explicit information processing such as True/False/Not Given, the scores had a significant correlation with vocabulary breadth but had no correlation with vocabulary depth. However, questions that generally required implicit information processing in IELTS academic reading tests, such as Matching Headings and Multiple Choice, correlated more highly with vocabulary depth than with vocabulary breadth. This was consistent with the findings reported by Li and Kirby (2015), who found that vocabulary breadth significantly correlated with multiple-choice performance, which require explicit information processing in their study; while vocabulary depth was more correlated with summary writing, as this requires

[2] We have followed Plonsky and Oswald's (2014) proposed interpretation of effect sizes: *r*s close to .25 should be considered small, .40 medium, and .60 large.

implicit information comprehension. The finding in this research also corroborated the study by Alavi and Akbarian (2012), who found that performance in Stated Detail questions had a higher correlation (.443) with vocabulary breadth than that in other question types including Main Idea (.208), Inference (.241), and Reference (.240).

Vocabulary breadth correlated with True/False/Not Given in the IELTS test because these questions involved explicit information comprehension without inferring or summarising, so test-takers could simply rely on their vocabulary breadth knowledge as they were only dealing with the surface meaning of the text. Multiple Choice, Matching Headings, and Sentence Completion were more related to vocabulary depth because they involved implicit information comprehension. Many of the Multiple Choice items asked questions about opinions and attitudes, which required readers to infer from the explicit information in the passage. For Matching Headings, however, readers must summarise the implicit main ideas of each paragraph by inferring from the explicit information. For these two types of questions, network word knowledge and precision of meaning both played a pivotal role because readers needed to link all the words together in an accurate manner so that they could infer a range of opinions and attitudes or extract main ideas from the passage.

Predicting IELTS reading test scores using vocabulary breadth and depth

To answer the last research question, this study conducted multiple regression analysis with dummy variables to examine how the increase in vocabulary breadth could predict the increase in IELTS reading test scores. The result of the multiple regression showed that once the vocabulary size reached a certain level (the threshold level of 6,000 vocabulary size in this study), a larger vocabulary size seemed to have limited effects on the improvement in reading comprehension. This threshold level of 6,000 vocabulary size concurred with Laufer and Ravenhorst-Kalovski's finding (2010); this is perhaps because beyond this level, other factors, such as test-taking strategies, grammar, and background knowledge, could become more important than vocabulary knowledge alone. This also partially aligned with Qian's finding (1999) that if learners' vocabulary size extended beyond the threshold level, vocabulary depth could make unique contributions to the prediction of reading comprehension in addition to vocabulary size.

Conclusion

This study has provided some insights into the relationship between vocabulary knowledge and learners' performance in IELTS academic reading tests. It has provided a clear and comprehensive conceptualisation of vocabulary depth knowledge based on the merits of previous conceptualisations and measured the construct with a combination of two empirically validated assessment tools. Correlation analysis showed that vocabulary depth is more important than breadth in terms of improving IELTS academic reading scores. This indicates that the development of vocabulary depth knowledge, which may have been neglected by many Chinese students and teachers, should be given more attention. Regression analysis revealed a vocabulary-size threshold level of 6,000 word families. Therefore, in order to improve students' reading comprehension, teachers are advised to help learners to develop their vocabulary breadth knowledge until they have reached the vocabulary size of 6,000 word families. Once this level has been attained, teachers should then focus more on developing learners' other skills such as test-taking strategies and grammar.

Some limitations in this study need to be noted. Firstly, although the sample size reaches the minimum required sample size for statistical analysis with multiple regression analysis, there may still be some bias if it is to be generalized across the population. If a larger sample size could be obtained, the models would be more robust. Secondly, this study only included vocabulary breadth and depth in the multiple regression analysis, and many variables—such as grammar, background knowledge, and first language reading strategies—were not measured and added into the regression model as control variables. In future research, the above issues need to be taken into consideration in data analysis.

References

Akbariam, I., & Alavi, S. M. (2013). Comparing the contribution of vocabulary breadth to IELTS and TOEFL reading subtests. *Porta Linguarum: Revista Internacional de Didáctica de Las Lenguas Extranjeras*, *20*, 135–151.

Alavi, S. M., & Akbarian, I. (2012). The role of vocabulary size in predicting performance on TOEFL reading item types. *System*, *40*, 376–385. doi:10.1016/j.system.2012.07.002

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the acquisition of word knowledge. In B. Hutson (Ed.), *Advances in reading/language research* (Vol. 2, pp. 231−256). Greenwich, CT: JAI Press.

