



The relationship between L2 vocabulary knowledge and listening proficiency:
The mediating effect of vocabulary fluency

Yanli Tong^a , Zuwati Hasim^{b1} , Huzaina Abdul Halim^c 

^a University of Malaya, Kuala Lumpur, Malaysia

Ningde Normal University, Ningde, China

^b University of Malaya, Kuala Lumpur, Malaysia

^c University of Malaya, Kuala Lumpur, Malaysia

APA Citation:

Tong, Y. L., Hasim, Z. & Abdul Halim, H. (2022). The relationship between L2 vocabulary knowledge and listening proficiency: The mediating effect of vocabulary fluency. *Journal of Language and Linguistic Studies*, 18(1), 427-446. Doi: 10.52462/jlls.192

Submission Date:05/08/2021

Acceptance Date:18/10/2021

Abstract

Based on 312 samples, this study investigated the relationship between two dimensions of L2 vocabulary knowledge (VK) and listening comprehension (LC) by comparing written and oral vocabulary measurements and further examined the mediating effect of L2 vocabulary fluency (VF) by using Bootstrapping and Monte Carlo Method. Results show that: 1) the regression coefficient from written receptive breadth and depth of VK to LC were $\beta = .29^{***}$ and $\beta = .22^{***}$; 2) the regression coefficient from oral receptive breadth and depth of VK to LC were $\beta = .34^{***}$ and $\beta = .31^{***}$; 3) the variance explanatory power of LC reached 64%; 4) VF played a partial mediating role between VK and LC; 5) on the condition of the mediating effect of VF, regression coefficient of written receptive vocabulary breadth to LC increased by 13%, reaching $\beta = .42^{***}$, that of oral one to LC raised by 17%, reaching $\beta = .51^{***}$; 6) regression coefficient of written receptive vocabulary depth to LC increased by 19%, reaching $\beta = .41^{***}$, that of oral one to LC raised by 15%, reaching $\beta = .46^{***}$. Finally, we discussed respective importance of different dimensions of VK in LC. The conclusion of the study was of significance to teaching of L2 VK and LC.

Keywords: written receptive vocabulary breadth and depth; oral receptive vocabulary breadth and depth; listening comprehension; vocabulary fluency; mediating effect

1. Introduction

Vocabulary Knowledge (VK) plays a crucial role in the process of language development (Laufer, 1998; Nation, 2013; Read, 2004). Many scholars emphasize the importance for L2 learners to command sufficient VK in connection with different language skills (Laufer, 1998; Nation, 2013; Schmitt, 2010). Since VK is an important predictor of Language Proficiency (LP), research on the relationship between VK and LP has attracted many scholars locally and globally. Although previous research on the relationship between VK and reading comprehension (RC) (Cheng & Matthews, 2018; Nation, 2001; Qian, 2002, 2004) yet studies on the correlation between VK and LC are rather limited (Cheng & Matthews, 2018; Stæhr, 2008, 2009). For most studies that examined the correlation

¹ Corresponding author

E-mail address: zuwati_hasim@um.edu.my

between VK and listening, the focus was mainly on the relationship between one dimension of VK and multiple language skills (Stæhr, 2008), or between multiple dimensions of VK and single language skill (Qian, 2002, 2004). Based on WOS database search of studies on VK that examined the breadth and depth in LC, it was found there are seven of studies (Stæhr, 2008; Teng, 2014). However, in China, there was only 2 studies that examined the breadth and depth in LC (Cheng & Matthews, 2018; Zhang, 2011). Especially, there was no empirical study that examined relative significance of two dimensions of VK covering breadth and depth in LC by comparing written and oral vocabulary measurements based on a relatively large sample pool. Furthermore, no scholars used Vocabulary Fluency (VF) as a mediating variable to study the correlation between VK and LC. To further understand the practical association between VK and LP, a more comprehensive VK measures should be employed to explore its strength of VK in LC. Hence, for this study, based on the conceptual framework of Figure 1, we investigate the relationship between L2 VK and LC through several listening tests and vocabulary measurements which require both written and oral responses from the students. By using the latest vocabulary test formats to measure the participants' lexical level and some constructs are developed to measure the effects and predictability of VK in LP based on a large group of L2 learners in China. The mediating effect of VF on VK and LC is also investigated.

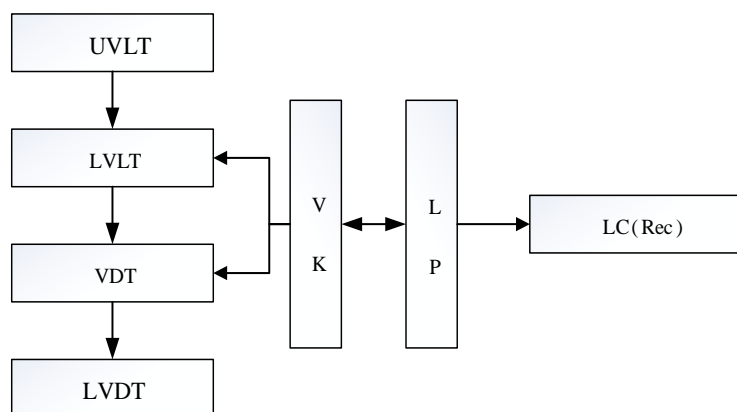


Figure 1. A conceptual framework

Note: UVLT=written receptive vocabulary breadth; LVL=listening receptive vocabulary breadth; VDT=written receptive vocabulary depth; LVDT=listening receptive vocabulary depth; VK=vocabulary knowledge; LP=language proficiency; LC=language comprehension

2. Theoretical Backgrounds

2.1 Concept Definition

2.1.1 Vocabulary Breadth

Vocabulary breadth is regarded as the size of words that L2 learners know at a certain level of LP (Nation, 2001). That is, learners acquire words' form-meaning knowledge and can easily recognize the word in the process of "listening" and "reading". According to Meara (1996), vocabulary breadth was the basic dimension of learners' vocabulary ability in which language learners with a large vocabulary size could use language more expertly as compared to those with smaller vocabulary size. Schmitt (2008) asserted that 95% of vocabulary recognition was a guarantee for LC. Also, Staehr (2009) pointed out that for academic listening, L2 learners need to master 98% of listening text vocabulary. On another note, Webb. et.al. (2017) claimed that approximate 65–85% of oral and written English were accounted for by the most frequent 1000-word and 5000-word families, and those were the most important and useful for L2 learners, that “represents the greatest range in vocabulary learning” (Webb & Sasao, 2013, p. 266). Based on this view, they developed an Updated Vocabulary Levels Test

(UVLT). McLean, Kramer, and Beglar (2015) also have developed vocabulary achievement instrument namely the Listening Vocabulary Levels Test (LVLT) to measure aural lexicons and the New Vocabulary Levels Test (NVLT) to measure lexical knowledge.

