une University of New England

Library

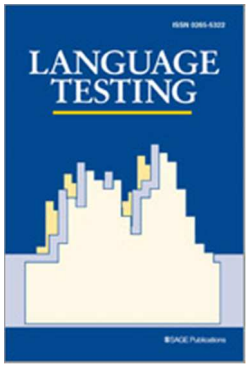
This is the pre-peer reviewed version of the following article:#

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

The final publication is available at Sage via

https://doi.org/10.1177/0265532216676851

Downloaded from e-publications@UNE the institutional research repository of the University of New England at Armidale, NSW Australia.



The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading

Journal:	*Language Testing*
Manuscript ID	LT-16-0023.R1
Manuscript Type:	Original Manuscript
Keywords:	Vocabulary knowledge, L2 listening, L2 reading, Vocabulary tests, Regression models
Abstract:	This study explores the constructs which underpin three different measures of vocabulary knowledge and investigates the degree to which these three measures correlate with, and are able to predict, measures of second language (L2) listening and reading. Word frequency structured vocabulary tests tapping receptive/orthographic (RecOrth) vocabulary knowledge, productive/orthographic (ProOrth) vocabulary knowledge and productive/phonological (ProPhon) vocabulary knowledge and tests measuring L2 listening and L2 reading were administered to 250 tertiary level Chinese learners of English as a foreign language (EFL). Results showed that ProPhon vocabulary knowledge correlated most strongly with L2 listening (r = .71) and ProOrth vocabulary knowledge correlated most strongly with L2 reading (r = .57). Factor analysis indicated that all subcomponents of the ProPhon vocabulary knowledge test loaded onto one factor and those of the RecOrth and ProOrth vocabulary knowledge tests loaded onto another. Regression modelling showed that ProPhon vocabulary knowledge explained 51% of the variance in L2 listening scores and that ProOrth vocabulary knowledge explained 33% of the variance in the L2 reading scores. Discussion addresses the varying importance of different dimensions of vocabulary knowledge in L2 listening and reading.

SCHOLARONE™
Manuscripts

The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading

Abstract

This study explores the constructs which underpin three different measures of vocabulary knowledge and investigates the degree to which these three measures correlate with, and are able to predict, measures of second language (L2) listening and reading. Word frequency structured vocabulary tests tapping *receptive/orthographic (RecOrth) vocabulary knowledge*, *productive/orthographic (ProOrth) vocabulary knowledge* and *productive/phonological (ProPhon) vocabulary knowledge* and tests measuring L2 listening and L2 reading were administered to 250 tertiary level Chinese learners of English as a foreign language (EFL). Results showed that ProPhon vocabulary knowledge correlated most strongly with L2 listening (r = .71) and ProOrth vocabulary knowledge correlated most strongly with L2 reading (r = .57). Factor analysis indicated that all subcomponents of the ProPhon vocabulary knowledge test loaded onto one factor and those of the RecOrth and ProOrth vocabulary knowledge tests loaded onto another. Regression modelling showed that ProPhon vocabulary knowledge explained 51% of the variance in L2 listening scores and that ProOrth vocabulary knowledge explained 33% of the variance in the L2 reading scores. Discussion addresses the varying importance of different dimensions of vocabulary knowledge in L2 listening and reading.

Keywords

Vocabulary knowledge, L2 listening, L2 reading, vocabulary tests, regression models

Introduction

Second language (L2) vocabulary knowledge is of fundamental importance to second language learners (Nation, 2001, 2006). The lexical level is the lowest level of representation at which stable links between form and meaning can be established (Hulstijn, 2002), and thus knowledge of the words in a given sample of speech or writing strongly facilitates comprehension. The strong link between a

For Peer Review

learner's vocabulary knowledge and the ability to successfully negotiate L2 reading and listening tasks (Matthews & Cheng 2015; Qian, 2002; Stæhr, 2008, 2009) highlights the importance of L2 learners having adequate levels of vocabulary knowledge in order to cope with the linguistic demands of these fundamental L2 skills (Nation 2001).

Although the general relationship between L2 vocabulary knowledge and the skills of L2 listening (Matthews & Cheng, 2015; Stæhr, 2009) and L2 reading (Li & Kirby, 2015; Qian, 2002) is relatively well established, the respective strength of association that different dimensions of vocabulary knowledge have with these skills is not as clearly understood. Despite the broad acknowledgement of the multidimensional nature of vocabulary knowledge (Henriksen, 1999; Nation 2001; Milton, 2009), previous studies aiming to explore the relationship between L2 vocabulary knowledge and L2 language skills have employed test instruments which tap a relatively constrained range of vocabulary knowledge dimensions. A majority of studies in this area have emphasised receptive and orthographic word knowledge, with only a few also investigating the link between productive and phonological dimensions of word knowledge and L2 listening and reading. Additionally, previous studies which have investigated the link between vocabulary knowledge and language skills have predominantly done so by either investigating the relationship between a single measure of vocabulary knowledge on multiple macro-skills (Stæhr, 2008) or by investigating multiple aspects of word knowledge on a single macro-skill (Li & Kirby, 2014; Matthews & Cheng, 2015; Qian, 2002). Few studies have investigated the relationship between multiple aspects of vocabulary knowledge and both L2 listening and reading among a single large cohort of L2 learners. As a consequence, a limited amount of empirical data exists which casts light on the specific nature of the constructs measured by various commonly used vocabulary tests and the relative strength of association these measures of vocabulary knowledge have on the language skills of L2 listening and reading. In order to more fully understand the practical implications of the relationship between vocabulary knowledge and L2 language skills, more research using multiple measures of vocabulary knowledge is required (Stæhr, 2009).

In this paper we build on the existing body of empirical knowledge around the relationship between different measures of L2 vocabulary knowledge and measures of L2 listening and reading.

http://mc.manuscriptcentral.com/LTJ

For Peer Review

To do so, we explore the constructs which underpin three vocabulary tests and investigate the degree to which these commonly used vocabulary test formats are differentially associated with and predictive of the measures of L2 listening and reading of a single cohort of L2 learners.

Multiple dimensions of word knowledge

In the earliest stages of learning a word, a language learner must be able to link the form of the word with the word's meaning. Although the ability to establish form-meaning links is a fundamental first step in gaining control over a particular word, the multidimensional nature of vocabulary knowledge determines that establishing this link is just part of the challenge faced by L2 learners. The multidimensionality of vocabulary knowledge, as overviewed below, contextualises the scale and specificity of this challenge.

