

L2 Vocabulary Knowledge and L2 Listening Comprehension: A Structural Equation Model

Yang Li

Guangdong University of Foreign Studies

Xian Zhang

University of North Texas

Abstract

The current study investigates how second language (L2) listening comprehension is associated with three dimensions of L2 vocabulary knowledge: size, depth, and fluency. Vocabulary knowledge tests administered to 290 participants measured L2 auditory vocabulary size, depth, and fluency. Afterward, participants took an International English Language Testing System (IELTS) listening test that measured L2 listening comprehension. Using a structural equation modeling technique, we found that all three dimensions of vocabulary knowledge are significant predictors of L2 listening comprehension. Size of auditory vocabulary in the L2 has the strongest predictive power over L2 listening comprehension. The results of the current study offer useful pedagogical implications for improving L2 listening comprehension.

Résumé

La présente étude examine comment la compréhension orale de la deuxième langue (L2) est associée à trois dimensions de la connaissance du vocabulaire de la L2 : la taille, la profondeur et la fluidité. Des tests de connaissance du vocabulaire mesurant la taille, la profondeur et la fluence du vocabulaire auditif de la L2 ont été administrés à 290 participants. Ensuite, les participants ont passé un test d'écoute *International English Language Testing System* (IELTS) qui mesure la compréhension à l'écoute de la L2. En utilisant une technique de modélisation par équation structurelle, nous avons constaté que les trois dimensions de la connaissance du vocabulaire sont des prédicteurs significatifs de la compréhension orale de la L2. La taille du vocabulaire auditif L2 a la plus forte puissance prédictive sur la compréhension orale de la L2. Les résultats de la présente étude offrent des implications pédagogiques utiles pour améliorer la compréhension orale L2.

L2 Vocabulary Knowledge and L2 Listening Comprehension: A Structural Equation Model

Introduction

L2 Listening Comprehension and L2 Vocabulary Knowledge

Various factors may influence second language (L2) listening comprehension. For example, L2 listening comprehension is found to be associated with first language (L1) listening comprehension (Vandergrift, 2006), L2 learner motivation (Vandergrift, 2005), and L2 learner meta-cognitive strategy use (Graham & Macaro, 2008; Vandergrift & Goh, 2012).

Among various aspects of L2 knowledge (e.g., L2 grammar knowledge, L2 pragmatic knowledge, L2 literacy), L2 vocabulary knowledge has been recognized as one of the most important aspects of L2 knowledge (Bernhardt & Kamil, 1995). Indeed, vocabulary is a basic unit for meaning construction and expression (Nation, 2001). Without sufficient L2 vocabulary knowledge, L2 listening and reading comprehension can be extremely difficult, if not impossible.

It is generally agreed that L2 vocabulary knowledge is important for L2 reading comprehension. This is because knowing the form and the meaning of a certain number of words is necessary for L2 reading comprehension (Nation, 2006). Nation (2006) showed that 8,000-9,000 word families are necessary to read different types of authentic texts fluently. Laufer and Ravenhorst-Kalovski (2010) argued that knowledge of 8,000 word families is the optimal threshold for adequate reading comprehension. These findings suggest that knowing more vocabulary items (or having a larger L2 vocabulary size) is beneficial for L2 reading comprehension. Evidence shows that vocabulary size is positively correlated with reading ability (e.g., Laufer, 1992; Horst, Cobb, & Meara, 1998; Qian, 2002). In a more recent study, Schmitt, Jiang, and Grabe (2011) reported a positive correlation between the percentage of known vocabulary items in a text and reading comprehension. In a large scale study, Alderson (2005) found that vocabulary size accounts for 40% of variance in reading performance.

The importance of L2 vocabulary knowledge is not limited to L2 reading comprehension. Knowledge of L2 vocabulary is also crucial for L2 listening comprehension. Knowledge of 3,000 word families is necessary to carry out successful everyday conversations (Adolphs & Schmitt, 2003). A 98% lexical coverage rate of L2 spoken texts is desirable for good listening comprehension (Stæhr, 2009). Similar to L2 reading comprehension, L2 listening comprehension also depends on vocabulary size. It has been found that L2 vocabulary size is positively correlated with L2 listening comprehension. Alderson (2005) reported that vocabulary size is significantly correlated with L2 listening performance (.60, $p < .01$). Stæhr (2009) also found a high correlation of .70 between L2 listening comprehension and L2 vocabulary size. In a more recent study,

Vandergrift and Baker (2015) used structural equation modeling to investigate the causal relationship between L2 listening comprehension and a number of variables such as L1 listening comprehension, working memory, auditory discrimination ability, and L2 vocabulary size. Vandergrift and Baker found that the correlation between L2 listening comprehension and L2 vocabulary size was .51 ($p < .01$). Moreover, L2 vocabulary size had a direct causal influence on L2 listening comprehension, lending support to the importance of L2 vocabulary size on L2 listening comprehension.