In view of the fact that breadth of VK is at the core of language ability, it is important to provide a reliable measurement instrument for assessing vocabulary breadth. It is identified that different researchers examined the learners' vocabulary breadth by using different instruments. There are various vocabulary size test that are available online (see <https://my.vocabularysize.com/>). Schmitt et al (2001) mentioned that “the Vocabulary Levels Test is designed to give an estimate of vocabulary size for second language (L2) learners of general or academic English” (p. 55). The most well-known instrument of vocabulary breadth is “the Vocabulary Levels Test” (VLT) designed by Paul Nation (1983, 1990), which is used to test L2 learners' recognition and understanding of VK. Another widely used one is “the Vocabulary Levels Test” (VLT) revised by Schmitt et al (2001). An Updated Vocabulary Levels Test (UVLT) proposed by Webb et al (2017) and the New Vocabulary Levels Test (NVLT) and Listening Vocabulary Levels Test (LVLT) developed by McLean et al (2015) focus on the first 1000-word frequency level was included in the testing and the target words were chosen from new BNC/COCA lists which typically represented current English used by native speakers today (Webb & Sasao, 2013, p:267). For the purpose of this study, we have chosen the UVLT developed by Webb et al and LVLT designed by McLean et al to check the participants' VK in current study. Choosing these two testing instruments are more reasonable for measuring the VK of the students in China since English is regarded as a foreign language and rarely used, thus it is reasonable to test the VK of the student participants from the 1000-word level.

It is noted that the Listening Vocabulary Levels Test (LVLT) was taken into consideration because the purpose of this current study was to explore the association between VK and LP, requiring to continue efficiently processing incoming information which stressed the role of processing speed and automaticity. Therefore, it is obviously necessary to measure VK by LVLT.

2.1.2 Vocabulary Depth

Just as its name implies, vocabulary depth reflects the quality of VK, which means how well learners understand words and whether they can use them correctly (Qian, 2002; Zhang, 2020). However, different scholars have different understanding on vocabulary depth. Henriksen (1999) proposed three aspects of vocabulary depth including partial-precise knowledge, depth of knowledge and receptive-productive based on vocabulary measurement methods. Read (2000) supported that Henriksen's three dimensions and asserted that they were meaningful because it helped us to better understand what the concept of quality of VK meant and what aspects of construct should be checked in the study. Read (2004), in his work, also proposed three dimensions on vocabulary depth: Precision of meaning, Comprehensive word knowledge and Network knowledge.

Webb (2013) indicated that the most comprehensive explanation about vocabulary depth was the details of knowing a word described by Nation (2001, 2013). Nation (2013, p. 49) asserted that learning a word requires not only to understand form and meaning, but also its usage (grammatical function, collocation, constraints on use). In other words, vocabulary depth included all vocabulary features. Generally, the framework of multiple components of VK has become the theoretical basis for scholars to study vocabulary depth. Schmitt (2014) argued that the best depth should include all aspects of VK, and the individual elements (e.g. polysemous meaning senses, derivative forms, collocation) should also contribute to vocabulary depth.

Nevertheless, both Read (2004) and Schmitt (2014) agree that there was no measure of depth as “whatever conceptualization or measure is used, it will only ever tap into limited facets of the overall quality of understanding of a lexical item. The size-depth relationship depends on various factors such

as the size of the learner's lexicon, the frequency level of the target words measured, and the learner's L1" (Schmitt, 2014, p.941). In this case, it is best to concentrate on the measures of the quality for VK (Read, 2004; Milton, 2009).

Also in view of the fact that vocabulary depth plays a great role for developing language ability, it is crucial to provide a reliable measurement instrument for assessing vocabulary depth. The research instrument which has been widely accepted for checking depth of VK is Word Associates Test (WAT) developed by Read (1993) and revised by Read (1998).

2.1.3 Reception and Production

VK is divided into two aspects: receptivity and productivity. In short, the former refers to understanding in theory and the latter means application in practice. Receptive vocabulary forms the number of vocabularies in the learners' mental lexical repertoire and needs to be further developed into productive vocabulary to be used for communication. The development of vocabulary from receptivity to productivity can be seen as a continuum, beginning with familiarity with the lexicon and ending with the capability to produce it freely. The process of making progress on this continuum is the development of the quality of learners' VK, also known as vocabulary depth development (Qian, 1999, 2002). Among four basic skills of LP, reading and listening belong to the receptive skills, and writing and speaking belong to the productive skills (Nation, 2013).

2.2 Listening Comprehension

LC is a basic language skill which is not easy to acquire skillfully. To successfully complete the listening task and accurately guess the words and sentences' meaning from the context, language learners need to have at least 95% vocabulary of listening materials (Nation, 2013). Staehr (2009) probed the role of VK in LC by using VLT and WAT to test vocabulary breadth and depth with 115 advanced Danish EFL learners and also pointed out that 89% lexicon coverage was necessary for dealing with spoken texts that made up the listening testing. However, the fleeting nature of spoken language requires the activation of short-term working memory and requires VK to play different roles in the listening process. This shows that learners with a similar breadth of VK have significant differences on LC scores for the same listening materials (Staehr, 2009; Van Zeland & Schmitt, 2013) since their general language proficiency levels can be different.

Staehr (2009) claimed that the correlation between vocabulary breadth and LC was .70, the correlation between vocabulary depth and LC was .65, and the correlation between breadth and depth was .80, and the variance explanatory power of LC reached 51% ($R^2 = .51$). Staehr (2009) declared that vocabulary breadth played more significant role in LC than depth does. However, Dabbagh (2016) conducted a survey of the LC on 73 college students in Iran, and received different results from Staehr's. Dabbagh found that only vocabulary depth had a significant predictive effect on learners' LC, and vocabulary breadth was not a powerful predictor of LC.

Cheng and Matthews (2018) recruited 250 Chinese college students as the subjects for their research to explore the relationship between VK and reading and listening using three measures of VK including productive phonemic VK (ProPhon), receptive orthographic VK (RecOrth), and productive orthographic VK (ProOrth). The results indicated that productive phonemic VK (ProPhon) had the strongest correlation with LC .71.

The differences and conflicts of these research results denote the importance of vocabulary breadth and depth for LC, that their correlation strength and predictive power need to be further verified. Moreover, no studies have been found to check the strength of correlation between VK and LC by comparing written and oral vocabulary measurements.

2.3 Vocabulary Fluency

Vocabulary fluency is the capability to listen or read a L2 texts and to speak or write in L2 accurately, smoothly and immediately. VF is absolutely significant as it provides a bridge between word recognition and comprehension. Nation (2013) pointed out that a well-balanced language course should consist of four major aspects, they are ‘meaning-focused input’, ‘meaning-focused output’, ‘form-focused instruction’ and ‘fluency development’. If there is no strong fluency strand in a course, then the learning process in the other three strands would be affected. Thus, it is believed that not less than 25 per cent of the class time should be given to achieve fluency, and if it is missing, the learning of the other three elements will not be achieved successfully. This indicates that fluency is very crucial in the process of language development. Learners who have achieved fluency level could recognize and process words automatically, effortlessly and naturally instead of concentrating on decoding words. Therefore, it is crucial to develop VF for achieving a certain level of LP.

What’s more, VF is the speed of recognizing and retrieving words, which is a key factor affecting L2 learners' performance in four different kinds of language skills (Segalowitz, 2005). It is not difficult to understand that as the speed of word recognition increases, learners can concentrate on dealing with higher levels of communication (Segalowitz, 2005). In the previous studies, positive correlations between VF and reading/writing/speaking were proved by Snellings et al. (2004), Schoonen et al. (2003) and Segalowitz and Freed (2004). Hui and Godfroid (2020) highlighted the significance of processing speed and accuracy at different levels of the linguistic hierarchy for L2 listeners. However, there is no research on the mediating effect of VF in relationship between VK and LC.