Although a universally accepted framework for vocabulary knowledge as yet does not exist (Daller, Milton, Treffers-Daller, 2007; Henriksen, 1999; Meara, 1996), Nation (2001) presents the most authoritative taxonomy currently available (Milton, 2013; Schmitt, 2008). This taxonomy categorises word knowledge under knowledge of word form, knowledge of word meaning and knowledge of word use. Knowledge of word form is subcategorised further as knowledge of the spoken form, knowledge of the written form and the knowledge of word parts. Knowledge of word meaning is subcategorised into knowledge of form-meaning connection, concepts and referents and knowledge of word associations. Knowledge of word use is subcategorised as knowledge of grammatical functions, collocations and knowledge of constraints on word use. Each subcategory of knowledge is additionally described in terms of receptive and productive dimensions.

The number of dimensions of word knowledge outlined in Nation's taxonomy (2001) highlight the fact that word learning is an incremental process (Schmitt, 2008). Word knowledge begins with vague categorisation of words, and if word learning is successful, will proceed to become precise and multidimensional in nature (Henriksen, 1999). To this end, successful word learning must necessarily include, but also go beyond, the consolidation of form-meaning links. Expanding the depth or quality of word knowledge will enable a learner to apply their word knowledge in a broader variety of contexts and facilitate skilled target language use (Schmitt, 2008). For example, possessing

For Peer Review

receptive knowledge of a word may enable the recognition of that item and facilitate fundamentally important form-meaning links. However, without possession of productive knowledge of that word, its active recall and productive use may not be available to the language learner. Similarly, knowledge of a word in the orthographic form is fundamental to the ability to comprehend and produce written language. However, without a sufficient depth of the phonological knowledge of words already known in the orthographic form the ability to process language while listening may be strongly inhibited (Goh, 2000).

Another important dimension of vocabulary knowledge relates to the fluency with which a known word can be recognised and used while engaged in the target language (Meara, 1996; Milton, 2013). The attribute of fluency is linked to the degree to which vocabulary knowledge can be processed with automaticity (Segalowitz & Hulstijn, 2005). Automatic processes are those which can be undertaken with accuracy, effortlessness and speed, and require the expenditure of very few cognitive resources (Fraser, 2007). So-called vocabulary fluency (Daller, et al. 2007) is particularly important when engaged in language tasks which require rapid processing in real time. In some instances a language learner may possess knowledge of a given range of words present in a text, but may not be able to apply this knowledge in communicative contexts as this word knowledge cannot be accessed with sufficient fluency. For example, the speed at which known vocabulary can be processed is particularly impactful in relation to listening (Hulstijn, 2003). The intrinsic nature of spoken language determines that the linguistic data encoded within speech is not temporally stable and is only available for processing by the listener for a short period of time. Without sufficient levels of vocabulary fluency, the listener typically encounters difficulty in accurately extracting meaning from spoken language (Field, 2008).

Receptive and productive word knowledge

It is broadly accepted that there is a distinction between receptive and productive vocabulary knowledge (Fan, 2000; Laufer & Goldstein 2004; Nation, 2001; Webb, 2005, 2008). Central to this distinction is the assumption that the ability to recognize a word requires less depth of knowledge than does accessing and producing that word. Vocabulary knowledge is generally assumed to start as

primarily receptive in nature and only later be of sufficient quality to facilitate productive word use (Read, 2000). It is for this reason that receptive and productive vocabulary knowledge can be considered to exist along a continuum (Henriksen, 1999; Melka, 1997). Indeed, research consistently shows that a learner's receptive vocabulary is larger in size than the learner's productive vocabulary (Fan, 2000; Laufer, 1998; Webb, 2008). In the earliest phases of knowing a word, a learner may only possess sufficient depth of knowledge for example, to know that he or she has seen the word before or perhaps to be able to link the word's form with the corresponding meaning. However, it may not be until the learner has encountered that word on a number of additional occasions and in a variety of contexts that he or she is able to appropriately produce that word in spontaneous speech or writing. Increased engagement with, and usage of a given word, is strongly associated with a learner's knowledge of that word (Ellis, 2002). With increased frequency of exposure to a word comes increased richness of representation in the learner's mental lexicon, which in turn is important for a learner to attain productive control over a given word (Henriksen, 1999).

The distinction between receptive and productive vocabulary has long been accepted by scholars and holds an important position in L2 vocabulary learning theory. Despite this acceptance and theoretical importance, inconsistencies exist in the criteria used to distinguish receptive and productive vocabulary knowledge (Henriksen, 1999; Meara, 1996). Further, for some time there has also been a lack of consistency in the manner by which receptive and productive vocabulary knowledge has been measured (Melka, 1997; Read, 2000). In light of these inconsistencies, Read (2000) suggests using the distinction between *word recognition* and *word recall* to define the constructs of receptive and productive word knowledge. According to Read, recognition is measured by test items which present the target word to the test taker and require the test taker to demonstrate understanding of the word. Recall on the other hand, is operationalised by test items which involve the provision of some stimulus in order to elicit the target word from the test taker's memory. This distinction between recall and recognition will be used to categorise the tests used as part of this investigation as either those which measure receptive or those which measure productive dimensions of vocabulary knowledge respectively.

For Peer Review

Phonological and orthographic word knowledge

A majority of previous research which has examined the relationship between vocabulary knowledge and language macro-skills has focussed on orthographic vocabulary knowledge. Here orthographic vocabulary knowledge is defined as word knowledge measured with test instruments delivered solely in the written modality. Orthographic vocabulary knowledge does have a strong relationship with both reading and listening (Stæhr, 2008, 2009), but as indicated in Nation's (2001) taxonomy, knowing the phonological form of words is also an important dimension of vocabulary knowledge. Milton and Hopkins (2006) demonstrate that although orthographic and phonological knowledge are correlated (r = .68, p < .01), the relationship between these constructs of vocabulary knowledge is not fixed and appears to depend on factors such as learners' first language background. Milton and Hopkins measured both vocabulary knowledge constructs among a group of 126 learners of English with a mix of Arabic and Greek native-speaking backgrounds, and found that overall the phonological vocabulary of learners was not as broad as the learners' orthographic vocabulary. This finding seems unsurprising considering the fixed nature of written words and the transient nature of words encoded in speech. The orthographic form is temporally stable and can be revisited repeatedly by the eye of the reader. In contrast, spoken words are transient and only remain available for perception for a limited period of time. As a consequence, listeners must be able to access knowledge of words presented in the spoken form in a highly time constrained manner. Accessing knowledge of words encoded in spoken language presents considerable difficulties for L2 learners (Field 2008; Goh 2000). Further, phonological word knowledge is strongly and positively associated with L2 listening comprehension (Matthews & Cheng, 2015). It is therefore surprising that so few studies have investigated the association between phonological vocabulary knowledge and language skills, especially those dependent on highly time constrained use of vocabulary knowledge such as L2 listening (Milton, 2013).