There is little doubt that vocabulary size plays a critical role for L2 comprehension. However, the significance of vocabulary knowledge is not limited to vocabulary size (i.e., the quantity of vocabulary knowledge). It has been argued that other dimensions of vocabulary knowledge (e.g., depth and fluency) are also important (Read, 2000). Read (2000) stated that vocabulary knowledge can be viewed as having three dimensions: breadth, depth, and fluency. Different from vocabulary size (i.e., vocabulary breadth), which refers to “the number of words the meaning of which one has at least some superficial knowledge” (Qian, 2002, p. 515), vocabulary depth refers to how well a lexical item is known (Qian, 2002). Vocabulary fluency refers to how fast vocabulary knowledge can be processed.

In addition to knowing more vocabulary items, knowing each individual item better can also benefit comprehension as L2 vocabulary knowledge is not seen as limited to knowing a word’s meaning and form (Nation, 2001). In addition to form and meaning, Nation (2001) suggested that there are other aspects of L2 vocabulary knowledge, such as grammatical function (part of speech), collocation (e.g., *scrambled* can collocate with *eggs* while *muddled* cannot), and constraints on use (e.g., in Chinese, *Nin [you]* is preferred over *Ni [another term for you]* when addressing seniors). In order to understand and use L2 vocabulary appropriately, one needs to have sufficient knowledge of L2 vocabulary items (Nation, 2001; Schmitt, 2010; Wesche & Paribakht, 1996). As de Bot, Paribakht, and Wesche (1997) have found, for L2 listening comprehension, vocabulary depth is beneficial because it can be linked to successful inferencing (i.e., guessing meanings of unknown word). Indeed, Qian (2002) showed that a greater level of vocabulary depth is associated with a higher success rate in inferencing. Moreover, L2 vocabulary depth has been found to be positively related to L2 listening comprehension (Stæhr, 2009).

Another dimension of vocabulary knowledge is vocabulary fluency. The fluency of L2 vocabulary can influence L2 listening comprehension as greater vocabulary fluency facilitates language processing (Segalowitz, 2005). It has been found that vocabulary fluency may also be a key factor related to performance in listening and reading (Segalowitz, 2005). This is because a greater vocabulary fluency is beneficial for L2 comprehension; if vocabulary knowledge is accessed faster, more attentional resources can be allocated to other levels of language processing (e.g., pragmatic and sociolinguistic levels, Segalowitz, 2005).

Purpose

Only a handful of studies have investigated how quality of vocabulary knowledge is associated with L2 listening comprehension. Although Stæhr (2009) found that both vocabulary size and depth are associated with L2 listening comprehension, a potential issue relates to the modality of the vocabulary knowledge tests. Stæhr used orthographic vocabulary size and depth tests. However, listening comprehension is an auditory task. Since an L2 learner's auditory and orthographic vocabulary sizes can be different (Milton & Hopkins, 2006), there is a need to investigate how L2 listening comprehension is associated with L2 auditory vocabulary knowledge.

In addition to size and depth, the relationship between vocabulary fluency and L2 listening comprehension has not yet been investigated. The current study attempts to fill this gap by seeking to answer the following questions:

1. How is L2 listening comprehension associated with auditory vocabulary size and depth?
2. How is L2 listening comprehension related to auditory vocabulary fluency?

Methodology

Participants

There were 290 participants in the current study. All participants were first year university students. There were 228 females and 62 males. All participants took regular English classes in their secondary school in China. In other words, all participants had learned English for at least 6 years before they joined university. All of them were required to take English courses regularly during each term of their 4-year study in university.

Instruments

Auditory vocabulary size test. The Peabody Picture Vocabulary Test (PPVT) 4th edition (Dunn & Dunn, 2007) was used to measure auditory vocabulary size. The PPVT has been proved to be a reliable test that measures auditory receptive vocabulary size (Dunn & Dunn, 2007). During this test, participants heard recordings of individual words. For each word they heard, participants were required to select a picture (from a panel of four pictures shown on a computer screen) to match the word. Since the test was not timed, participants were allowed to control the flow of the test and they were allowed to replay the recordings. The scoring of the test was relatively straightforward. Each correct response received one point. The final score was calculated as the total number of correct answers.