It was highlighted from the literature review that firstly, VK has a strong correlation with LC and is also a powerful predictor of LC. Secondly, different dimensions of VK play different roles in LC. Third, there is no empirical data that explore the correlation between two dimensions of VK and LC by comparing written and oral vocabulary measurements based on a large group of samples. Finally, there is no study on the mediating effect of VF on the correlation between each dimension of VK and LC. With these in view, we would like to carry out a study to see if there would be similar results when VK study is conducted

2.4 Research Purpose

The aim of this current study is to investigate the relationship between L2 VK and LC through several listening tests and vocabulary measurements which require both written and oral responses from the students. Hence among the research objectives for the study are to measure the effect and predictability of VK in LP, to analyze the correlation between two dimensions and LC, and to investigate the mediating effect of VF on the association between VK and LC.

2.5 Research Questions and Hypothesis

Based on these objectives, several research questions (RQ) are developed:

RQ1. What effect can the two dimensions of L2 VK have on LC?

RQ2. To what extent LC can be predicted by L2 VK?

RQ3. Does vocabulary fluency have a mediating effect on the correlation between between VK and LC?

Based on the research questions, the following hypotheses are outlined:

H1: Receptive breadth of VK measured by written vocabulary measurement has a positive effect on LC

H2: Receptive depth of VK measured by written vocabulary measurement has a positive effect on LC

H3: Receptive breadth of VK measured by oral vocabulary measurement has a positive effect on LC

H4: Receptive depth of VK measured by oral vocabulary measurement has a positive effect on LC

H5: VF has a mediating effect on the relationship between VK and LC

3. Method

This study is quantitative in nature where AMOS24.0 was used to analyze the selected data. Then, a Structural Equation Model (SEM) was established based on the theoretical model to test the model fit and to do bi-variate correlation analysis. Details of the methodology and procedures are described in the following sections.

3.1 Subjects

The sample of participants in the present study comprised of 312 first year university students enrolled in the first semester of their study in 2019, in Ningde Normal University. There are female and male students whose average age is 20.5 year-old and none of them has oversea living or learning experiences. The selection criteria guarantees that the participants might share similar English learning experience and English proficiency level, age and educational background and have taken their University Entrance Exam for accessing to Ningde Normal University, which is a medium-ranked Chinese university located in Ningde, the medium city of Fujian province. Ranked at 465 top universities among 1243 undergraduate universities in China, Ningde Normal University is considered a very typical representative of tertiary institutions in China. so, Ningde Normal University is considered a very typical representative of tertiary institutions in China.

3.2 Instruments

To carry out the study and to answer the research questions, several testing instruments are adopted and used for the purpose of this study. A total of six instruments are used. They are: UVLT, LVLTL, VDT, LVDT, VFT and LC.

3.2.1 Measurement for Written Receptive Breadth of VK (UVLT)

The measurement tool of receptive vocabulary breadth used in this study is the Updated Vocabulary Levels Test (UVLT) that was improved by Webb et al (2017) based on the Vocabulary Level Test (VLT) designed by Nation (1983) and the revised version by Schmitt, Schmitt, & Clapham (2001). The reason why UVLT is chosen in the current study is that it includes the most frequency 1000-word families which accounts for 65-85% of spoken and written English and includes the most useful and frequent 5000-word families for learners (Webb & Nation, 2017). At the same time, items within word families in UVLT are derived from texts from current vocabulary (BNC/COCA).

UVLT is a three-line table where six items (three keys and three distractors) are presented in bold in the first line on the right and three definitions are presented in the first column on the left. The participants who choose a correct item for each definition will receive a point for each correct response in a three-line table. There are 10 three-line tables per level, so the maximum score at each level is 30. There are five frequency word levels including (1,000, 2,000, 3,000, 4,000 and 5,000) on

behalf of the most frequent 5000-word families, therefore, the maximum total score of five levels is 150. The sample items are as follows:

	ear	gold	lake	letter	office	people
information sent to people						
men and women						
place for working						

On academic vocabulary measurement, we use Academic Word List (AWL) (Coxhead, 2000) because words in AWL vary greatly in their value based on Webb. It is noted that although word items came from WAL, we still follow the UVLT’s testing format for standardization. There are 30 target words, so the maximum score at each academic level is 30. The sample items of WAL are as follows:

	benefit	labor	source	survey	percent	principle
work						
part of 100						
general idea used to guide one’s actions						

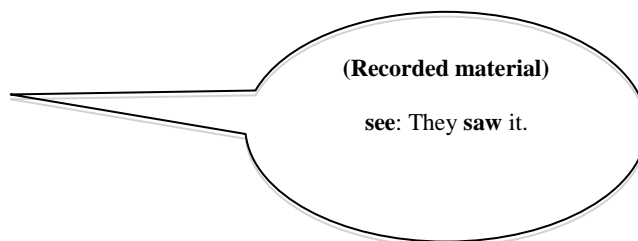
Therefore, the total score of receptive vocabulary breadth is 180. The total time limit for five levels and academic words, including the presentation of the stimulus, was 35 minutes based on a pilot study.

3.2.2 Measurement for Listening Receptive Breadth of VK (LVLV)

The measurement instrument of oral receptive vocabulary breadth used in this study is LVLV created by McLean, Kramer and Beglar (2015) (see www.lvlt.info). It was developed based on written receptive vocabulary breadth measurement instrument (NVLV), which is a new vocabulary levels test representing a receptive test of the most frequent 5,000 word families in Nation’s (2012) BNC/COCA word list and providing an instrument with greater pedagogical utility compared to previous VLV format. As Webb and Macalister (2013) stated, “mastery of the 5,000 word level may be challenging for all but advanced learners, so assessing knowledge at five most frequent levels may represent the greatest range in vocabulary learning for the majority of L2 learners” (p. 266). In addition to a thirty-item section to measure AWL knowledge, the test includes five levels with 24-item per level and together it measures the most commonly used 5000-word families. Based on a pilot study, 35 minutes were required to complete a total of 150 testing items. The sample items are as presented below.

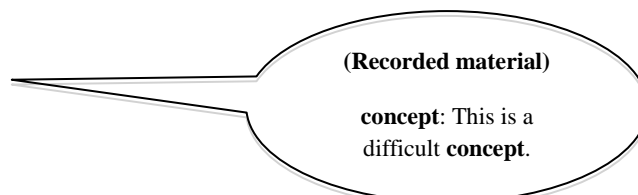
Please select the closest meaning from a, b, c, and d to match the target word you heard.

- a. cut
- b. waited for
- c. Looked at
- d. Started



Also, for the study, 5000-word level of academic words from Academic Word List (Coxhead, 2000) are also used to check the participants’ listening vocabulary ability. Below is the sample of the items:

- a. legal agreement
- b. idea about what something is
- c. way of doing things



d. a written explanation of a law

Based on a pilot study, 8 seconds is needed for each item, including the presentation of the stimulus. There are five levels with 24 items per level and 30 academic word items. The participant will receive a point for each correct response, so a total score of all items is 150 points. Each word and the sentence made by this word were recorded at a normal speed by a native male speaker of American English. The test items and audio are freely available and can be downloaded from (www.lvlt.info).