Bardakci, M. (2016). Breadth and depth of vocabulary knowledge and their effects on L2 vocabulary profiles. *English Language Teaching*, *9*, 239–250. doi: 10.5539/elt.v9n4p239

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, *27*, 101–118. doi:10.1177/0265532209340194

Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, *42*, 390–405. doi:10.1006/jmla.1999.2681

Cakir, A., Unaldi, I., Arslan, F. Y., & Kilic, M. (2016). Effects of reading strategies and depth of vocabulary knowledge on Turkish EFL learners' text inferencing skills. *English Language Teaching*, *9*, 11–20. doi:10.5539/elt.v9n8p11

Cambridge ESOL. (2011). *Cambridge IELTS 8 student's book with answers: Official examination papers from University of Cambridge ESOL Examinations* (Vol. 8). Cambridge: Cambridge University Press.

Cambridge ESOL. (2013). *Cambridge IELTS 9 student's book with answers: Official examination papers from University of Cambridge ESOL Examinations* (Vol. 9). Cambridge: Cambridge University Press.

Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage.

Chiang, H. H. (2018). English vocabulary size as a predictor of TOEIC listening and reading achievement among EFL students in Taiwan. *Theory and Practice in Language Studies*, *8*, 203–212. doi:10.17507/tpls.0802.04

Cobb, T. (2015). VocabProfile [Computer program]. Retrieved from https://www.lextutor.ca/vp/

Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *The Journal of Educational Research*, *36*, 206–217

Dóczi, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization*. Oxford: Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780190210274.001.0001/acprof-9780190210274

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: SAGE.

Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners* (Unpublished doctoral dissertation). Lunds University, Media-Tryck, Lund, Sweden.

Hsueh-Chao, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–430.

IELTS numbers rise to three million a year. (2018). Retrieved 14 October 2018, from https://www.ielts.org/news/2017/ielts-numbers-rise-to-three-million-a-year

Karakoç, D., & Köse, G. D. (2017). The impact of vocabulary knowledge on reading, writing and proficiency scores of EFL learners. *Journal of Language and Linguistic Studies*, *13*, 352–378.

Karami, H. (2012). The development and validation of a bilingual version of the Vocabulary Size Test. *RELC Journal*, *43*, 53–67. doi:10.1177/0033688212439359

Kremmel, B. (2015). *The more, the merrier? Issues in measuring vocabulary size*. Paper presented at the 2015 Language Testing Research Colloquium, Toronto.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). UK: Palgrave Macmillan. doi:10.1007/978-1-349-12396-4_12

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size. *Reading in a Foreign Language*, *22*, 15–30.

Li, M., & Kirby, J. R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics*, *36*, 611–634. doi:10.1093/applin/amu007

Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174. doi:10.1007/BF02296272

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, *19*, 741–760. doi: 10.1177/1362168814567889

Meara, P. (1992). Network structures and vocabulary acquisition in a foreign language. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 62–70). London: Palgrave Macmillan. doi:10.1007/978-1-349-12396-4_6

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35–53). Cambridge: Cambridge University Press.

Meara, P., & Fitzpatrick, T. (2000). Lex30: an improved method of assessing productive vocabulary in an L2. *System*, *28*, 19–30. doi:10.1016/S0346-251X(99)00058-5

Meara, P., & Jones, G. (1990). *Eurocentre's Vocabulary Size Test: User's guide*. Zurich: Eurocentres.

Miles, J. & Shevlin, M. (2001). *Applying regression & correlation: A guide for students and researchers*. London: SAGE.

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. del Mar Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters.

Mokhtari, K., & Thompson, H. B. (2006). How problems of reading fluency and comprehension are related to difficulties in syntactic awareness skills among fifth graders. *Reading Research and Instruction*, *46*, 73–94. doi: 10.1080/19388070609558461

Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, *1*(5), 12–25.

Nation, P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary in a second language* (pp. 33–43). Cambridge: Cambridge University Press.

Nation, P. (2013). *Learning vocabulary in another language* (2nd ed.). New York: Cambridge University Press.

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Nguyen, L. T. C., & Nation, P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal*, *42*, 86–99. doi:10.1177/0033688210390264

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning. In J. Coady & T. N. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174−200). Cambridge: Cambridge University Press.

Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices—Past, present, and future. In S. G. Paris, & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13−70). London: Routledge.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. doi: 10.1111/lang.12079

Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and Writing*, *25*, 1635–1664. doi:10.1007/s11145-011-9336-5

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in ESL reading comprehension. *The Canadian Modern Language Review*, *56*, 282–308.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*, 513–536. doi:10.1111/1467-9922.00193

Rashidi, N., & Khosravi, N. (2010). Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension of Iranian EFL learners. *Journal of Pan-Pacific Association of Applied Linguistics*, *14*, 81–108.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, *10*, 355–371. doi:10.1177/026553229301000308

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41–60). Mahwah, NJ: Lawrence Erlbaum.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209–228). Amsterdam: John Benjamins.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, *10*, 77–89. doi:10.2307/3585941

Schmitt, N. (1998). Quantifying word association responses: What is native-like? *System*, *26*, 389–401. doi:10.1016/S0346-251X(98)00019-0

Schmitt, N. (2010a). Key issues in teaching and learning vocabulary. In R. Chacón Beltrán, C. Abello-Contesse, & M. del Mar Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 28–40). Bristol: Multilingual Matters.