L2 word knowledge and L2 listening and reading comprehension

Several previous studies have provided valuable insight into the relationship between L2 vocabulary knowledge and L2 reading. Qian (2002) investigated the relationship between reading and vocabulary

For Peer Review

knowledge among 217 students of English as a second language (ESL) with a broad range of first language backgrounds including Korean, Japanese, Spanish, Chinese, Tajik, Arabic, Portuguese, Russian, Italian and 10 other languages. Vocabulary knowledge was measured with multiple test instruments including instruments aimed at measuring vocabulary depth and size. Vocabulary size was measured with the Vocabulary Levels Test (Nation, 1983) which measures receptive knowledge of words from the second, third, fifth and tenth thousand frequency levels, as well as a group of words which appear regularly in academic text books. Vocabulary depth was measured with an adapted version of the Word Associates Test (Read, 1998), which measured receptive knowledge of synonyms, polysemes and collocations. Each vocabulary test used was administered solely in written form and thus each is assumed to have only tapped orthographic dimensions of vocabulary knowledge. Reading comprehension was measured with a version of the Test of English as a Foreign Language (TOEFL) reading comprehension subtest. Results indicated that a strong and significant correlation existed between vocabulary depth and reading comprehension (r = .77, p < .01) and vocabulary breadth and reading comprehension (r = .74, p < .01) Vocabulary depth alone was shown to be highly predictive of the variance observed in the reading comprehension scores (R^2 = .59, p < .01).

Li and Kirby (2015) also investigated the relationship between vocabulary knowledge and reading comprehension among 246 Chinese middle school students who were learning English in China. Multiple measures of reading comprehension were attained including one which was a traditional multiple choice reading comprehension test and one which was a summary writing test which involved students reading a composition and summarising it in English afterwards without having access to the original text. Multiple measures of vocabulary knowledge were used including vocabulary breadth and depth. Receptive vocabulary breadth was measured with a test which involved matching contextualised target words with their closest synonym via a multiple choice format. Vocabulary depth was measured with three separate measures of knowledge: multiple meanings, morphological awareness, and word definitions. Knowledge of multiple meanings and morphological awareness were both measured in a receptive manner via linguistic information mediated in the written form. Knowledge of word definitions was measured by asking test takers to provide

For Peer Review

definitions of those words and to provide as much information as possible about each word. The final measure of vocabulary depth tapped a productive dimension of vocabulary knowledge. Results indicated that correlations between the measures of vocabulary knowledge and scores for reading were positive and in the most part statistically significant. Vocabulary breadth was more predictive of reading comprehension as measured by the test with the multiple choice format and vocabulary depth was more predictive of comprehension as demonstrated by summary writing. Li and Kirby also found that different elements of vocabulary knowledge could be combined and used in regression models to more powerfully predict reading comprehension.

The two studies reviewed have a similar methodological structure in that they explored the relationship between multiple dimensions of vocabulary knowledge and used measures of L2 reading as an outcome variable. These studies are significant in that they provide empirical data demonstrating the magnitude of association between vocabulary knowledge and reading and furthermore demonstrate that different dimensions of vocabulary knowledge are differentially correlated with and predictive of reading comprehension. However, these studies focussed strongly on receptive, orthographic vocabulary knowledge. As a consequence these results provide limited information relating to the strength of association productive and phonological forms of vocabulary knowledge may have with L2 reading.

A limited number of studies have applied a similar methodological approach to those reviewed above, but with a focus on the relationship between L2 vocabulary knowledge and L2 listening comprehension. Stæhr (2009) investigated the relationship between vocabulary knowledge and listening comprehension of a cohort of 115 advanced Danish EFL learners. Both vocabulary breadth and vocabulary depth were measured. Receptive vocabulary depth was measured with a version of the Vocabulary Levels Test (Schmitt, Schmitt & Clapham, 2001) which measured knowledge of words from the second, third, fifth and tenth thousand frequency levels. Receptive vocabulary depth was also measured with an adapted version of Read's Word Associates Test (1993, 1998). Again, both vocabulary tests used only measured receptive, orthographic knowledge. Listening comprehension was measured with the relevant section of the Cambridge certificate of proficiency in English and involved testing a range of listening skills such as listening for gist, detailed information,

http://mc.manuscriptcentral.com/LTJ

For Peer Review

making inferences and interpreting opinions. Results indicated a strong and significant relationship between receptive, orthographic vocabulary size and listening comprehension (r = .70, p < .01) and receptive, orthographic vocabulary depth and listening comprehension (r = .65, p <.01). Multiple regression analysis indicated that the measure of vocabulary size could predict 49% of the variance observed in the listening comprehension scores. The measure of vocabulary depth only contributed an additional 2% of unique predictive power to the model. Stæhr (2009) affirms that vocabulary knowledge is an important construct in predicting L2 listening comprehension. Further, Stæhr shows that different constructs of vocabulary knowledge, in this case vocabulary breadth and depth, have overlapping but measurably different levels of predictive strength in relation to L2 listening comprehension. However, a major limitation of this study was that both dimensions of vocabulary knowledge were measured in the orthographic form. This limitation is of strong importance considering the modality specific nature of word knowledge and the difficulty L2 learners have in recognizing words from L2 speech (Goh, 2000).

Only a very limited number of studies have systematically sought to investigate the strength of association between phonological word knowledge and L2 listening comprehension. Matthews and Cheng (2015) investigated the strength of association between knowledge of high frequency L2 words presented in the phonological form and L2 listening comprehension. The study was undertaken among 167 native Chinese speaking students studying within a university in China. Word recognition from speech was measured with a partial dictation test which involved test takers transcribing words from speech into the written form. This productive measure of knowledge of the phonological form of words was found to be strongly correlated with L2 listening comprehension as measured by an International English Language Testing System (IELTS) listening test (r = .73, p < .05). Further, phonological word knowledge as measured by the partial dictation test was able to predict 54% of the variance observed in the L2 listening comprehension scores. These results affirm the importance of measures of phonological word knowledge in relation to the skill of L2 listening comprehension.

Few studies have investigated the relationship between vocabulary knowledge and the skills of both reading and listening among a single cohort of L2 learners. Stæhr (2008) investigated the strength of associated between the vocabulary knowledge of 88 Danish EFL students' L2 listening

For Peer Review

and reading comprehension. Receptive, orthographic vocabulary knowledge was measured with a version of the Vocabulary Levels Test (Schmitt, Schmitt & Clapham, 2001). Reading was assessed with a 40 minute test containing 25 multiple choice questions and multiple-matching format. The reading test measured a range of constructs including reading for gist, reading to extract specific information, and making inferences. Listening was assessed with a 20 minute test involving 16 multiple choice questions requiring test takers to listen for specific details, gist, and to use information understood to make inferences. Correlational analysis indicated that receptive, orthographic vocabulary knowledge and listening (r = .69, p < .01) and receptive, orthographic knowledge and reading scores (r = .83, p < .01) were both strongly and positively correlated. These results provide empirical data which demonstrates that a single measure of vocabulary knowledge may be differentially associated with the skills of reading and listening.