3.2.3 Measurement for Written Receptive Depth of VK (VDT)

The measurement instrument of receptive vocabulary depth used in this study is Word Associate Test (WAT) designed by Read (1993, 1998). The version designed in 1993 is only for academic vocabulary testing, resulting in uneven test results. To maintain consistency of testing, the developed version in 1998 only uses adjectives as testing items. The test form contains a target word and eight candidate words, four of which are distractors. Read (1993, 1998) proves that the instrument is reliable and valid from the perspective of testing. Below is the sample of items:

Sound

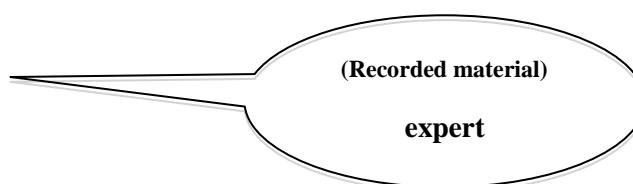
(A)logical(B)healthy(C)bold(D)solid	(E)snow(F)temperature(C)sleep(D)dance
-------------------------------------	---------------------------------------

There are eight words evenly distributed in the left and right boxes. Words in the left box refer to meaning of **sound**. Words in the right box refer to collocation of **sound**. There are forty target words and four correct keys per target word, so a total score is 160 points. Including the presentation of the stimulus, the total time limit for written VDT is 30 minutes based on a pilot study.

3.2.4 Measurement for Listening Receptive Depth of VK (LVDT)

To develop oral VDT, 40 target words which are also derived from the Word Associate Test (WAT) are recorded reading using a male native speaker of American English. The participants only heard each target word once, then chose the right answer. We changed the format slightly, in which meaning of four words in the upper box refer to the meaning of **expert** and four words in the bottom box refer to collocation of **expert**. There are forty target words and four correct keys per target word, so a total score is 160 points. The time limit for each item is 30 seconds based on a pilot study. The total time of the test is 30 minutes. The following is the sample of the test item:

(A) overseas (B) mend
(B) uniform (D) spring
(E) reinforcing (F) ideas
(F) commodity (H) market



The Listening Receptive Vocabulary Depth (LVDT) was conducted two weeks later than Receptive Vocabulary Depth (VDT) because both written and oral vocabulary depth testing items are derived from Word Associate Test. We collected the participants' response using a questionnaire to investigate if they knew that both written and oral vocabulary depth measurement have the same items after LVDT testing. The result was that 95% participants did not know that both of vocabulary depth testing were the same items and only 5% test takers said that there was something vaguely familiar about part of items. One possible reason was that on the first testing day, they took part in the 6 measurements. This might be too many for them to remember all the items given in a day. Another possible reason was that we conducted LVDT two weeks later to avoid the participants from remembering the testing items.

3.2.5 Measurement for Vocabulary Fluency (VFT)

VF measurement instrument is a dictation test taping aspects of phonetic, morphology and pragmatic of VK. The dictation test requires test takers to listen to a short passage and asks them to recognize and produce the 20 missing target words per passage according to the provided context. Because of the fleeting characteristics of listening, if test takers can produce the missing words in limited time, it reflects to some extent the speed of retrieving vocabulary which indicates perfect vocabulary fluency. The sample item is as follows:

A modern ____ wanting to ____ privately can use ____ _____. Three hundred years ago, ____ would have been a better _____. Before ____ ____ became ____ in the 1800s, ____ were ____ with no ____ _____. Privacy-minded ____ ____ instead on ____ ____ of folds, tucks, ____ and _____. a practice Jana Dambrogio at the Massachusetts Institute of Technology has dubbed “_____”.

The time limit for each passage is five minutes, including the presentation of the stimulus based on a pilot study. There are four passages with 80 missing words and the participant receives a point for each correct word, so a total score of four passages is 80 points. Because VF testing aim is to check the participants' ability of the speed of processing vocabulary information and retrieving vocabulary on controlling over the speed of the incoming speech stream, minor vocabulary spelling errors are ignored. The total time of the test is 25 minutes.

3.2.6 Measurement for Listening Comprehension

In order to check all 312 participants' L2 listening ability, we choose IELTS listening test version which contains forty task items and all test-takers are required to listen to two monologues and two dialogues which are used to check their comprehension. There are a variety of testing questions such as multiple choices, filling in forms, completing sentences and so on. After listening to four sections, participants have ten minutes to finish and check all the answers. The total time of the test was 40 minutes. And according to their self-report, they have not attended IELTS test before.

3.3 Procedure

There are six testing tasks altogether for the study: ①written receptive vocabulary breadth ②oral receptive vocabulary breadth ③written receptive vocabulary depth ④oral receptive vocabulary depth ⑤vocabulary fluency and ⑥IELTS listening. The testing tasks were conducted in 3 lecture halls. Firstly, an information sheet on the overall testing procedure was presented and explained to every participant before testing. Then, all measurements were conducted in 3 lecture halls equipped with high-quality audio devices which was used for listening testing, and we spent 70 minutes to finish the first two tests (①and②). After a short break of 10 minutes, they went through a receptive vocabulary depth testing in 30 minutes (③). After that, IELTS listening test (⑥) was administered. The whole process lasted about 2.5 hours including 5 minutes for examination introduction before testing and a break of 10 minutes between the former two tasks and the latter two tasks. Two weeks later, both ④oral VDT and ⑤VFT were conducted on the same classroom. Finally, a questionnaire with only one question was administered, which is about if they met these items before oral VDT. And two measurements were completed successfully in 55 minutes. Then, we collected and analyzed the data.

3.4 Data Analysis

All measurements were scored by getting one point for one correct answer. Before conducting SEM regression analyses, a set of data checking analyses were conducted. First of all, we check the reliability and validity of the data by applying AMOS24.0 with value of composite reliability (CR=Cronbach's α), square multiple correlations (SMC) and average of variance extracted (AVE).

Then, SEM correlation analysis was conducted. Finally, the mediating effect analyses by Bootstrapping and Monte Carlo Method were administered.

4. Results

4.1 Confirmatory Factor Analysis of Variable Reliability and Validity

The value of composite reliability (CR) is the combination of the reliability of all measurement variables, representing internal consistency of the constructs, which is equivalent to Cronbach's α . Based on the standard of CR proposed by Fornell and Larcker (1981), all CR value comply with the standard because they are higher than 0.7. Average of variance extracted (AVE) is the average of the explanatory power of calculating the latent variables to the measured variables. The higher AVE value is the greater convergence validity of the constructs is. Fornell and Larcker (1981) suggested that it had better be higher than 0.5, and 0.36 is the acceptable threshold. To obtain composite reliability and convergence validity, all variables were not be pulled correlation and the data of Table 1 was acquired by Confirmatory Factor Analysis.