Schmitt, N. (2010b). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*, 913–951. doi:10.1111/lang.12077

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*, 55–88. doi:10.1191/026553201668475857

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, *36*, 139–152. doi:10.1080/09571730802389975

Stewart, J., Batty, A. O., & Bovee, N. (2012). Comparing multidimensional and continuum models of vocabulary acquisition: An empirical examination of the Vocabulary Knowledge Scale. *TESOL Quarterly*, *46*, 695–721. doi: 10.1002/tesq.35

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *International Journal of Applied Linguistics*, *168*, 33–69. doi: 10.1075/itl.168.1.02web

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *The Canadian Modern Language Review*, *53*, 13–40.

Zhang, D., & Koda, K. (2017). Assessing L2 vocabulary depth with word associates format tests: Issues, findings, and suggestions. *Asian-Pacific Journal of Second and Foreign Language Education*, *2*. doi:10.1186/s40862-017-0024-0

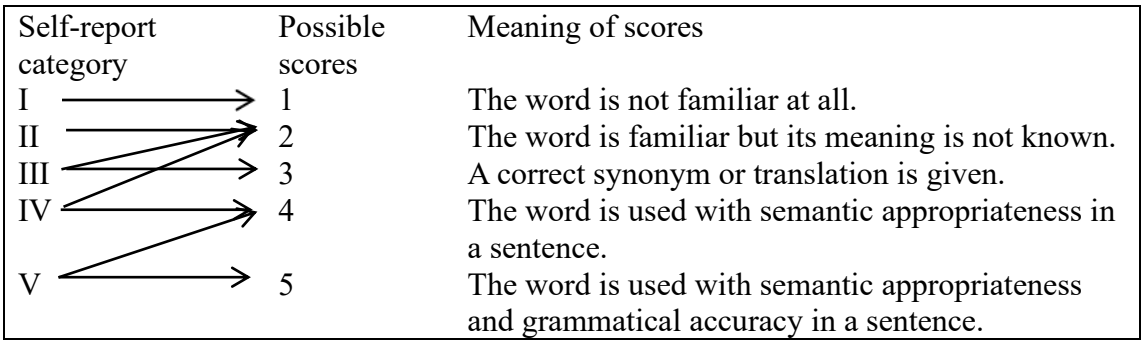
Zhang, L. J., & Anual, S. B. (2008). The role of vocabulary in reading comprehension: the case of secondary school students learning English in Singapore. *RELC Journal*, *39*, 51–76. doi:10.1177/0033688208091140

Appendix A Words selected for Vocabulary Knowledge Scale

1. investigate
2. recognise
3. demand
4. bacteria
5. component
6. enthusiasm
7. provide
8. invest
9. potential
10. survey
11. manufacture
12. perceive
13. psychology
14. substitute
15. artificial
16. odour
17. evoke
18. cue
19. synthetic
20. vaccine
21. intimate
22. elusive
23. chemist
24. immerse
25. sensor

Appendix B Vocabulary Knowledge Scale scoring system

Category 1 and 2 are scored with 1 and 2 points respectively. If the test-taker's answer for category 3 or 4 is mistaken, he or she can also receive 2 points; if the answer is correct, a score of 3 is given for category 3 and 4 for category 4. If an acceptable sentence is provided in category 5, a score of 5 is rewarded, if not the test-taker will receive 4 points.



Self-report category	Possible scores	Meaning of scores
I	1	The word is not familiar at all.
II	2	The word is familiar but its meaning is not known.
III	3	A correct synonym or translation is given.
IV	4	The word is used with semantic appropriateness in a sentence.
V	5	The word is used with semantic appropriateness and grammatical accuracy in a sentence.