Milton, Wade and Hopkins (2010) makes an important contribution toward more fully understanding the relationship between vocabulary knowledge and L2 listening and reading. An important point of difference of their study is that both orthographic and phonological vocabulary knowledge were measured. Additionally, the relationship between these two measures of vocabulary knowledge and *both* reading and listening performance was investigated. The vocabulary knowledge of 30 learners (10 Chinese, 10 Japanese, and 10 from European countries) studying English as a second language was measured with a yes/no computerised test which measured self-reported knowledge of words from the first, second, third, fourth and fifth thousand word frequency levels. Significant correlations were observed between orthographic vocabulary knowledge and both reading and listening as measured by IELTS. Phonological knowledge was shown to have a significant correlation with listening but this was not the case with reading. Although Milton et al. (2010) provide important insight into the association that vocabulary knowledge, both orthographic and phonological, has with listening and reading, there are two important limitations of this study. Firstly, the very small sample size of participants used in the study (N=30) means that only very tentative conclusions from this study can be drawn. Further, although the format of the tests used tapped phonological and orthographic dimensions of vocabulary knowledge, the yes/no test answer formats, as with a majority of the previous studies reviewed above, strongly emphasises receptive aspects of vocabulary

For Peer Review

knowledge. As such the results of the study may not provide insight into the relationship between productive elements of vocabulary knowledge and the skills of L2 listening and reading.

The preceding reviews highlight a number of aspects relevant to the present study. Firstly, it is evident that vocabulary knowledge is a strong correlate and predictor of L2 listening and reading. Secondly, different dimensions of vocabulary knowledge possess varying strengths of association and predictive value in relation to L2 listening and reading. Further, a majority of the vocabulary test instruments used in the studies reviewed only measured orthographic and receptive dimensions of vocabulary knowledge. This trend is of note in light of the known multidimensionality of vocabulary knowledge: a multidimensionality which is known to encompass both phonological and productive dimensions (Nation, 2001). Finally, none of the studies reviewed above have measured multiple dimensions of vocabulary and both L2 listening and reading comprehension among a single large cohort of L2 learners.

The current study aims to provide additional information about the relationship between different dimensions of vocabulary and L2 listening and reading. Three different forms of vocabulary knowledge are measured among a single large cohort of L2 learners: receptive/orthographic (RecOrth) vocabulary knowledge, productive/orthographic (ProOrth) vocabulary knowledge, and productive/phonological (ProPhon) vocabulary knowledge. The constructs underpinning these vocabulary measures are analysed and their relative correlation with and prediction of L2 listening and reading are assessed and discussed.

Research questions

RQ1: What is the relative magnitude of correlation of receptive/orthographic (RecOrth) vocabulary knowledge, productive/orthographic (ProOrth) vocabulary knowledge and productive/phonological (ProPhon) vocabulary knowledge with L2 reading and L2 listening?

RQ2: How many discrete factors underpin the subcomponents of the RecOrth, ProOrth, and ProPhon vocabulary knowledge tests, and what commonalities do the variables which load onto these factors share?

For Peer Review

RQ3: To what extent can RecOrth, ProOrth, and ProPhon vocabulary knowledge, either individually or in combination, predict the variance observed in L2 listening and L2 reading?

Method

Participants

The participants were 250 tertiary level EFL students at a large university in the People's Republic of China. The participant group was made up of 7 different class groups possessing a range of 24 different majors. The total group consisted of 159 males and 91 females, with ages between 18 and 23 with an average age of 19. All possessed Mandarin Chinese as their first language.

Vocabulary test development

The formats of the vocabulary tests used to measure RecOrth, ProOrth, and ProPhon vocabulary knowledge were based on the formats of three pre-existing and broadly used vocabulary test types: Vocabulary Levels Tests (Nation, 2001; Schmitt, Schmitt & Clapham, 2001), controlled-production vocabulary levels test (Laufer & Nation, 1999), and partial dictation (Matthews & Cheng, 2015).

The target words to be used in the vocabulary tests were chosen from four categories of word frequency level as determined by comparison with word lists generated with the British National Corpus. These four categories included words from the first thousand frequency level (1K), the second thousand frequency level (2K), the third thousand frequency level (3K) and a combined category which included words from both the fourth and fifth thousand frequency level (4/5K). The rationale for selecting target words up to and including the five thousand frequency level was that knowledge of this range of words is likely to make up the core of lexical knowledge needed for adequate levels of L2 listening and L2 reading (Adolphs & Schmitt, 2003; Laufer & Ravenhorst-Kalovski, 2010; Webb & Rodgers, 2009a, 2009b).

Target words for each of the three tests were unique to ensure that learning or priming effects between the different tests were not a factor. Groups of words were selected to ensure that the corresponding frequency level subcomponents of the three test formats contained words of similar length and form. Each subcomponent for each vocabulary test contained a set ratio of word categories

http://mc.manuscriptcentral.com/LTJ

For Peer Review

(50% nouns, 25% verbs, 25% adjectives). Care was taken to ensure that the lexical items used as part of the contextual sentences or word meaning used in the tests were drawn from the same or higher frequency level than that or the target word for each item. All test items were piloted in two stages, first with a panel of native speakers and secondly with a group of 40 EFL students representative of the population from within which the test sample was taken.

Testing Receptive/Orthographic (RecOrth) Vocabulary Knowledge

The instrument used to measure RecOrth vocabulary knowledge had an item structure identical to that of the Vocabulary Levels Tests (Nation, 2001; Schmitt, Schmitt & Clapham, 2001). Each item requires three target words from a list of six words to be matched with their meanings (See Figure 1). This test format presents stimulus only in the written modality and thus taps orthographic vocabulary knowledge. Additionally, the test is assumed to tap receptive vocabulary knowledge as the target words are presented to the test taker. The RecOrth vocabulary knowledge test had 32 items, with a total of 96 target words. There were 24 target words from each of the four frequency levels tested (1K, 2K, 3K, 4/5K).