Auditory vocabulary depth test (AVDT). The auditory vocabulary depth test (AVDT) in the Word Associates Test (WAT) format originally developed by Read (1993) is viewed as a measure of vocabulary depth. Read (2000) suggested that the WAT has a high reliability. A number of studies have reported that the test has a high reliability (e.g. Qian, 2002; Stæhr, 2009). In a more recent validation study by Schmitt, Ching, and Garras (2011), the WAT was found to be a valid measure of vocabulary depth.

The AVDT in the current study was adopted from Qian (2002). There are 40 blocks in the test. Each block consists of one target word (an adjective) followed by eight options (see Example 1 below). For any one block, a total of four out of eight options are associated with the target adjective. One to three options (in the left box) can be synonymous to the target word and one to three options (in the right box) may collocate with the target word.

Example 1

Sound

logical	healthy	bold	solid	snow	temperature	sleep	dance
---------	---------	------	-------	------	-------------	-------	-------

Presentation of the stimuli. In order to measure auditory vocabulary depth, we revised the test so that stimuli were presented orally. A native English speaker read and recorded all the target words and the options. For each block, participants heard the recording of a target word (e.g., sound) first. A screen was also displayed with eight checkboxes and eight buttons (see Example 2 below). Each button played a recording of an option (without showing the option on the screen). Participants were required to choose four options that might be associated with the target word. Participants were allowed to replay the recordings of the target words and options.

Example 2

Item 1							
<input type="checkbox"/>	A	<input type="checkbox"/>	B	<input type="checkbox"/>	E	<input type="checkbox"/>	F
<input type="checkbox"/>	C	<input type="checkbox"/>	D	<input type="checkbox"/>	G	<input type="checkbox"/>	H

Scoring. Scoring of the AVDT followed Schmitt, Ning, et al.'s (2011) recommendation. Schmitt et al. suggested two reliable scoring approaches. The first approach is known as the “all-but-nothing” approach. This approach only counts responses that hit all the correct answers. Although the all-but-nothing approach is effective in reducing guessing, it does not take into account partial knowledge. The second approach

counts responses that hit more than two correct answers. We adopted the second approach in the current study. Responses that hit all four correct answers received four points as these responses demonstrated full knowledge. Responses that hit three correct answers received two points and responses that hit two correct answers received one point as these responses demonstrated partial knowledge. Responses that hit only one or zero correct answers received zero points.

Auditory vocabulary fluency test (AVFT).

Stimuli. Auditory vocabulary fluency measured the speed of lexical access using the auditory vocabulary fluency test (AVFT) that was in the Yes/No test format. In a Yes/No test, participants were required to make fast judgments on whether the words they heard were real words or not by pressing the F or J button (F for yes and J for no).

There were a total of 160 stimuli in the AVFT, including 100 real words and 60 pseudowords. The 100 real words were vocabulary items chosen from the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001). Among the 100 words, 20 words are from the most frequent 5,000 words (extracted from Corpus of Contemporary American English, Davies, 2008-2017). Another 20 words were chosen from the Academic Word List (Coxhead, 2000).

The 60 pseudowords were created by the pseudowords generator (Keuleers & Brysbaert, 2010), controlling for word length and number of syllables so that these pseudowords were similar to the target real words (e.g., *hurked, brye, lorched, loof, steath, garced*). Pseudowords make up 37.5% of all stimuli in the AVFT. Although there is no consensus in the literature on how many pseudowords should be included in a Yes/No test of this type, it is recommended that a typical percentage falls between 25% and 50% (Beeckmans, Eyckmans, Janssens, Dufranne, & Van de Velde, 2001). Meara (as cited in Beeckmans et al., 2001) suggested that 33% is a reasonable percentage. Pellicer-Sánchez and Schmitt (2012) used 29%.

Presentation of the stimuli. All the words in the AVFT were read by a native English speaker and recorded. The 180 words were randomly played one by one. After hearing a word, the participants needed to judge whether the word was a real word or not by pressing the appropriate buttons. If there was no response after 3,000 milliseconds (ms), the next stimulus was played. Each word was played only once. Following common practice in psycholinguistics, we eliminated responses with a reaction time (RT) either shorter than 500 ms or longer than 2,500 ms. Vocabulary fluency was computed using the mean RT of hits (a hit is a Yes response to a real word).

