Vocabulary Size Tests of Different Modality and Their Relationships with L2 Reading and Listening Comprehension by Korean EFL Learners in Middle School

Myoungjin Kim†
Seoul National University

ABSTRACT

The present research intends to compare the orthographic and phonological vocabulary sizes of Korean EFL students in middle school and to address the relationship between the two different types of vocabulary size tests and L2 reading and listening comprehension. The participants were found to have greater orthographic vocabulary knowledge than phonological knowledge. Specifically, their vocabulary knowledge, regardless of its modality, drastically decreased in frequency level from 1,000 to 2,000 and from 3,000 to 4,000. Although both types of vocabulary knowledge exhibit a correlation with each other, as well as with reading and listening comprehension, orthographic vocabulary size was shown to be the most predictive in terms of the variance found in both reading and listening. The findings of the study contribute to existing research on L2 vocabulary acquisition by providing further evidence of the non-parallel development of phonological and orthographic vocabulary knowledge by EFL students, and by suggesting the significant predictive value that orthographic vocabulary knowledge has on the performance of students in reading and listening comprehension tests employed in Korea.

Keywords: vocabulary knowledge, orthographic vocabulary, phonological vocabulary, reading comprehension, listening comprehension

1. Introduction

Vocabulary knowledge has been recognized as one of the most significant aspects of language, which plays a pivotal role in L2 learning (Nation 2001, 2006). In this regard, it is a primary goal for a myriad of L2 learners to acquire a sizeable amount of lexical items as they develop their proficiency. In order to monitor their lexical progress and to diagnose what they are missing, implementing vocabulary assessment

† Corresponding author: lucy84@snu.ac.kr

is essential for teachers, students, and even researchers. Moreover, building upon the well-grounded relationship between vocabulary and language sub-skills, the information on one's vocabulary knowledge may allude the overall developmental status of L2 knowledge. It should be noted that it is more challenging to directly measure one's language proficiency given its abstract quality, and thus, learners are often required to prove their knowledge in some other countable or measurable ways. Since lexical knowledge is regarded as more manageable in terms of countability, at least on the surface, assessing vocabulary enables learners to estimate their overall language proficiency (Milton 2009).

Meanwhile, according to common consensus, vocabulary knowledge involves multidimensional constructs (Nation 2001) and learning lexical items is an incremental procedure where learners gain knowledge from wide-ranging aspects of vocabulary as they develop their proficiency (Schmitt 2008). Seen in this perspective, it must be admitted that assessing vocabulary knowledge is not a straightforward task in reality. In attempting to measure vocabulary knowledge, it should be primarily decided in which specific type of knowledge is to be assessed among those various dimensions. In particular, vocabulary "breadth" is regarded as a significant indicator of not only one's lexical competence but also his or her overall language skills. As the term "breadth" denotes the number of words that one can map the meaning onto the "form", learners are expected to acquire knowledge on both written and spoken form when expanding their size of the vocabulary. Despite this broad agreement on the complex nature of vocabulary knowledge and its measurement, the majority of learning and teaching practice in an EFL setting appears to highlight only a limited aspect of vocabulary size, namely, orthographic vocabulary size. At this point, it is speculated that the students in EFL context might not have developed another important aspect of vocabulary form, that is, a phonological dimension of vocabulary.

Not only has the language classroom failed to distribute equal attention to different types of vocabulary, but most of the L2 research has also focused predominantly on orthographic vocabulary knowledge. Specifically, the previous research investigating the relationship between vocabulary and language sub-skills, such as reading and listening, has mostly ignored the phonological aspects of vocabulary (Bonk 2000; Kelly 1991; Koda 1989; Laufer and Revenhorst-Kalovski 2010; Li and Kirby 2015; Schmitt, Jiang and Grabe 2011; Stæhr 2008, 2009; Qian 1999, 2002; Van Zeeland and Schmitt 2012). Although such studies were able to find the strong relationship between vocabulary and language skills, the majority of assessment employed in the