The VKS Scoring Category (Paribakht & Wesche, 1997, p. 181)

Appendix C Word Associates Format

1. beautiful

A. enjoyable B. expensive C. free D. loud	E. education F. face G. music H. weather

2. bright

A. clever B. famous C. happy D. shining	E. colour F. hand G. poem H. taste

3. calm

A. open B. quiet C. smooth D. tired	E. cloth F. day G. light H. person

4. natural

A. expected B. helpful C. real D. short	E. foods F. neighbours G. parents H. songs

5. fresh

A. another B. cool C. easy D. raw	E. cotton F. heat G. language H. water

6. general

A. closed B. different C. usual D. whole	E. country F. idea G. reader H. street

7. common

A. complete B. light C. ordinary D. shared	E. boundary F. circle G. name H. party

8. complex

A. angry B. difficult C. necessary D. sudden	E. argument F. passengers G. patterns H. problem

9. broad

A. full B. moving C. quiet D. wide	E. night F. river G. shoulders H. smile

10. convenient

A. easy B. fresh C. near D. suitable	E. experience F. sound G. time H. vegetable

11.dense

A. crowded B. hot C. noisy D. thick	E. forest F. handle G. smoke H. weather

12.curious

A. helpful B. interested C. missing D. strange	E. accident F. child G. computer H. steel

13. direct

A. honest B. main C. straight D. wide	E. fence F. flight G. heat H. river

14. favorable

A. helpful B. legal C. possible D. positive	E. habit F. response G. teacher H. weather

15. secure

A. confident B. enjoyable C. fixed D. safe	E. game F. job G. meal H. visitor

16. tight

A. close B. rough C. uncomfortable D. wet	E. bend F. pants G. surface H. wood

17. violent

A. expected B. smelly C. strong D. unlucky	E. anger F. death G. rubbish H. storm

18. domestic

A. home B. national C. regular D. smooth	E. animal F. movement G. policy H. speed

19. formal

A. fast B. loud C. organised D. serious	E. bomb F. education G. growth H. statement

20. independent

A. changed B. equal C. important D. separate	E. child F. country G. ideas H. prices

21. sensitive

A. feeling B. interesting C. sharp D. thick	E. clothes F. instrument G. skin H. topic

22. professional

A. paid B. public C. regular D. religious	E. advice F. manner G. musician H. transport

23. critical

A. clear B. dangerous C. important D. rough	E. festival F. illness G. time H. water

24. liberal

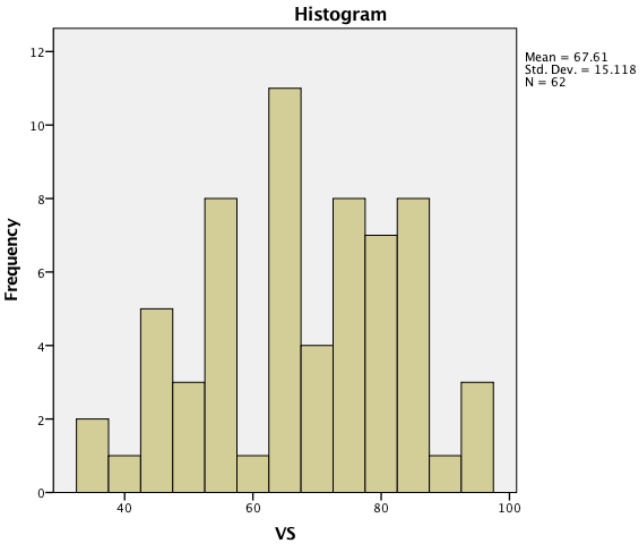
A. free B. moderate C. plenty D. valuable	E. crops F. furniture G. parents H. transport

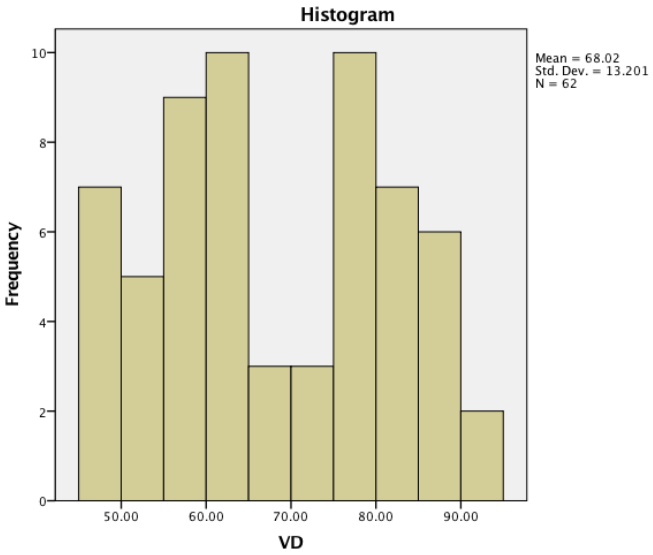
25. dramatic

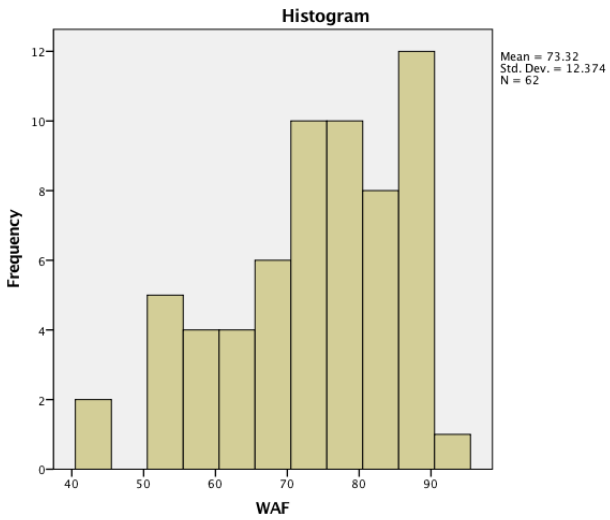
A. exciting B. official C. surprising D. worried	E. adventure F. change G. patient H. salary