Listening component of an International English Language Testing System (IELTS) practice test. The listening component of an International English Language Testing System (IELTS) practice test was selected to measure listening comprehension as

the IELTS test has been proven to be a valid and reliable test to measure listening proficiency (Clapham & Alderson, 1997; Davies, 2008). The listening test had four parts with 40 questions, including eight multiple choice questions and 32 fill-in-the-blank questions. Part I was a recording of a conversation between a customer and a salesperson in a car showroom. Part II was a talk given to graduate students by an admissions officer in a British university. Part III was a conversation between two students about a course they took. Part IV was a presentation given by a student concerning household waste recycling.

Procedure

The participants took part in the study 3 weeks after they joined the university. All the tests were administered on the same day in computer labs. The tests were completed on computers. Since there were not enough computers to accommodate all participants at the same time, participants were divided into two groups to take the tests following the same procedure: the AVFT was first administered, followed by the PPVT and the AVDT. The time to finish all three tests was roughly 70 minutes. After the three tests, participants took a 20-minute break and then completed the IELTS listening practice test. The listening test lasted about 45 minutes.

Results

Descriptive Statistics

In this section we present descriptive statistics of the the four tests (see Table 1 below), including reliability, mean, and standard deviation. The reliability of the IELTS listening test (Cronbach's alpha) was .85. The PPVT had a reliability of .88. The AVDT and the AVFT also had good reliability of .90 and .91.

The mean score of the IELTS listening test was 22.4 ($SD = 4.13$). The PPVT had a mean score of 86.9 points with a standard deviation of 11.4. The AVDT had a mean score of 73.9 points with a standard deviation of 21.0 points. Vocabulary fluency was measured by RT. The mean RT of the AVFT was 952 ms, with a standard deviation of 200 ms. A smaller RT represented a greater fluency level because it required less time to recognize a word.

Table 1
Descriptive Statistics

Test	Reliability	<i>M</i>	<i>SD</i>
PPVT	.88	86.9	11.4
ADVT	.90	73.9	21.0
AVFT	.91	952	200
Listening test	.85	22.4	4.13

Note. PPVT = Peabody Picture Vocabulary Test (the auditory vocabulary size test); ADVT = auditory depth of vocabulary knowledge test; AVFT = auditory vocabulary fluency test.

The Structural Equation Model (SEM)

The structural equation modeling technique was used to investigate how the three dimensions of vocabulary knowledge are associated with L2 listening comprehension. To construct the model, AMOS 16 was used. As shown in Figure 1, the structural equation model (SEM) is made up of seven observed variables and one latent variable (i.e., variables that are not directly observable). The observed variables, represented by rectangles in Figure 1, include the IELTS listening score (listening), the AVST score (size), the AVDT score (depth), and the AVFT score (RT) at four frequency levels (Fluency-2000, Fluency-3000, Fluency-AWL, and Fluency-5000). The latent variable fluency, represented by the ellipse in Figure 1, is measured by the reaction time at four frequency levels. Double-headed arrows estimate the correlation between variables (size ↔ fluency; size ↔ depth; depth ↔ fluency). Single-headed arrows (size → listening; fluency → listening; depth → listening) evaluate the impacts of the three dimensions of vocabulary knowledge on L2 listening comprehension.