Testing Productive/Orthographic (ProOrth) Vocabulary Knowledge

The instrument used to measure ProOrth vocabulary knowledge had an item structure identical to that of the controlled-production vocabulary levels test (Laufer & Nation, 1999). This test format requires test takers to produce a single target word in a blank space of a contextual sentence (See Figure 2). This format presents stimulus solely in the written modality and therefore taps orthographic vocabulary knowledge. This test is assumed to tap a form of controlled productive vocabulary knowledge, as the test taker is required to produce the form of the target word, albeit only partially due to the first letters of the target word being provided. Following Laufer and Nation (1999) a majority of the items were supplied with either two or three of the initial letters of the target word, and in some instances four of the initial letters were provided. The rationale for providing the initial letters of the target word was to ensure that only the target words were elicited. The ProOrth vocabulary

http://mc.manuscriptcentral.com/LTJ

For Peer Review

knowledge test instrument had 64 items each with one target word. There were 16 words from each of the four frequency levels tested (1K, 2K, 3K, 4/5K).

Figure 1. Example item used to test RecOrth vocabulary knowledge

Example question:

words	**Meanings**
1 conference 2 economy 3 item 4 percentage 5 prospect 6 quotation	_____ meeting _____ object _____ part of 100

Example answer:

words	**Meanings**
1 conference 2 economy 3 item 4 percentage 5 prospect 6 quotation	___1__ meeting ___3__ object ___4__ part of 100

Figure 2. Example item used to test ProOrth vocabulary knowledge

Example question: He was riding a bic____

Example answer: He was riding a bicycle

Testing Productive/Phonological (ProPhon) Vocabulary Knowledge

The instrument used to measure ProPhon vocabulary knowledge was a partial dictation test (Matthews & Cheng, 2015; Matthews, Cheng & O'Toole, 2015) which requires test takers to produce a single target word after listening to a spoken stimulus sentence once (See Figure 3). This test format taps the test taker's ability to recognize the phonological form of the target word and produce that target word in the written form. As such this test taps a productive construct of vocabulary knowledge. Further, the language stimulus which enables the test taker to produce the target word is in the aural modality and thus taps a phonological dimension of word knowledge. As it is important that such tests are not

For Peer Review

primarily a measure of spelling (Buck, 2001) a scoring rubric based on the principles outlined in Matthews, O'Toole and Chen (2016) was implemented to ensure that minor spelling errors were not penalised when scoring these tests.

Figure 3. Example item used to test ProPhon vocabulary knowledge

> *Example question*: Her two favourite subjects at university were _____and computer studies.
>
> *AURAL STIMULUS IS HEARD ONCE*
>
> *Example answer*: Her two favourite subjects at university were __finance____and computer studies.

Listening Tests

A published practice version of an IELTS listening test was used to measure all 250 participants' L2 listening. The test contained 40 questions and required participants to listen to two monologues and two dialogues and evidence their comprehension in a number of ways including multiple choice, labelling a diagram, short answer responses, and completing tables, notes and sentences. The stimulus was heard only once although the test takers were able to answer while they listened. Participants were also given ten minutes after hearing the listening text to check and complete their answers. According to self-report, none of the participants had attempted this version of the test before.

Reading Tests

Reading was measured with a test developed in-house similar to those routinely used to test reading comprehension in the research setting. The test contained 8 written paragraphs each of approximately 500 words, with each paragraph having 5 associated multiple choice questions (40 questions in total). Each group of five multiple choice questions for each paragraph had a standard structure: three questions dependent on the literal comprehension of the text (one question about the main idea of the paragraph and two questions relating to details in the text), and two questions which depended on being able to make inferences which went beyond a literal reading of the text.

For Peer Review

The stimulus material for both the listening test (transcribed) and reading test were analysed using *Vocabprofile* located on the *Compleat Lexical Tutor* website (Cobb, n.d.) against frequency ordered word lists extracted from the British National Corpus (BNC). Over 97% of the words in the listening test stimulus and over 98% of the reading test stimulus were within the five thousand word frequency range (See Table 1).

Table 1. Frequency analysis of the lexical content of listening and reading tests

Frequency level (K)	Word families (%)		Tokens (%)		Cumulative tokens (%)	
	Listening test	Reading test	Listening test	Reading test	Listening test	Reading test
1	394 (69.24)	516 (57.27)	2025 (86.58)	3928 (81.63)	86.58	81.63
2	92 (16.17)	200 (22.20)	156 (6.67)	483 (10.04)	93.25	91.67
3	31 (5.45)	60 (6.66)	52 (2.22)	140 (2.91)	95.47	94.58
4	17 (2.99)	63 (6.99)	21 (0.90)	127 (2.64)	96.37	97.22
5	14 (2.46)	24 (2.66)	18 (0.77)	67 (1.39)	97.14	98.61

Procedure

Participants undertook the three vocabulary knowledge tests in a single session which was part of regular English language classes (N = 250). Listening and reading tests were undertaken after the vocabulary tests in an additional single session. Three students were unable to undertake the reading test (N = 247). Tests were undertaken silently and without interaction between participants. Tests were scored by a team of scorers under the training and instruction of the principal investigator.

Data analysis

Four stages of data analysis were undertaken. Firstly, test scores derived from the vocabulary, listening and reading tests were analysed to produce a range of descriptive and reliability statistics.

For Peer Review

Next correlational analysis was undertaken to determine the magnitude of the relationship between measures of vocabulary knowledge and measures of L2 listening and reading. Thirdly, factor analysis was used to investigate constructs underpinning the subcomponents of each vocabulary test format. Finally, two hierarchical multiple regression models were built to quantify the degree to which total scores from the three vocabulary tests contributed to the prediction of variance observed within listening and reading scores.

Results

Descriptive statistics

The minimum, maximum, mean scores and standard deviations for each test administered are shown in Table 2. Values of skewness and kurtosis show that scores were normally distributed to an acceptable level. Mean scores suggest ProPhon and ProOrth vocabulary knowledge tests were more demanding for the participant group than was the RecOrth vocabulary knowledge test. These results support previous findings that the productive control of vocabulary knowledge is a developmentally more advanced state of word knowledge than is receptive vocabulary knowledge.

Table 2. Descriptive statistics for scores obtained from the test instruments

Test	N	Min (%)	Max (%)	Mean (%)	SD	Skewness	Kurtosis	Cronbach's α
RecOrth	250	42.71	100	82.42	10.04	-.73	1.20	.781
ProOrth	250	3.13	90.63	54.79	16.61	-.17	-.11	.879
ProPhon	250	4.69	82.81	47.47	15.77	-.21	-.07	.902
L2 Listening	250	0	86.25	39.46	15.21	.07	-.08	.807
L2 Reading	247	32.50	90.00	67.03	11.07	-.37	-.03	.615

RQ1: What is the relative magnitude of correlation between RecOrth, ProOrth and ProPhon vocabulary knowledge and L2 reading and L2 listening?