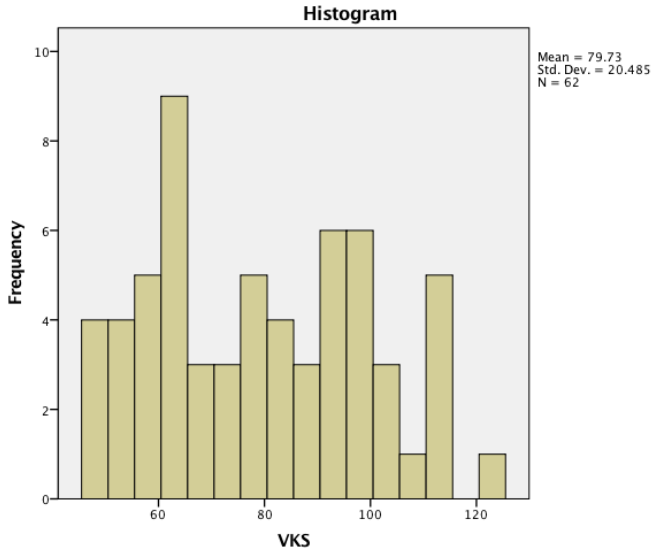
Appendix D Normality test results

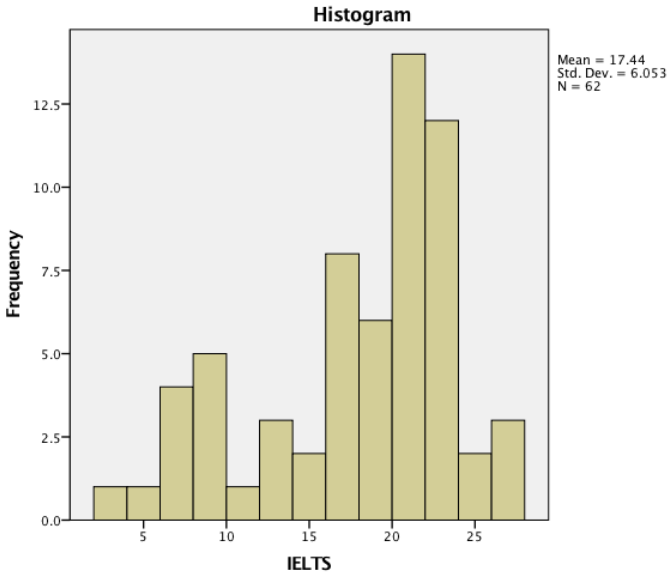
	Tests of Normality									
	Kolmogorov-Smirnov[a]		Shapiro-Wilk		Skewness			Kurtosis		
	Statistic	Sig.	Statistic	Sig.	Statistic	Std. Error	z-value	Statistic	Std. Error	z-value
VS	.083	.200*	.971	.152	-.262	.304	-0.86	-.810	.599	-1.35
VD	.119	.029	.946	.009	.105	.304	0.35	-1.239	.599	-2.07
WAF	.087	.200*	.953	.018	-.538	.304	-1.77	-.436	.599	-0.73
VKS	.119	.030	.953	.018	.206	.304	0.68	-1.088	.599	-1.82
IELTS	.166	.000	.915	.000	-.778	.304	-2.56	-.302	.599	-0.50
TF	.143	.003	.941	.005	-.440	.304	-1.45	-.204	.599	-0.34
SAQ	.235	.000	.814	.000	-1.231	.304	-4.05	.867	.599	1.45
MH	.232	.000	.857	.000	-.689	.304	-2.27	-.836	.599	-1.40
MC	.247	.000	.837	.000	-.418	.304	-1.38	-1.234	.599	-2.06
SC	.167	.000	.883	.000	.287	.304	0.94	-.990	.599	-1.65