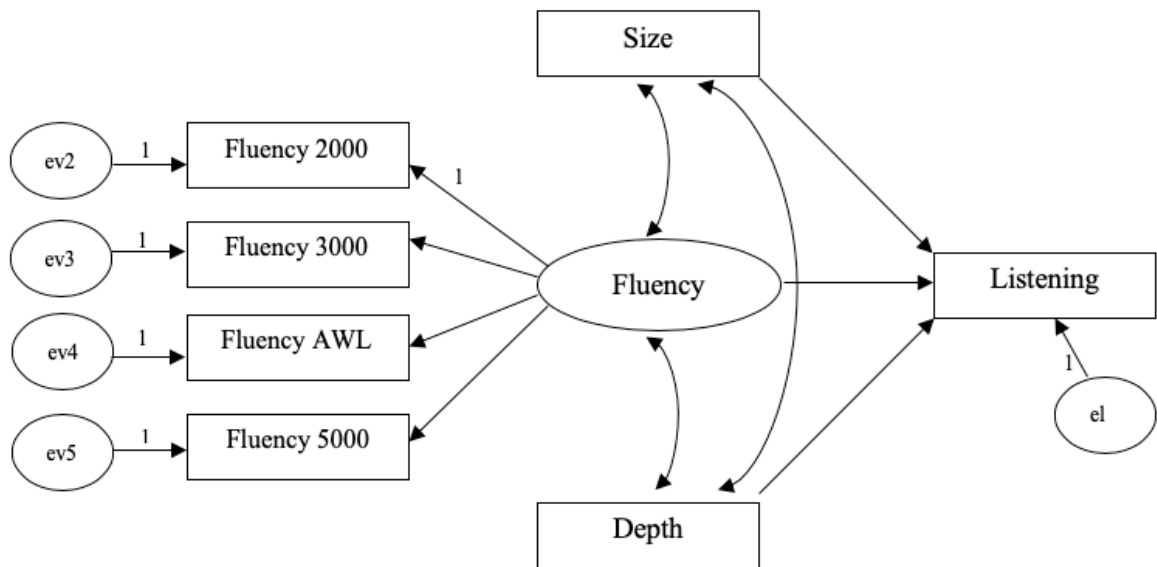


Figure 1. The Structural Equation Model (SEM).

Testing Model Assumptions

Collinearity among observed variables was inspected. Table 2 shows the collinearity estimates. Since all variance inflation factor (VIF) statistics were smaller than 10 (Myers, 1990) and all tolerance statistics were bigger than 0.2 (Menard, 1995), we believe that multi-collinearity did not present in the model.

Table 2
Multicollinearity Diagnostics

Variable	Tolerance	Variance Inflation Factor (VIF)
Size	.63	1.58
Depth	.71	1.41
Fluency 2000	.54	1.84
Fluency 3000	.38	2.67
Fluency AWL	.42	2.40
Fluency 5000	.63	1.59

One important assumption for SEM analysis is that data used in a structural equation model is multivariate normal (Byrne, 2001). In order to check whether this assumption was accurate, we calculated the multivariate kurtosis statistics to evaluate the multivariate distribution of our samples. The value of the normalized estimate of multivariate kurtosis (Mardia, 1970) is 14.6. This large value (i.e., large being defined as an absolute value greater than 5, Bentler, 2005) indicates that there was a significant kurtosis,

a sign of nonnormality in the dataset (Bentler, 2005). One way to address this issue is to use an asymptotic distribution-free (ADF) method for parameter estimation (Browne, 1984). It should be noted that the ADF estimation cannot be trusted if the sample size is not 10 times greater than the total number of estimated parameters (Raykov & Marcoulides, 2000). The sample size of the current study is 290 and the total number of estimated parameters in the SEM is 24. Since our sample size is 10 times larger than the number of estimated parameters, we believe that the result of the ADF estimation can be trusted.

Model fit and model evaluation. Commonly used model fit statistics of the model are given in Table 3. Both the chi-square/degrees of freedom ratio (χ^2/df) and the root mean square error of approximation (RMSEA) reached the acceptable threshold (Byrne, 2001; Hu & Bentler, 1999). The comparative fit index (CFI) is slightly below the acceptable level but it is approaching the threshold. Given the strength of the fit indices, the result of the model can be safely interpreted (Byrne, 2001).

Table 3
Commonly Used Fit Indices

Index	χ^2/df	CFI	RMSEA
Acceptable Level	< 3	> .95	< .08
The SEM	3.05	.92	.08

Note. df = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation.