Table 1. Convergence Validity

Construct	Item	Sig. test of Parameters				Std.	Item Reliability SMC	Composite Reliability CR	Convergence Validity AVE	
		Unstd.	S.E.	z-value	P					
UVLT	A1	1.000				.752	.566	.912	.635	
	A2	1.191	.085	14.003	***	.778	.605			
	A3	1.145	.088	13.046	***	.730	.533			
	A4	1.615	.100	16.169	***	.885	.783			
	A5	1.714	.112	15.358	***	.844	.712			
	A6	1.724	.122	14.080	***	.781	.610			
VDT	B1	1.000				.807	.651	.879	.646	
	B2	1.327	.086	15.353	***	.821				.674
	B3	1.372	.092	14.856	***	.797				.635
	B4	1.471	.100	14.679	***	.789				.623
LVLT	C1	1.000				.769	.591	.881	.553	
	C2	.998	.086	11.601	***	.665				.442
	C3	1.136	.088	12.881	***	.731				.534
	C4	1.367	.100	13.704	***	.774				.599
	C5	1.276	.097	13.094	***	.742				.551
	C6	1.335	.098	13.689	***	.773				.598
LVDT	D4	1.000				.729	.531	.855	.597	
	D3	1.125	.086	13.063	***	.822				.676
	D2	.914	.073	12.580	***	.781				.610
	D1	.768	.063	12.213	***	.755				.570
LC	L1	1.000				.676	.457	.866	.621	
	L2	1.323	.117	11.269	***	.713				.508
	L3	2.079	.155	13.411	***	.917				.841
	L4	2.022	.159	12.750	***	.824				.679
VF	F4	1.000				.754	.569	.815	.525	
	F3	.951	.080	11.850	***	.782				.612
	F2	.768	.072	10.713	***	.680				.462
	F1	.665	.062	10.665	***	.676				.457

Note: *** indicates that the result is significant at $p < 0.001$

As shown in Table 1, the values of non-standardized points estimation are all positive, z -value is higher than 1.96 and $P < 0.000$ which is lower than the level of significance and thus the null-hypothesis can be rejected. The standardized points estimations are factor loading which are all higher than 0.6 which indicate the items of constructs have a good reliability. SMC is the square of factor loading which are higher 0.36, also indicating that the items of constructs have a good reliability. The composite reliability is above 0.8, and the convergent validity is above 0.5, which totally meet the required standard proposed by statistic scholar (Fornell & Larcker, 1981). In order to test the discriminate validity between various variables, all variables were made to correlate. After analyzing, the data of Table 2 was obtained.

Table 2. Discriminant Validity

CONSTRUCT	AVE	VF	LC	LVDT	LVLT	VDT	UVLT
VF	.525	.725					
LC	.621	.525	.788				
LVDT	.597	.241	.571	.773			
LVLT	.553	.181	.593	.335	.744		
VDT	.646	.213	.516	.328	.260	.804	
UVLT	.635	.294	.541	.230	.290	.346	.797

As shown in Table 2, AVE square root values in bold are larger than all relevant values below. Therefore, the study has good discriminating validity. Based on Confirmation Factors Analysis (CFA), the data quality check indicates that both reliability and validity of variables are perfect.

4.2 Structural Equation Modelling

SEM hypothesis: sample covariance matrix (S) = model covariance matrix $\Sigma(\theta)$.

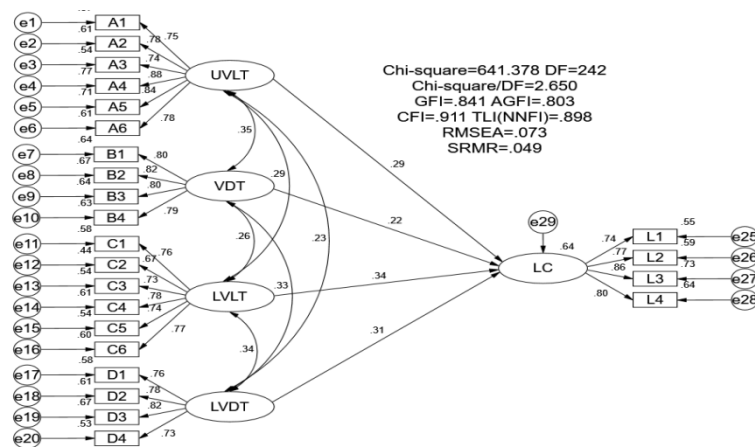


Figure 2. The Structural Equation Model of the Relationship between VK and LC

In SEM analysis, the endogenous factor was L2 listening comprehension, which was regressed on four exogenous factors including UVLT, LVLT, VDT and LVDT. Figure 2 is a research model of the relationship between two dimensions of VK and LC. From the analysis of SEM, it can be seen that all four aspects of VK have a positive effect on LC. The interpreted power of LC is $R^2 = .64$.

The research results above verify that the hypothesis of H1, H2, H3, and H4 are valid, providing support for verification of the hypothesis. And the RQ1 & RQ2 can be answered as well.

Table 3. Structural equation model fit

Index	Criteria	Model fit	Result
Chi-square	lower, better	641.378	
DF	higher, better	242	
Chi-square/DF	<3	2.650	meet criteria
GFI	>0.9	0.841	may accept
AGFI	>0.9	0.803	may accept
RMSEA	<0.08	0.073	meet criteria
SRMR	<0.08	0.049	meet criteria
CFI	>0.9	0.911	meet criteria
TLI(NNFI)	>0.9	0.898	meet criteria

Note: model fit criterion is proposed by some scholars like MacCallum et al.

It can be seen from Table 3 that the model fit index meets the requirements, indicating that SEM assumption is valid.

4.3 Mediating Effect

To check whether VF has a certain influence on the relevance of various dimensions of VK on LP, we adopted an indirect effect test strategy to verify the mediating effect of VF. We chose the most advanced methods of the 21st century: Bootstrapping and Monte Carlo Method, using repeated sampling to calculate sampling allocation, which can calculate the asymmetric confidence interval by calculating 5000 times. The statistical test power is higher than both *B-K* and Sobel *z* test. The second method is Monte Carlo Method with a bootstrap method of parameter estimation. There is a more symmetrical *a*b* confidence interval. Monte Carlo Method is faster and more accurate than bootstrap.

4.3.1 Testing the Mediating Effect of Vocabulary Fluency (UFLT vs. LC)

Figure 3a (standardized) and Figure 3b (non-standardized) are relationship diagrams of SEM between written receptive vocabulary breadth and LC on the condition that VF plays a role of a mediating variable.

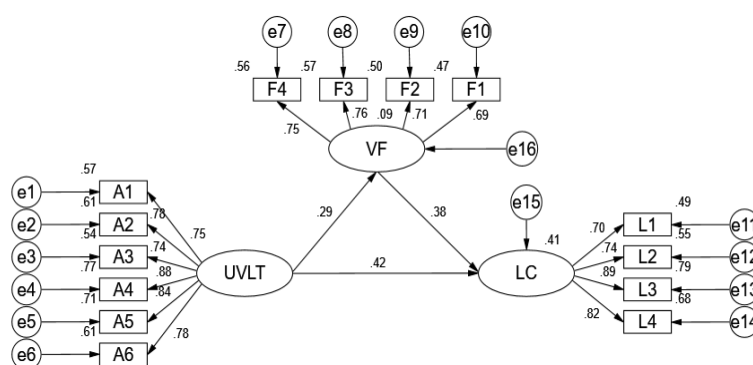


Figure 3a. (Standardization) Mediating effect of vocabulary fluency (UFLT vs. LC)

As shown in Figure 3a (standardized), correlation between written receptive vocabulary breadth and LC increase to .42 which is higher than that before intermediary, that shows VF has contributed to relationship between written receptive vocabulary breadth and LC.