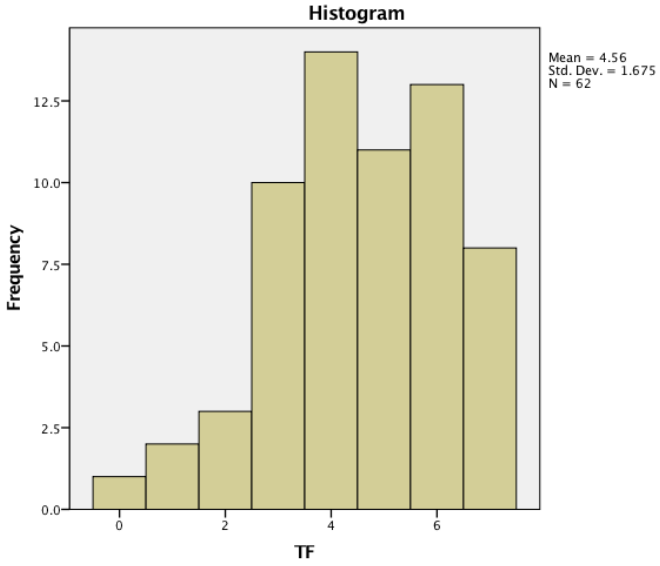


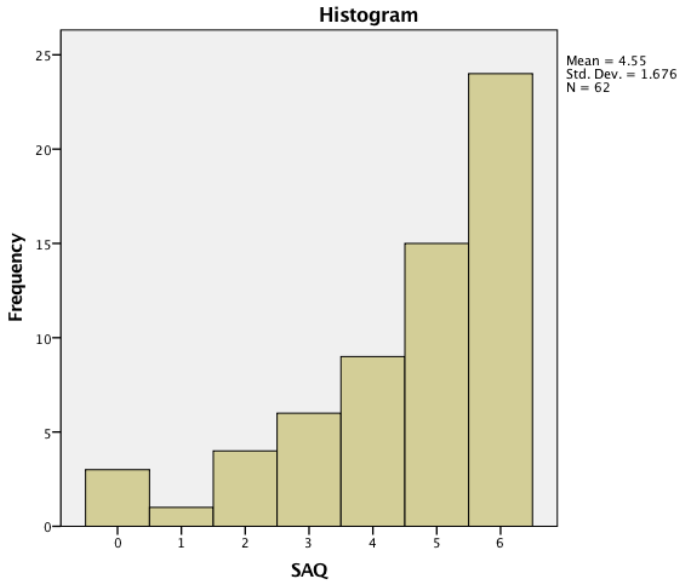


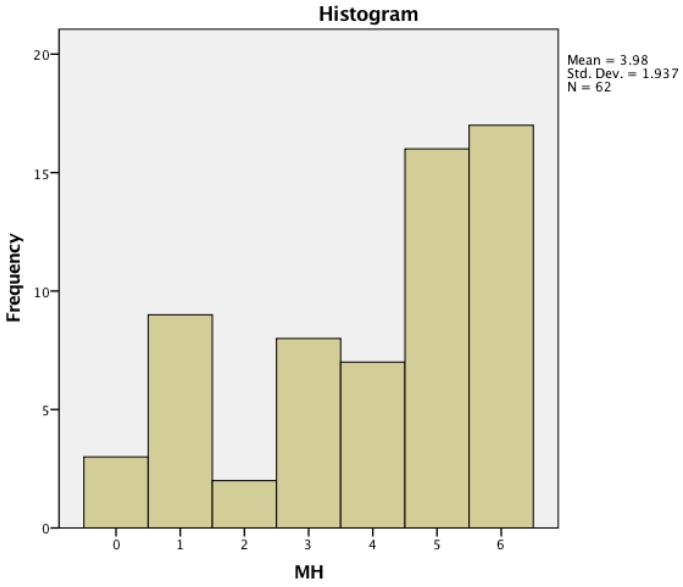


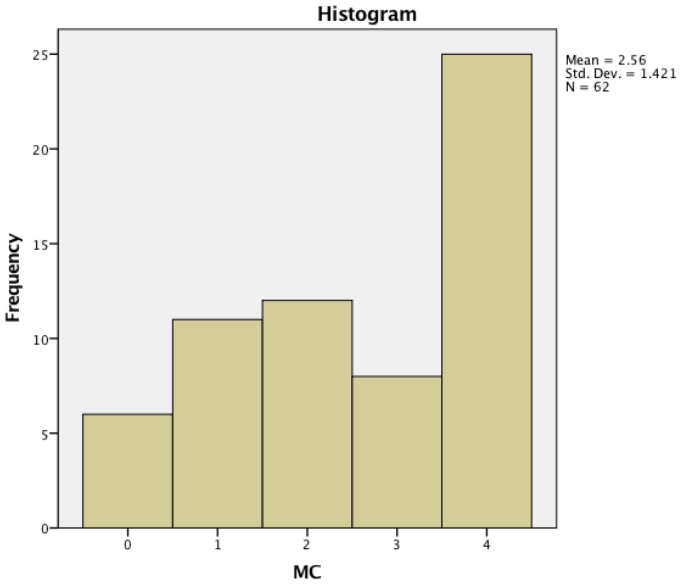


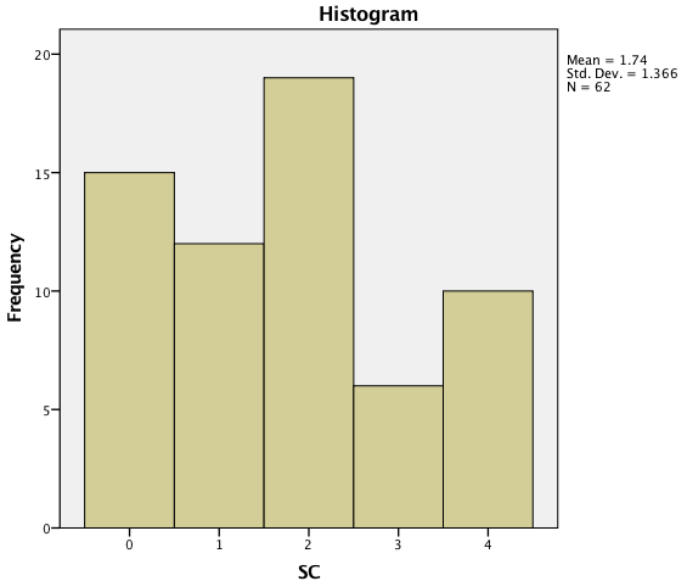

Histogram
Mean = 17.44
Std. Dev. = 6.053
N = 62
Frequency
IELTS


Histogram
Mean = 4.56
Std. Dev. = 1.675
N = 62
Frequency
TF

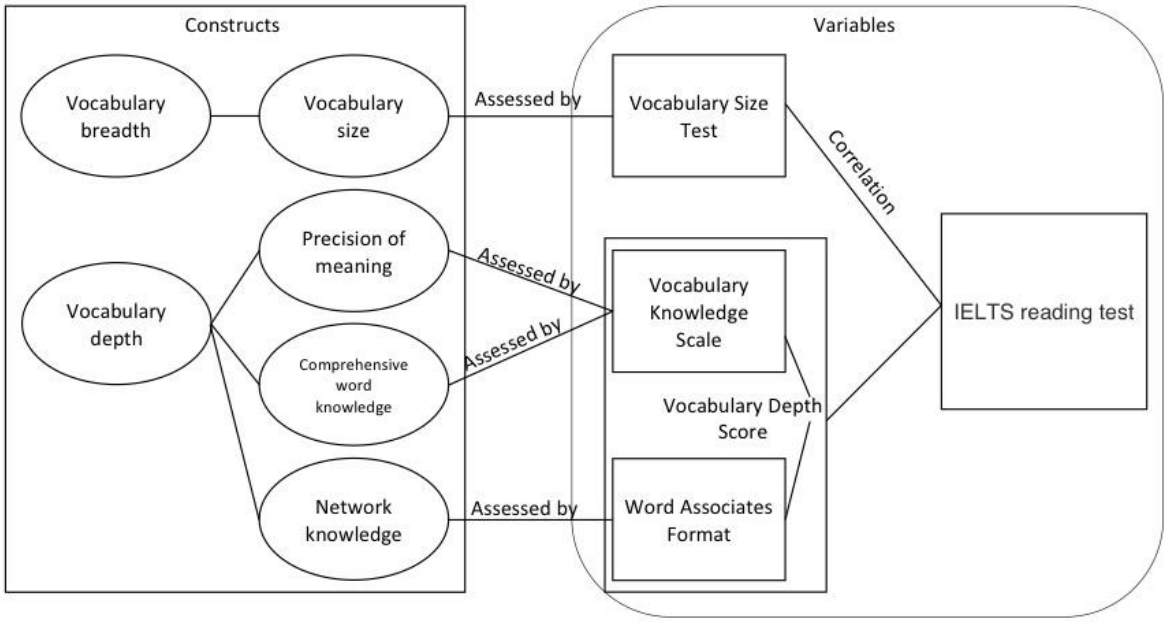

Histogram
Mean = 4.55
Std. Dev. = 1.676
N = 62
Frequency
SAQ


Histogram
Mean = 3.98
Std. Dev. = 1.937
N = 62
Frequency
MH

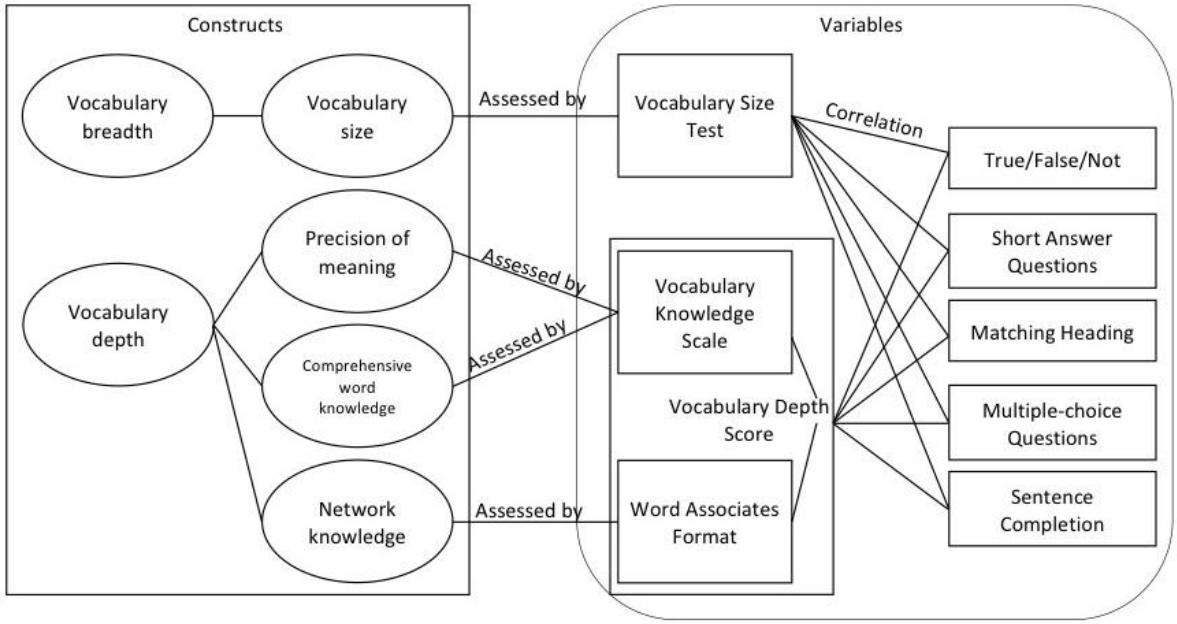

Histogram
Mean = 2.56
Std. Dev. = 1.421
N = 62