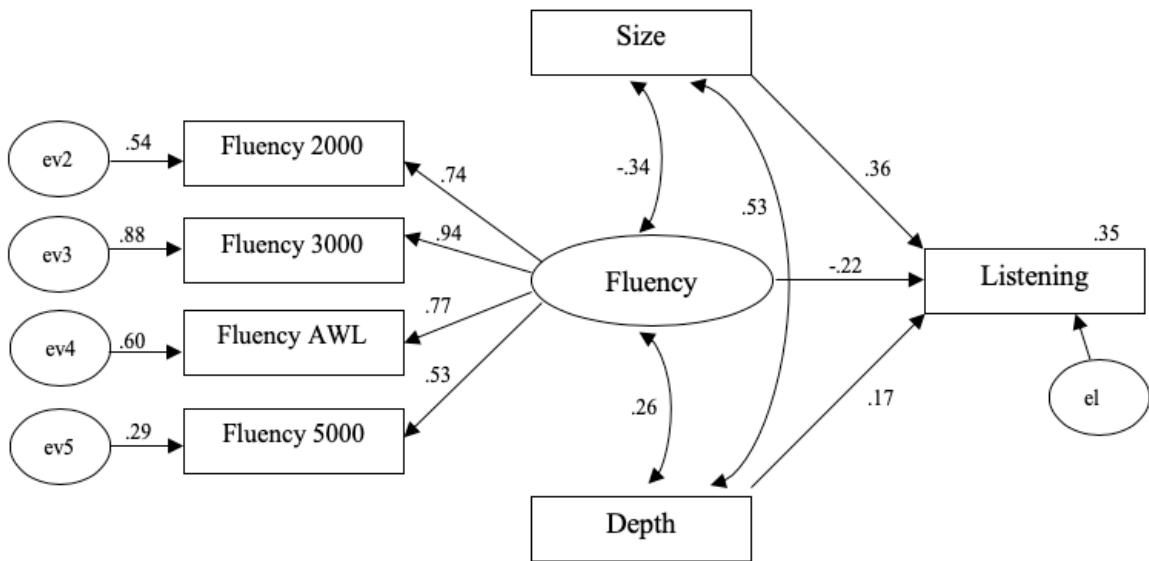


Figure 2. Estimation result of the structural equation model (SEM).

The SEM evaluated the relationship between the three dimensions of vocabulary knowledge (Figure 2). Table 4 summarizes their correlation coefficients. Vocabulary size and vocabulary depth has a correlation of .53 ($p < .01$). Vocabulary size and fluency has a negative and significant correlation of -.34 ($p < .01$). It should be pointed out that a smaller vocabulary fluency score represents a greater fluency level. This result suggests that participants with a bigger vocabulary size also tended to have a greater vocabulary depth and fluency. Vocabulary depth and vocabulary fluency has a moderate and negative correlation ($-.26, p < .01$); a higher level of vocabulary depth is associated with a greater vocabulary fluency.

Table 4
Correlations Between Size, Depth, and Fluency

	Size	Depth
Depth	.53**	
Fluency	-.34**	-.26**

Note. ** significant at .01 level.

Regression coefficients of listening comprehension on the three dimensions of vocabulary knowledge are summarized in Table 5. Vocabulary size had a predictive power of .36 ($p < .01$) and vocabulary depth had a predictive power of .17 ($p < .01$) over L2 listening comprehension. Vocabulary fluency also had a significant predictive power of -.22 ($p < .01$) over L2 listening. In total, the three dimensions of vocabulary knowledge account for 35% of the variance in L2 listening comprehension.

Table 5
Regression Coefficients

Predicted Variable	Size	Depth	Fluency
Second language (L2) listening comprehension	.36**	.17**	-.22**

Note. ** significant at .01 level.

Discussion

The recent study by Vandergrift and Baker (2015) confirmed that quantity of L2 vocabulary (i.e., vocabulary size) is directly related to L2 listening comprehension. Stæhr (2009) was the first to investigate how L2 listening comprehension is associated with quantity (size) and quality (depth) of vocabulary knowledge. There are two major differences between the current study and the above two studies. The current study used auditory vocabulary size and depth to predict listening comprehension, whereas Stæhr's study used orthographic vocabulary size and depth as predictors. Since listening comprehension depends largely on auditory information, we believe that it is more

appropriate to use an auditory form of vocabulary size and depth as the predictors. The second major difference is that auditory vocabulary fluency is included in the current study to predict L2 listening comprehension.

L2 Vocabulary Size is Most Important for L2 Listening Comprehension

The current study provides additional evidence to support the role of L2 vocabulary size in L2 listening. We found that vocabulary size has a significant predictive power (.34, $p < .01$) over L2 listening performance. Participants with a bigger vocabulary size tend to have a better performance in listening comprehension. A larger auditory vocabulary size offers a higher coverage of listening discourses. A higher coverage facilitates listening comprehension (Adolphs & Schmitt, 2003; Stæhr, 2009). It should be noted that the predictive power of vocabulary size on L2 listening comprehension in the current study is weaker when compared to previous studies: .49 in Vandergrift and Baker (2015) and .70 in Stæhr (2009). Various factors may contribute to the differences. First, the correlation between L2 listening comprehension and L2 vocabulary knowledge can vary between participants. In fact, the correlation varies by as much as 12% among the two subgroups of the participants in Vandergrift and Baker's (2015) study. Another explanation is that the current study includes two additional predictors (i.e., depth and fluency) in the structural equation model, which may reduce the predictive power of L2 vocabulary size.

Compared to vocabulary depth and fluency, L2 vocabulary size plays the most important role in predicting L2 listening comprehension. Indeed, our model shows that vocabulary size explains the largest proportion of variance in L2 listening comprehension. In the case of L2 comprehension, quantity of vocabulary seems to outweigh quality of vocabulary.