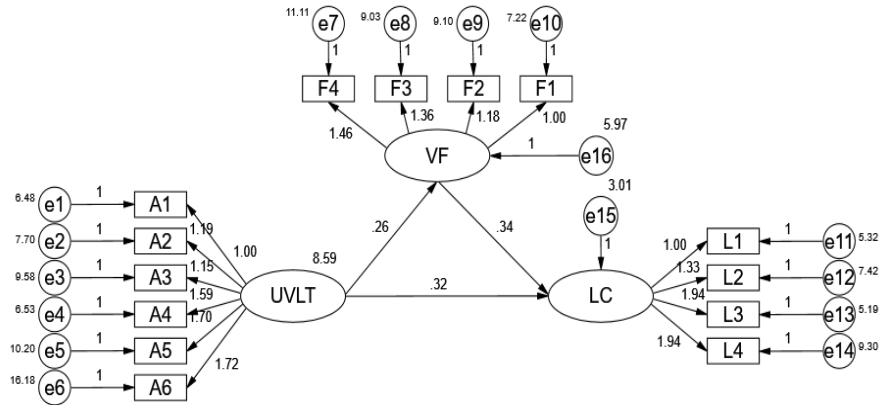


Figure 3b. (Un-standardization) Mediating effect of vocabulary fluency (UVLT vs. LC)

According to the non-standardized results in Figure 3b, Bootstrapping is used to calculate the total effect, direct effect and indirect effect of the relationship between written receptive vocabulary breadth and LC based on the mediating effect of VF.

Table 4. Mediated Variable on relationship between UVLT and LC

Relationship	Points Estimate	Product of Coefficients	Bootstrapping				
			Bias-Corrected 95% CI		Percentile 95% CI		
		SE	Z	Lower	Upper	Lower	Upper
Total Effects							
UVLT→LC	.410	.057	7.193	.306	.532	.305	.532
Direct Effects							
UVLT→LC	.324	.052	6.231	.228	.434	.227	.432
Indirect Effects							
UVLT→LC	.086	.025	3.440	.044	.143	.044	.141

Note: 5000 bootstrapping samples

As shown in Table 4, z-value is higher than 1.96 and zero is not in the confidence interval, which means that we can be confident that the indirect effect is different from zero, that is, the mediating effect exists and belongs to partial mediating.

4.3.2 Testing the Mediating Effect of Vocabulary Fluency (LVLT vs. LC)

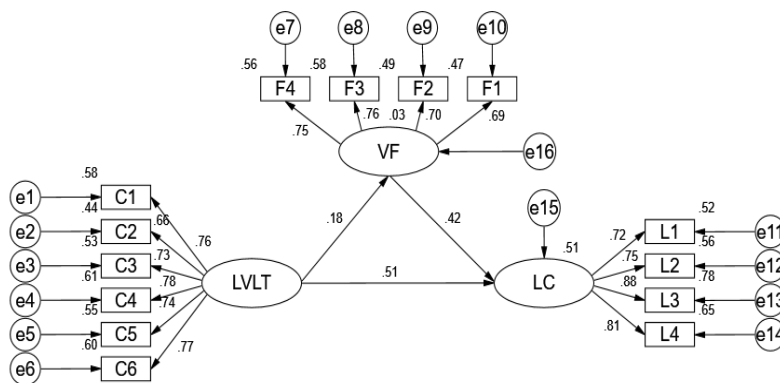


Figure 4. (standardization) Mediating effect of vocabulary fluency (LVLT vs. LC)

Figure 4 is a relationship diagram of SEM between oral receptive vocabulary breadth and LC on the condition that VF plays a role of a mediating variable.

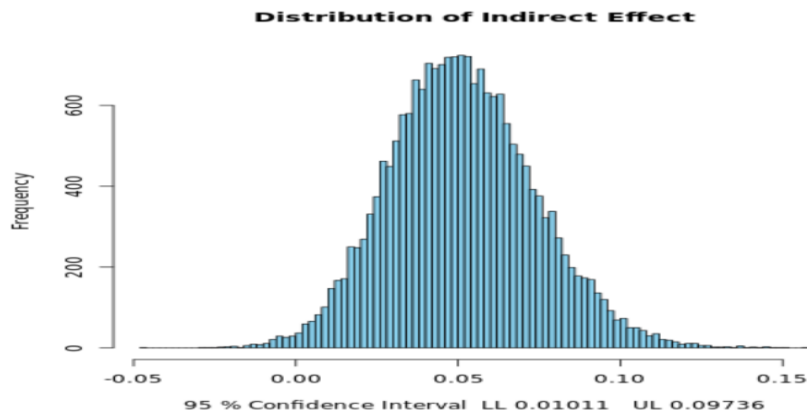


Figure 5. Mediating effect of vocabulary fluency (LVLV vs. LC)

Entering the corresponding data into the calculation area of the confidence interval on the Internet <http://quantpsy.org/medmc/medmc.htm>, as shown in Figure 5, it is calculated that there is no “0” included between in the lower bounds (LL 0.01011) and in the upper bounds (UL 0.9736), which indicates that indirect effects exist. The mediating effect size is 0.129 which indicates that the calculated mediating effect accounts for 12.9% of the total effect. That is, VF plays a certain role as a booster in the relationship of the two.

4.3.3 Testing the Mediating Effect of Vocabulary Fluency (VDT vs. LC)

Based on the mediating effect theory diagram of VF, Figure 6a (standardized) and Figure 6b (non-standardized) are the relationship diagrams of SEM between written receptive vocabulary depth and LC on the condition that VF plays a role of a mediating variable.

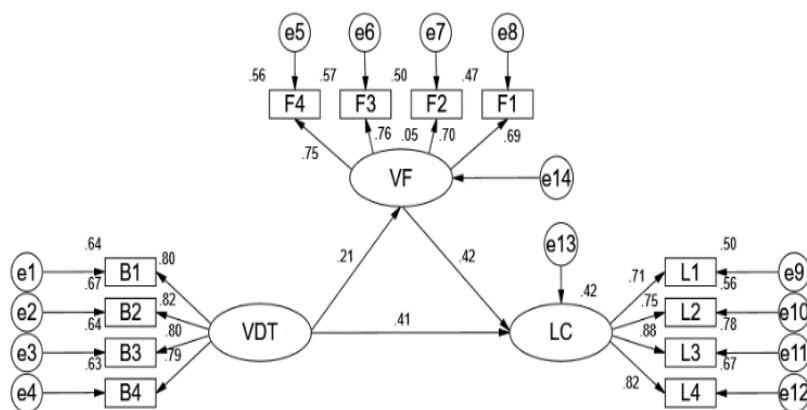


Figure 6a. (Standardization) Mediating effect of vocabulary fluency (VDT vs. LC)

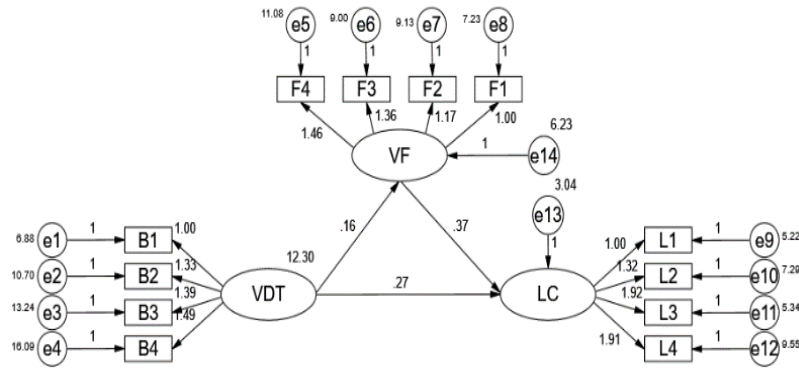


Figure 6b. (Un-standardization) Mediating effect of vocabulary fluency (VDT vs. LC)

According to the non-standardized results, Bootstrapping is used to calculate the total effect, direct and indirect effect of the relationship between written receptive vocabulary depth and LC based on the mediating effect of VF.