Frequency
MC


Histogram
Mean = 1.74
Std. Dev. = 1.366
N = 62
Frequency
SC

Appendix E Variables and correlations this study sought to investigate

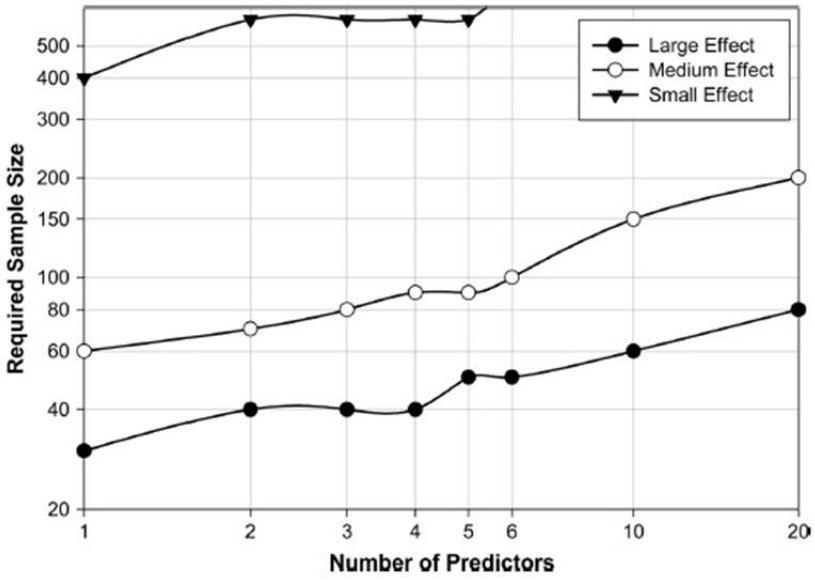


Constructs and variables for the correlation between vocabulary breadth and depth and IELTS reading test scores



Constructs and variables for the correlation between vocabulary breadth and depth and different types of IELTS reading questions

Appendix F Required sample size



Sample size required in multiple regression based on the number of predictors and size of expected effect (Field, 2009, p. 223)

About the authors

Chen Chen is an English Language Tutor at Xi'an Jiaotong-Liverpool University in China. She is also a part-time PhD student at the University of Exeter. Her research interests include L2 vocabulary, corpus linguistics, and English for Academic Purposes. Email: chen.chen@xjtlu.edu.cn

Yongcan Liu is Reader in Applied Linguistics and Languages Education at the Faculty of Education, University of Cambridge. His research interests lie in sociocultural theory of mind, school multilingualism, and heritage language acquisition and education. E-mail: yl258@cam.ac.uk