For Peer Review

The correlation between all measures of vocabulary and L2 listening and reading are shown in Table 3. All measures of vocabulary knowledge are significantly correlated with both L2 listening and reading. Listening was strongly correlated with ProPhon vocabulary knowledge (r =.71, p <.001), and moderately correlated with RecOrth vocabulary knowledge (r =.39, p <.001) and ProOrth vocabulary knowledge (r =.55, p <.001). Reading was moderately correlated with each of the measures of vocabulary knowledge, with the strongest correlate of L2 reading being ProOrth vocabulary knowledge (r = .57, p <.001). The correlations between the total scores for each of the vocabulary tests were moderate to strong.

Table 3. Correlation between scores obtained from test instruments

Measures	1	2	3	4	5
1. RecOrth	-				
2. ProOrth	.81**	-			
3. ProPhon	.59**	.72**	-		
4. L2 Listening	.39**	.55**	.71**	-	
5. L2 Reading	.46**	.57**	.46**	.44**	-

Notes.
N = 250, except for L2 reading N = 247.
** p <.001 (2-tailed)

RQ2: How many discrete factors underpin the subcomponents of the RecOrth, ProOrth, and ProPhon vocabulary knowledge tests, and what commonalities do the variables which load onto these factors share?

In exploring the relative importance of different measures of L2 vocabulary knowledge it is important to establish the validity of the constructs the test instruments are assumed to measure. Factor analysis was thus used as a statistical approach to investigate the constructs which underpinned the three vocabulary tests used. An assessment to ensure that the data satisfied the assumptions needed for factor analysis was first undertaken. Examination of the correlation matrix between the

http://mc.manuscriptcentral.com/LTJ

For Peer Review

subcomponents of the three tests (twelve variables) indicated a positive manifold in the data. This finding suggested a general factor in the data which was as expected as each test measured different aspects of L2 vocabulary knowledge. All twelve subsections were correlated to at least .40 with at least one other of the vocabulary test subsection scores, suggesting an adequate level of factorability of the variables.

Table 4. Pattern matrix for factor analysis

	Factors	
Variable (test sub-component)	1	2
RecOrth 1K	**.50**	.12
RecOrth 2K	**.69**	-.02
RecOrth 3K	**.81**	-.04
RecOrth 4/5K	**.86**	.03
ProOrth 1K	**.53**	-.23
ProOrth 2K	**.64**	-.18
ProOrth 3K	**.72**	-.22
ProOrth 4/5K	**.68**	-.24
ProPhon 1K	.05	**-.78**
ProPhon 2K	.05	**-.83**
ProPhon 3K	-.06	**-.87**
ProPhon 4/5K	.24	**-.69**

Note.
N = 250.
Factor loadings greater than .30 are shown in bold text.

A maximum likelihood extraction method with direct oblimin rotation was used to conduct the factor analysis. Twelve variables were used in the analysis, namely the percentage scores for each of the four word frequency level subsections (1K, 2K, 3K, 4/5K) which were subcomponents of the RecOrth, ProOrth and ProPhon vocabulary knowledge tests. Examination of the resultant initial Eigen

For Peer Review

values generated indicated the presence of two factors with Eigen values with a magnitude greater than one. Visual appraisal of the most evident point of inflection on the resultant scree plot also indicated the presence of two factors.

The two factors explained 62.57% of the cumulative variance. The first factor explained 54.70% of that variance and the second factor explained an additional 7.87%. The pattern matrix for the factor analysis is provided in Table 4. Overall these results indicate that all vocabulary tests subcomponents with formats mediated solely through the orthographic modality loaded onto one factor (*orthographic vocabulary factor*). All subcomponents with a phonological component loaded onto another factor (*phonological vocabulary factor*). All variables had a minimum primary loading of .50, with no cross-loading values exceeding .30.

RQ3: To what extent can RecOrth, ProOrth and ProPhon vocabulary knowledge, either individually or in combination, predict the variance observed in L2 listening and L2 reading?

In order to determine the extent to which the predictor variables of RecOrth, ProOrth and ProPhon vocabulary knowledge could explain the variance observed in L2 listening or reading, hierarchical multiple regression analyses were used. Each of the hierarchical models was built using the same logic for the entry order of variables. The predictor variables with the strongest magnitude of correlation with the outcome variable were added in the first step (ProPhon vocabulary knowledge for listening and ProOrth vocabulary knowledge for reading). The remaining two predictor variables were then added in the second step. For each multiple regression analysis undertaken, tolerance levels were all well above .20 indicating that multicollinearity was not an issue. Visual appraisal of the scatterplot between residuals and the predicted values clearly indicated the assumption of homoscedasticity was met.

For Peer Review

Table 5. Hierarchical regression model addressing the predictive value of vocabulary knowledge on listening comprehension (model 1)

				unstandardized		standardized
	R	R^2	ΔR^2	*B*	SE *B*	β
Step 1	.711	*.506***				
Constant				6.887	2.152	
ProPhon				.686	.043	.711***
Step 2	.720	.518***	.012*			
Constant				19.936	6.809	
ProPhon				.638	.062	.662***
ProOrth				.189	.081	.207*
RecOrth				-.257	.115	-.169*

Note.
N = 250.
* *p* < .05, *** *p* <.001

The first model (see Table 5) sought to determine the relative predictive value of RecOrth, ProOrth and ProPhon vocabulary knowledge on L2 listening scores. Thus, model 1 was built by entering ProPhon vocabulary knowledge in the first step and then by entering ProOrth and RecOrth vocabulary knowledge in the second step. The first step explained 50.6% of the variance in L2 listening. After entry of RecOrth and ProOrth vocabulary knowledge in the second step, the total variance explained was 51.8%. In summary, ProPhon vocabulary knowledge provided the greatest unique contribution in predicting variance in L2 listening. Entering RecOrth and ProOrth vocabulary knowledge in step two added a small, yet statistically significant contribution of 1.2% to the predictive power of the model.

Standardized beta weights (β) provide a measure of the degree to which a dependent variable, in this case L2 listening comprehension test scores, changes when an independent variable increases by one standard deviation while the other independent variables are kept constant. The beta weights presented in Table 5 reaffirm the comparative strength of the relationship between ProPhon

For Peer Review

vocabulary knowledge and L2 listening comprehension, when compared to that of both ProOrth and RecOrth vocabulary knowledge and L2 listening comprehension.

Model 2 (see Table 6) sought to determine the relative predictive value of RecOrth, ProOrth and ProPhon vocabulary knowledge test scores on L2 reading scores. This model was built by entering ProOrth vocabulary knowledge in the first step and ProPhon and RecOrth vocabulary knowledge in the second step. The first step explained 32.8% of the variance in L2 reading. Entry of the other two variables into the second step did not provide a statistically significant unique contribution to the predictive power of the model.