Table 5. Mediated Variable on relationship between VDT and LC

Relationship	Points Estimate	Product of Coefficients		Bootstrapping			
		SE	Z	Bias-Corrected 95% CI		Percentile 95% CI	
				Lower	Upper	Lower	Upper
VDT→LC	.328	.053	6.189	.226	.436	.236	.442
			Total Effects				
VDT→LC	.269	.048	5.604	.177	.368	.180	.371
			Direct Effects				
VDT→LC	.058	.021	2.762	.025	.108	.023	.105
			Indirect Effects				

Note: 5000 bootstrapping samples

As shown in Table 5, z-value is higher than 1.96 and zero is not in the confidence interval, indicating that we can be confident that the indirect effect is different from zero, that is, the mediating effect exists and belongs to partial mediating. As shown in Figure 8a (standardized), the correlation between written receptive vocabulary depth and LC raise to .41 which is higher than that before intermediary, indicating that VF has contributed to the relationship between written receptive vocabulary depth and L2 LC.

4.3.4 Testing the Mediating Effect of Vocabulary Fluency (LVDT vs. LC)

Figure 7 is a relationship diagram of SEM between oral receptive vocabulary depth and LC on the condition that VF plays a role of a mediating variable. Entering the corresponding data into the calculation area of the confidence interval on the Internet <http://quantpsy.org/medmc/medmc.htm>, as shown in Figure 8, it is calculated that there is no “0” contained between in the lower bounds (LL 0.0223) and in the upper bounds (UL 0.122), indicating that the indirect effects exist. Based on the calculation, mediating effect size is 0.175 which shows that the calculated mediating effect accounts for 17.5% of the total effect. That is, VF plays a certain role as a helper in the relationship of the two.

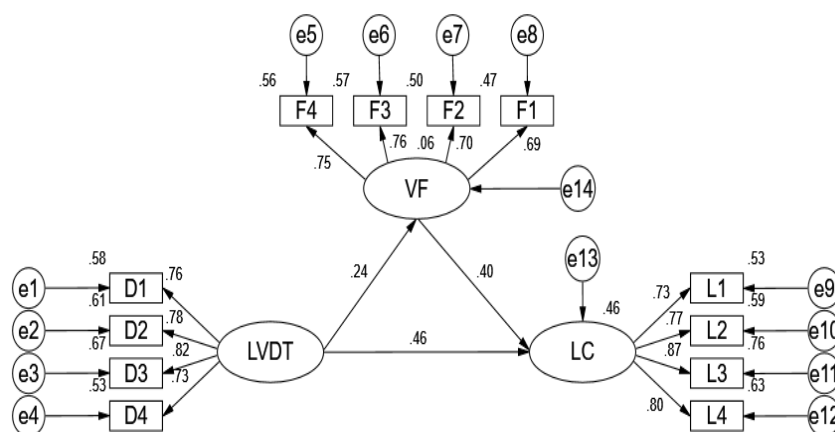


Figure 7. (standardization) Mediating effect of vocabulary fluency (LVDT vs. LC)

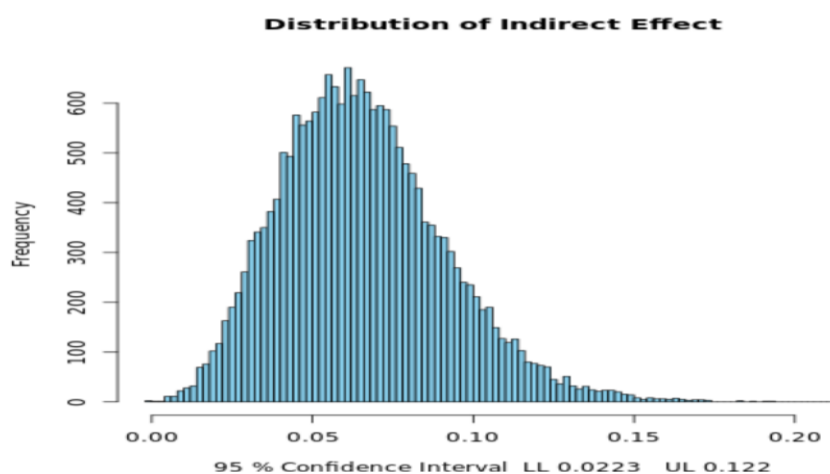


Figure 8. Mediating effect of vocabulary fluency (LVDT vs. LC)

The mediation effect of VF on the relationship between two dimensions, including four aspects, of VK and LC is tested using Bootstrapping and Monte Carlo (MC). The results show that the mediation effect of VF exists. Therefore, the hypothesis of H5 is valid, thus provides support for verification of the hypothesis, and the answer to the RQ3 is offered as well.

5. Discussion and Conclusion

5.1 Discussion

The aim of the current study was to examine the role of L2 VK in LP, as well as their relative significance. This study found that there is a positive correlation between VK and LC, but different dimensions of VK had different roles on LC. Additionally, the written and oral vocabulary measurements had different effects on LC. Below are the details:

1) The regression coefficient between written receptive vocabulary breadth and LC is roughly close to the results of Nasir (2017) with .33 and lower to the results of Cheng (2018) and Favardin and (2017) with .39 and .47 respectively. However, it is far from the results of other scholars. For instance, Staehr (2009) pointed out that the influence of vocabulary breadth on LC was .70 and Wolf (2019) with .70. The reason why the research results are quite different may be due to the different measurement instrument used, the different age group and level of subjects involved.

2) The regression coefficient between written receptive vocabulary depth and LC is lower than Favardin's (2017) with .43 and Nasir's (2017) with .43. But the research results of some scholars are very high, such as Staehr (2009) with .65 and Wolf (2019) with .75. Some scholars believe that vocabulary breadth has a high degree of correlation with LC, while others believe that vocabulary depth is more relevant to LC. However, we believe that written receptive vocabulary breadth and depth are almost the same important for LC because there is only difference of 0.07 between the two.

3) The influence of oral receptive vocabulary breadth and depth on LC are relatively higher than that of written receptive vocabulary breadth and depth on LC. This result is consistent with language nature, that is, oral vocabulary knowledge is much more related to listening. Therefore, it sounds completely reasonable that oral vocabulary measurements play a more significant role on LC than written one.

4) The multiple regression analysis shows that receptive vocabulary breadth and depth together account for 64% of variance in LC, which demonstrates that independent variables could together predict more than half of LC. This result illustrates an important contribution and shows a moderate relationship between learners' VK and the quality of their LC.

5) This study found that VF has a partial mediating effect on the relevance of VK and LC. Whether it is written or oral receptive vocabulary breadth or depth, the importance of VF is self-evident. Daller et al. (2007) divided VK into breadth, depth, and fluency. The development of fluency is essential for listening, speaking, reading, and writing, as Schmitt (2014) proposed that simple vocabulary learning can be transferred into practical applications with the help of fluency.

6) On the condition of VF as a mediating variable, the correlation degree of VK to LC is increased whether it is the written or oral receptive vocabulary breadth or depth.

In short, oral receptive VK is more important for LC than written receptive VK; Receptive vocabulary breadth and depth are nearly the same important for LC; VF, as an intermediary variable, can improve the correlation degree between four aspects of VK and LC.