Vocabulary Depth Benefits L2 Listening Comprehension

The structural equation model shows that L2 vocabulary depth is a significant predictor of L2 listening comprehension. The regression load of listening comprehension on vocabulary depth was .17 ($p < .01$). This finding is consistent with Stæhr (2009), who also found that greater vocabulary depth is associated with better listening comprehension. Knowing individual word items better improves listening comprehension.

L2 Vocabulary Fluency Contributes to L2 Listening Comprehension

Vocabulary fluency has a significant predictive power over L2 listening comprehension, suggesting that faster lexical access is beneficial to listening comprehension. As listening discourse is presented in sequential order, listening comprehension can be quite demanding on vocabulary fluency. A fast retrieval of vocabulary form and meaning is desirable for listening comprehension. In other words, a

higher level of vocabulary fluency allows listeners to keep track of the oral discourse. During listening, if the meaning of an oral vocabulary item is not retrieved fast enough, comprehension of subsequent elements of the discourse may be affected since cognitive resources are still focusing on meaning retrieval of the vocabulary item in question.

Pedagogical Implications

Nation (2001) argued that activities involving explicit vocabulary teaching and learning can enhance listening comprehension. We believe that vocabulary learning activities should focus first on enhancing vocabulary size because vocabulary size explains the largest proportion of variance in listening comprehension.

Different types of vocabulary knowledge usually develop at a different pace (Schmitt, 2000). Laufer and Nation (2001) proposed that the development of vocabulary fluency may lag behind the development of vocabulary size. Zhang and Lu (2014) found supporting evidence to show that vocabulary fluency can develop at a high rate even when vocabulary size reaches a high level. When a form-meaning connection is first established, the speed of meaning retrieval may still be slow. Afterward, the speed of meaning retrieval gradually develops as a result of more encounters and practice. However, the form-meaning connection has to be established first, after which vocabulary fluency can develop. Since vocabulary fluency development must be based on meaning recognition, vocabulary size—which concerns form and meaning recognition—needs to be regarded as the foundation of other types of vocabulary knowledge.

It has been argued that form and meaning are easier to teach (Schmitt, 2010). Vocabulary size development can benefit from explicit teaching and learning (Nation, 2001). Until such time as vocabulary size reaches a high level (e.g., a vocabulary size of 8,000-9,000 word families, Nation, 2006), the focus of teaching can shift to promote other types of vocabulary knowledge.

Some aspects of vocabulary knowledge such as collocation are more difficult to teach and learn (Schmitt, 2010). However, collocational knowledge can still be enhanced through input. In fact, vocabulary learning and comprehension are mutually beneficial to each other. As Koda (2005) argued, L2 comprehension and vocabulary knowledge are dynamically related to each other. Vocabulary knowledge plays a crucial role in understanding L2 discourse. In return, reading and listening can facilitate vocabulary knowledge acquisition (e.g. Krashen, 1989; Pigada & Schmitt, 2006; Saragi, Nation, & Meister, 1978). Therefore, learners should be encouraged to listen and read extensively so as to enhance vocabulary size and depth, which improve overall L2 comprehension.

Given that vocabulary fluency is associated with listening comprehension, listening instruction should incorporate vocabulary fluency training. Nation (2007) suggested that pedagogical activities designed to promote vocabulary fluency development should include large quantities of language input and output that are familiar to learners. Some specific activities for promoting vocabulary fluency development include reading graded readers

and listening to stories (Nation, 1990). With sufficient training that enhances vocabulary fluency, L2 listening comprehension can also be improved.

Limitations

One limitation of the current study is related to the AVDT. First, the change of the test format (from visual to auditory) might have affected how the participants responded. Further investigation is needed to evaluate the impact of the change in test format. Second, although it has been shown that the AVDT is a reliable measure of vocabulary depth (Schmitt, Ning, et al., 2011), this test only measures paradigmatic and collocation knowledge. It should be pointed out that vocabulary depth is not limited to these two types of knowledge. Other aspects of vocabulary knowledge such as constraints and grammatical functions are also important (Nation, 2001). We recommend that future studies to investigate the role of vocabulary depth include other aspects of vocabulary knowledge.

Conclusion

The current study investigated how three dimensions of vocabulary knowledge are associated with L2 listening comprehension. We used auditory tests to measure auditory vocabulary size, depth, and fluency. We found that all three dimensions of vocabulary knowledge are significant predictors of L2 listening comprehension. Among the three dimensions of vocabulary knowledge, vocabulary size plays the most important role in L2 listening. In addition to vocabulary size, vocabulary fluency is also important given that these two dimensions of vocabulary knowledge explain a significant proportion of variance in L2 listening comprehension.