Table 6. Hierarchical regression model addressing the predictive value of vocabulary knowledge on L2 reading (model 2)

				unstandardized		standardized
	R	R^2	ΔR^2	*B*	SE *B*	β
Step 1	*.573*	*.328***				
Constant				45.973	2.010	
ProOrth				.383	.035	.573***
Step 2	.577	.333***	.005			
Constant				45.665	5.847	
ProOrth				.338	.071	.506***
ProPhon				.068	.053	.098
RecOrth				-.006	.099	-.005

Note
$N = 247$.
*** $p < .001$

Discussion

The relationships between measures of vocabulary knowledge

The effective and valid use of tests strongly depends on an empirical basis from which to understand the constructs which underpin them. In this study, factor analysis clearly indicated that

For Peer Review

subcomponents of the RecOrth and the ProOrth vocabulary knowledge tests loaded onto the same factor, and that ProPhon vocabulary knowledge test subcomponents loaded onto another. This finding is strongly suggestive of the functional importance of distinguishing between *phonological vocabulary factors* and *orthographic vocabulary factors*. This finding supports previous research which suggests that measures of orthographic and phonological vocabulary tap different constructs (Milton & Hopkins, 2006; Milton et al., 2010). The very strong correlation between RecOrth and ProOrth vocabulary knowledge (r = .81, p < .001) is another indication of the substantive overlap between orthographic vocabulary factors.

Although factor analysis showed ProOrth and ProPhon vocabulary knowledge loaded onto different factors, their strong correlation (r = .72, p < .001) is suggestive of a behavioural similarity among items which tapped productive dimensions of vocabulary knowledge. The ProOrth vocabulary knowledge test (M = 54.79%, SD = 16.61) and ProPhon vocabulary knowledge test (M = 47.47%, SD = 15.77) were clearly more demanding for the participant group than the RecOrth test (M = 82.42%, SD = 10.04). The two vocabulary tests with a productive dimension apparently provided a more stringent gauge of vocabulary knowledge.

The relationship between vocabulary knowledge and L2 listening

The finding that L2 listening was most strongly correlated with ProPhon vocabulary knowledge (r = .71, p < .001), and more moderately correlated with RecOrth vocabulary knowledge (r = .39, p < .001) and ProOrth vocabulary knowledge (r = .55, p < .001), strongly suggests the specificity of vocabulary knowledge mediated through the aural modality and its importance in relation to L2 listening. The practical significance of phonological vocabulary knowledge is also reinforced by the substantive predictive capacity ProPhon vocabulary knowledge had in accounting for the variance observed in L2 listening scores (R^2 = .506, p < .001). These results show that the type of word knowledge tapped by the ProPhon vocabulary knowledge test, namely that which is phonological in nature and that which can also be accessed under time constraints, strongly supports effective L2 listening. The finding that orthographic vocabulary knowledge offered a very small unique contribution to the prediction of L2 listening over and above that offered by ProPhon vocabulary

For Peer Review

knowledge (ΔR^2 = .012, p < .05) again speaks to the specificity of phonological vocabulary knowledge in relation to L2 listening. Further, it is of note that the contribution RecOrth vocabulary knowledge made in predicting variance observed in L2 listening was associated with a negative standardised beta weight (β = -.169, p < .05) (See Table 5). This indicates that for each unit increase in RecOrth vocabulary knowledge, while ProOrth and ProPhon vocabulary knowledge are kept constant, a small but statistically significant decrease in L2 listening score is evident. Although speculative, it seems possible that the receptive and relatively automaticity-independent form of word knowledge which may support achievement on RecOrth vocabulary knowledge tests may in fact have a small but measurable supressing effect on the ability to process and comprehend spoken language. Although again highly speculative, a possible explanation for this occurrence may relate to the mode of vocabulary learning known to be undertaken by some members of the sample group. Leaners that have emphasized the development of vocabulary knowledge through highly controlled cognitive processes, such as translation from L2 to L1 and back again, may have a relatively strong capacity to form receptive L2 form-meaning linkages as indicated by vocabulary tests which are receptive and orthographic in nature. However, it could be hypothesised that an emphasis on such approaches to developing vocabulary knowledge may, in a practical sense, preclude the development of vocabulary fluency, a dimension of vocabulary knowledge known to be critically important for language skills such as listening (Hulstijn, 2003).

The relationship between vocabulary knowledge and L2 reading

Of the three measures of vocabulary knowledge investigated, ProOrth vocabulary knowledge was most strongly correlated with L2 reading (r = .57, p < .001). The individual strength of correlation between ProPhon and RecOrth vocabulary knowledge and L2 reading was of an equivalent magnitude (r = .46, p < .001). Although factor analysis indicated that ProOrth and RecOrth each loaded onto the same factor, hierarchical multiple regression indicated that ProOrth vocabulary knowledge was the only variable which offered a unique contribution in explaining the variance observed within L2 reading scores (R^2 = .33, p < .001). These results are in contrast to those of Qian (2002) who found that receptive measures of vocabulary size predicted 54% of the variance observed in L2 reading

http://mc.manuscriptcentral.com/LTJ

For Peer Review

scores. A possible explanation for the differences between these results may relate to the broader range of word frequency levels tested by Qian: namely the 2K, 3K, 5K *and* ten thousand (10K) frequency ranges. We suggest that receptive word knowledge at the lower frequency levels, such as those beyond the 10K level, may provide a more rigorous measure of the word knowledge or general language proficiencies associated with successful L2 reading.

The finding that ProPhon vocabulary knowledge did not provide a unique contribution to the model built to predict L2 reading again reiterates the modality specific nature of word knowledge.

Limitations

This study has several limitations. Firstly, due to issues of feasibility only a relatively limited range of word frequency levels were investigated (1K – 5K). We suspect that including test items which tapped knowledge of words of lower frequency levels, such as those from the 10K range would have added additional explanatory power to the predictive models used in this study. To assess the magnitude of this limitation, future investigations could examine the relative predictive value of word knowledge of different frequency levels on L2 listening and reading. Secondly, only a relatively limited range of vocabulary measures were used in this study. To more fully understand the relationship between vocabulary knowledge and L2 listening and reading, it is suggested that traditional measures of vocabulary depth (Read, 1993) and those which tap constructs of vocabulary fluency be included in future analyses. Thirdly, it is necessary to mention the difficulties involved in differentiating and adequately operationalizing the constructs of vocabulary knowledge used in this study. For example, although our measure of phonological knowledge was defined as one which involved processing information presented in the phonological form, test takers were required to evidence their phonological knowledge through orthographic means. As noted, steps were taken to reduce the threat to validity presented by this issue (see page 15); however, operationalizing the construct of phonological vocabulary knowledge without also engaging other dimensions of word knowledge presented a strong practical challenge. Indeed, developing alternative approaches to operationalizing discrete constructs of vocabulary knowledge, which are by necessity applied in a highly integrated manner during language use, presents an area for future research effort.