5.2 Conclusion

Based on the analysis above, the following conclusions could be drawn: 1) The oral receptive vocabulary breadth has the strongest correlation with LC and the second rank is oral receptive vocabulary depth; 2) The correlation between receptive vocabulary breadth and LC are slightly higher than those between receptive vocabulary depth and LC; 3) 64% explanation power of variance indicates that the predictive power of VK for LC reach over medium; 4) VF plays a partial mediating role between four aspects of VK and LC; 5) On the condition of the mediating effect of VF, the influence of four aspects of VK on LC increased by 13%, 19%, 17% and 15% respectively. Therefore, to improve LC, language learners should make their advantages and avoid disadvantages in vocabulary learning by not memorizing words given to them.

6. Implications and Limitations

6.1 Implications

The current study was conducted to fill the gap of the research on the relationship between written and oral VK and LC, and to bridge the gap of the research on the mediating effect of VF on the relationship between VK and LC. There are some teaching and testing implications of VK and LC based on the conclusions of this research. Firstly, the main purpose of measuring VK is to accurately know about the learner's current VK level supporting basic language skills such as "LC". It is

recommended that a multi-dimensional measurement mode of VK should be used instead of only one dimension. Secondly, this study found that due to the intervention of VF, the influence of exogenous variables (four aspects of VK) on endogenous variables (LC) increases accordingly. Therefore, we suggest that teachers should allocate 1/4 of classroom teaching time to develop vocabulary fluency in teaching design. Thirdly, the research results show that receptive vocabulary breadth and depth are almost equally important on LC. Therefore, it is recommended that vocabulary teaching should not favor one aspect over another. For example, learners should be encouraged to read extensively, being exposed to a lot of new words and strengthen the training of vocabulary depth in the context, and retrieving the words learned in reading and listening repeatedly and fluently, so as to achieve the three-pronged development of "breadth", "depth" and "fluency" simultaneously. Fourthly, the oral receptive VK is more important for LC than written VK. Therefore, for L2 learners who want to improve single listening skill, it is recommended to consciously develop VK by listening and speaking, such as increasing the training of some exercises like "meaning recognition" and "meaning recall" in the process of listening.

6.2 Limitations

This study has several limitations. First, it only explores the correlation between the multiple dimensions of L2 VK and LC. Further study needs to focus on multiple dimensions of VK and multiple language skills at the same time. Second, this study only explores VF as a mediating variable to influence the importance of independent variables to dependent variables. Further study attempts to examine other mediating factors and moderating factors. Third, productive depth of VK measurement instrument and VF measurement tool have no ready-made versions, but based on the theories of related scholars, the two tools are developed by ourselves and used as a testing instrument for latent variables after passing exploratory and confirmatory factor analysis.

Funding

1. This work was financed by the Major Project of Ningde Normal University [grant number 2018ZDS04]
2. This work was financed by the Scientific Research Innovation Team of Ningde Normal University [grant number 2018T08]

References

- 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>
- Dabbagh, A. (2016). The predictive role of vocabulary knowledge in listening comprehension: depth or breadth? *International Journal of English Language and Translation Studies*, 4(3), 1-13. <https://www.eltjournal.org>
- Daller, H., & Xue, H. (2007). Lexical richness and the oral proficiency of Chinese EFL students. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 150-164). Cambridge University Press.
- Farvardin, M. T., & Valipouri, L. (2017). Probing the relationship between vocabulary knowledge and listening comprehension of Iranian lower-intermediate EFL learners. *International Journal of Applied Linguistics and English Literature*, 6(5), 273-278.

<https://doi.org/10.7575/aiac.ijalel.v.6n.5p.273>

- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(2), 39-50. <https://doi.org/10.2307/3151312>
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in second language acquisition*, 21(2), 303-317. <https://doi.org/10.1017/S0272263199002089>
- Hui, B. and Godfroid, A. (2020). Testing the role of processing speed and automaticity in second language listening. *Applied Psycholinguistics*, 1-27. <https://doi.org/10.1017/S0142716420000193>
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2), 255-271. <http://doi.org/10.1093/applin/19.2.255>
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening Vocabulary Levels Test. *Language Teaching Research*, 19(6), 741-760. <https://doi.org/10.1177/1362168814567889>
- Meara, P. (1996). *The dimensions of lexical competence: performance and competence in second language acquisition*. Cambridge University Press.
- Nasir, N. F. W. M., Manan, N. A. A., & Azizan, N. (2017). Examining the relationship between vocabulary knowledge and general English language proficiency. *Journal of Social Sciences and Humanities*, 1, 15-22. <http://ir.uitm.edu.my/id/eprint/29794>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review-Revue Canadienne Des Langues Vivantes*, 56(2), 282-307. <https://doi.org/10.3138/cmlr.56.2.282>
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning*, 52(3), 513–536. <https://doi.org/10.1111/1467-9922.00193>
- Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21(1), 28–52. <https://doi.org/10.1191/0265532204lt273oa>
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(4), 355-371. <https://doi.org/10.1177/026553229301000308>
- Read, J. (1998). *Validating a test to measure depth of VK: Validation in language assessment*. Erlbaum.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In B. L. P. Bogaards (Ed.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209–227). John Benjamins.
- Schoonen, R., Gelderen, A. v., Gloppe, K. d., Hulstijn, J., Annegien Simis, Snellings, P., & Stevenson, M. (2003). First language and second language writing: the role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, 53(1), 165-202. <https://doi.org/10.1111/1467-9922.00213>
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language*

- Teaching Research*, 12(3), 329–363. <https://doi.org/10.1177/1362168808089921>
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/s0261444812000018>
- 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. <https://doi.org/10.1177/026553220101800103>
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 173–199. <https://doi.org/10.1017/s0272263104262027>
- Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. DeGroot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford University Press.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139–152. <https://doi.org/10.1080/09571730802389975>
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(4), 577–607. <https://doi.org/10.1017/S0272263109990039>
- Webb, S., & Macalister, J. (2013). Is text written for children appropriate for L2 extensive reading? *TESOL quarterly*, 47(2), 300–322. <https://doi.org/10.1002/tesq.70>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: developing and validating two new forms of the VLT. *International Journal of Applied Linguistics*, 168(1), 34–70. <https://doi.org/10.1075/itl.168.1.02web>
- Wolf, M. C., Muijselaar, M. M., Boonstra, A., & de Bree, E. H. (2019). The relationship between reading and listening comprehension: shared and modality-specific components. *Reading and Writing*, 32(7), 1747–1767. <https://doi.org/10.1007/s11145-018-9924-8>

AUTHOR BIODATA

Yanli Tong is a PhD student of University of Malaya as well as an associate professor of Ningde Normal University, China. Her main research interest is L2 acquisition and vocabulary teaching and research.

Zuwati Hasim is a senior lecturer at the University of Malaya, Kuala Lumpur. She completed her PhD in Applied Linguistics from the University of Waikato, New Zealand where her research focused on understanding ESL writing and its assessment practice through the sociocultural lens. She is an associate member of the Centre for Civilisational Dialogue of University of Malaya and an editor to KATHA Journal. Her current research interests include assessment, action research, pedagogy, teacher education, reflective practice, higher education and school-based assessment.

Huzaina Abdul Halim is a PhD Supervisor of University of Malaya. Her main research interest is vocabulary teaching and research.