Since there was only a moderate correlation between vocabulary size and depth, we believe that vocabulary size and depth are two different constructs of vocabulary knowledge (Qian, 2002; Schmitt, 2014), which contradicts Vermeer's (2001) argument that vocabulary size and depth belong to a same construct. Although such a finding is not the focus of the current study, the moderate correlation between size and depth is a good indicator to suggest that teachers and learners should not pay attention only to vocabulary size. We believe that vocabulary depth also deserves attention.

Correspondence should be addressed to Xian Zhang.

Email: xian.zhang@unt.edu

Acknowledgement

This study was supported by the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities in P.R. China (Project No. 17JJD740004) and the National Key Research Center for Linguistics and Applied Linguistics of Guangdong University of Foreign Studies.

References

- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24, 425-438.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M., & Van de Velde, H. (2001). Examining the Yes/No test: Some methodological issues in theory and practice. *Language Testing*, 18(3), 235-274.
- Bentler, P. M. (2005). *EQS 6 Structural equations program manual*. Encino, CA: Multivariate Software Inc.
- Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics*, 16(1), 15-34.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37(1), 62-83.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. London: Erlbaum.
- Clapham, C. M., & Alderson, J. C. (1997). *Constructing and trialing the IELTS test: IELTS research report 3*. Cambridge, UK: The British Council, University of Cambridge Local Examinations Syndicate and International Development Program of Australian University and Colleges.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Davies, A. (2008). *Assessing academic English: Testing English proficiency 1950-89: The IELTS solution*. Cambridge, UK: Cambridge University Press.
- Davies, M. (2008-2017). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-2017*. Retrieved from <https://corpus.byu.edu/coca/>.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. Minneapolis, MN: Pearson Assessments.
- de Bot, K., Paribakht, T. S., & Wesche, M. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from ESL reading. *Studies in Second Language Acquisition*, 19(3), 309-329.

- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223.
- Hu, L-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge, UK: Cambridge University Press.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal, 73*(4), 440-464.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627-633.
- 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). London: Macmillan.
- Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA Yearbook, 1*(1), 7-28.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15-30.
- Graham, S., & Macaro, E. (2008). Strategy instruction in listening for lower-intermediate learners of French. *Language Learning, 58*(4), 747-783.
- Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519-530.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*(2), 142-154.
- Menard, S. (1995). *Applied logistic regression analysis. Quantitative applications in the social sciences; Sage university papers no. 07-106* Thousand Oaks, CA: Sage.
- Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *Canadian Modern Language Review, 63*(1), 127-147.
- Myers, R. (1990). *Classical and modern regression with applications (2nd ed.)*. Boston, MA: Duxbury.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. New York: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82.
- Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching, 1*(1), 2-12.

- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3), 513-536.
- Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring Yes-No vocabulary tests: Reaction time vs. nonword approaches. *Language Testing*, 29(4), 489-509.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18, 1-28.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371.
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.
- Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling*. Mahwah, NJ: Erlbaum.
- Saragi, P., Nation, P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6(2), 72-78.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, UK: Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913-951.
- Schmitt, N., Ching, J. W., & Garras, J. (2011). The Word Associates Format: Validation evidence. *Language Testing*, 28(1), 105-126.
- Schmitt, N., Jiang, X., and Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95(1), 26-43.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.
- Segalowitz, N.S. (2005). Automaticity and second languages. In C. J. Doughty & M. H. Long, (Eds.), *The handbook of second language acquisition* (pp. 382-408). Malden, MA: Blackwell.
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(4), 577-607.
- Vandergrift, L. (2005). Relationships among motivation orientations, metacognitive awareness and proficiency in L2 listening. *Applied Linguistics*, 26(1), 70-89.
- Vandergrift, L. (2006). Second language listening: Listening ability or language proficiency? *The Modern Language Journal*, 90(1), 6-18.
- Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, 65(2), 390-416.
- Vandergrift, L., & Goh, C. (2012). *Teaching and learning second language listening: Metacognition in action*. New York: Routledge.

- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22(2), 217-234.
- Wesche, M., & Paribakht, T. S. (1996). Assessing L2 vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, 53(1), 13-40.
- Zhang, X., & Lu, X. (2014). A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. *Applied Linguistics*, 35(3), 283-304.