For Peer Review

Another limitation of the study was that the measurement of productive vocabulary knowledge was highly controlled in order to measure the test takers' knowledge of a prescribed range of words from various frequency levels. Although this approach enabled the collection of the data required for the objectives of this research, such modes of measurement may be considered a threat to the validity of the construct of productive vocabulary knowledge. Including productive measures of word knowledge as indicated by the range and control of words used by participants during spontaneous speech and writing, although not without its own logistical challenges, is recommended for future analyses. Lastly, it is important to note that this study was undertaken among a homogenous sample group in terms of contextual language learning factors and first language background. As such, caution needs to be applied when generalizing these results to other groups of learners. Indeed, future research which investigates the influence that first language background has on the relationship between various measures of L2 vocabulary knowledge and L2 macro-skills is of strong interest.

Implications and conclusions

The findings presented here have a number of implications for the testing and teaching of L2 vocabulary in real language learning contexts. Firstly, it is assumed that the primary objective of measuring L2 vocabulary knowledge is to acquire accurate information about the degree to which learners' current level of vocabulary knowledge supports fundamental language skills such as listening and reading. If this assumption holds true, then the use of test formats which tap productive dimensions and phonological dimensions of word knowledge should be encouraged. Extrapolation from our results would suggest that test instruments which only measure orthographic and receptive forms of vocabulary knowledge are likely to be less predictive of listening and reading than equivalent forms of vocabulary tests which also encompass both productive and phonological vocabulary knowledge. Based on this assertion, we suggest a move away from vocabulary tests which only measure orthographic and receptive knowledge. More widespread use of vocabulary tests which tap productive/orthographic and productive/phonological vocabulary knowledge is likely to have important pedagogical implications. For example, such measures are likely to provide a more robust metric for diagnosing which areas of lexical knowledge are most likely to be inhibiting L2 listening

For Peer Review

and reading performance. Additionally, positive washback is likely to result from increasing the degree to which productive and phonological dimensions of vocabulary knowledge are emphasised in language tests. Vocabulary tests which require the fluent and productive control of vocabulary knowledge are likely to encourage both teachers and students to emphasise the dimensions of vocabulary knowledge required to recall words more precisely from memory. Similarly, tests which measure knowledge of the phonological form of target words and which require this knowledge to be accessed and applied under time constraints are also likely to yield positive washback. In light of the specificity of phonological vocabulary knowledge and its strong positive relationship with L2 listening, it is important to have testing systems in place which emphasise this fundamental dimension of vocabulary knowledge.

This study has addressed the relationship between three different measures of vocabulary knowledge and the skills of L2 listening and reading. Productive phonological and productive orthographic vocabulary dimensions loaded onto different factors, and were each shown to be the most predictive in terms of the variance observed in L2 listening and L2 reading respectively. The receptive orthographic measure of vocabulary knowledge contributed either very little or not at all to the predictive power of regression models seeking to explain variance observed within L2 listening and reading scores. Pedagogical approaches which emphasise the development of both productive and phonological vocabulary knowledge are therefore recommended. Such approaches will assist L2 learners to develop vocabulary knowledge which extends beyond receptive form-meaning links and will enhance the dimensions of vocabulary knowledge which most strongly support L2 listening and reading.

For Peer Review

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage in spoken discourse. *Applied Linguistics*, *24*, 425-438.

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

Cobb, T. (n.d.). Web Vocabprofile. Retrieved from http://www.lextutor.ca/vp/available

Daller, H., Milton, J. & Treffers-Daller, J. (2007). Editors' introduction: conventions, terminology and an overview of the book. in H. Daller, J. Milton and J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 1-32). Cambridge, England: Cambridge University Press.

Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicitand explicit lnguage acquisition. *Studies in Second Language Acquisition, 24*, 143-188.

Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, *31*, 105-119.

Field, J. (2008). Revising segmentation hypotheses in first and second language listening. *System*, *36*, 35-51.

Fraser, C. A. (2007). Reading rate in L1 Mandarin Chinese and L2 English across five reading tasks. *The Modern Language Journal*, *91*, 372-394.

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*, 55-75.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*, 303-317.

For Peer Review

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research, 18*, 193-223.

Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning*, *5*, 413-425.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics, 19*, 255-271.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength and computer adaptiveness.' *Language Learning, 54*, 399-436.

Laufer, B., & Nation, I.S.P. (1999). A vocabulary-size test of controlled productive ability.' *Language Testing, 16*, 33-51.

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

Li, M., & Kirby, J. R. (2015). The effect of vocabulary breadth and depth on English reading. *Applied Linguistics, 36,* 611-634.

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System, 52*, 1-13.

Matthews, J., Cheng, J., & O'Toole, J.M. (2015). Computer-mediated input, output and feedback in the development of L2 word recognition from speech. *ReCALL,* 27, 321-339.

Matthews, J., O'Toole, J. M., & Chen, S. (2016). The impact of word recognition from speech (WRS) proficiency level on interaction, task success and word learning: design implications for CALL to develop L2 WRS. *Computer Assisted Language Learning*, 1-26. http://dx.doi.org/10.1080/09588221.2015.1129348

For Peer Review

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

Melka, F. (1997). Receptive versus productive aspects of vocabulary. In N. Schmitt and M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy*. (pp. 84-102), Cambridge, Engand: Cambridge.

Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol, England: Multilingual Matters.

Milton, J. (2013). L2 vocabulary acquisition, knowledge and use. *EUROSLA Monographs Series* 2, 57-78.

Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: do vocabulary tests underestimate the knowledge of some learners. *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 and M. López-Jiménez (Eds.), *Further insights into non-native vocabulary teaching and learning. (*pp. 83-97). Bristol, England: Multiligual Matters.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines, 5*, 12-25.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge, England: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*, 59-82.

For Peer Review

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning, 52*, 513-536.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*, 355-371.

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, England: Cambridge University Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*, 329-363.

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.

Segalowitz, N., & Hulstijn, J. (2005). Automaticty in bilingualism and second language leaerning. In J. F. Kroll & A. M. B. DeGroot (Eds.), *Handbook of bilingualism* (pp. 371-388). Oxford, England: Oxford University Press.

Stæhr, L. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal, 36*, 139-152.

Stæhr, L. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*, 577-607.

Webb, S. (2005). Receptive and productive vocabulary learning - The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*, 33-52.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30*, 79-95.

For Peer Review

Webb, S., & Rodgers, P. (2009a). Vocabulary demands of television programs. *Language Learning, 59*, 335-366.

Webb, S., & Rodgers, P. (2009b). The lexical coverage of movies. *Applied Linguistics, 30*, 407-427.

For Peer Review