studies exclusively measured one's orthographic knowledge. Given the multifaceted aspects of vocabulary knowledge, a single measurement cannot comprehensively encompass the different dimensions of lexical competence. Accordingly, a variety of assessing instruments should be further added to the investigation of one's vocabulary competence to provide a more comprehensive picture of L2 lexical development. Despite the growing acceptance on the importance of vocabulary in language, only a few studies to date have investigated phonological vocabulary knowledge of L2 students and its relationship with reading and listening comprehension development (Cheng and Matthews 2018; Matthew and Cheng 2015; Milton et al. 2010; Mizumoto and Shimamoto 2008).

In the light of these considerations, the primary purpose of the present study is to examine the difference between the orthographic vocabulary size (OVS) and phonological vocabulary size (PVS) of five frequency levels (i.e., 1,000~5,000 levels based on the British National Corpus) with a particular focus on Korean EFL students in middle school. Since no previous study has investigated the PVS of secondary school students, the outcomes of the present study would contribute to a deeper understanding of L2 lexical knowledge development. The study also aims to explore the relationship between OVS/PVS and reading/listening comprehension. Although a few studies have implied the modality-specific nature of vocabulary and its effect on the comprehension processing (Cheng and Matthews 2018; Matthew and Cheng 2015; Milton et al. 2010; Mizumoto and Shimamoto 2008), the scant amount of empirical data seem to be insufficient to determine the relative effect size of each vocabulary knowledge on reading and listening comprehension.

Before discussing the key topics and relevant previous research, it should be addressed at the outset that, this paper has limited to the discussion of receptive dimension in exploring two types of L2 vocabulary size, since the purpose of the study is to investigate the relationship between vocabulary knowledge and "receptive" skills (i.e., reading and listening). In addition, although the productive dimension should be regarded as crucial in discussing one's lexical knowledge, there is a threshold of ability in recognizing words and their meanings before learners attempt to use the words in a productive way (Palmberg 1987). Given that the focus of the study is learners at a relatively early stage of acquisition, it seems plausible to limit the attention of the research mainly to the receptive dimension.

To this end, the present study addresses the following research questions:

1. Do Korean EFL middle school students' orthographic and phonological vocabulary sizes exhibit a significant difference?

2. To what extent do orthographic vocabulary size, phonological vocabulary size, reading and listening comprehension correlate with one another?
3. To what extent do orthographic vocabulary size and phonological vocabulary size contribute to predicting the performance of reading and listening comprehension tests?

2. Literature Review

2.1. Assessing receptive vocabulary size

As Nation (2001) and many other researchers have posited, there is a wide range of dimensions involved in vocabulary knowledge and whether one "knows" a certain word cannot be judged with ease by a single criterion. Among those different aspects, one conventional distinction has been made between the dimensions of vocabulary "breadth" and "depth". To put it simply, "breadth" of word knowledge denotes the number of words a learner knows; "depth" refers to how well the learner knows the word. In particular, much academic research has been conducted on vocabulary breadth up to date (e.g., Al-Hazemi 2000; Laufer and Nation 1995; Read 2000; Schmitt 2010; Vassiliu 1994), given the importance of vocabulary size in representing one's lexical knowledge. That is, vocabulary breadth can be regarded as a characteristic of one's entire lexicon, whereas vocabulary depth seems to be characteristics only of certain words, and therefore, it might be more plausible to examine one's vocabulary size in order to approach his overall vocabulary development (Meara and Wolter 2004).

With this in view, numerous efforts to measure one's vocabulary size lead to the development of instruments which are reliable and of high validity. Milton (2013) posited that vocabulary size in English is now readily measurable employing test instruments which have normalized scores. For instance, one of the most widely-employed vocabulary size tests (VST) is Vocabulary Level Test (VLT) , which is a multiple-choice format test where test takers select the matching explanation for the test item. Of important feature is the fact that the given explanation is a single word form instead of its lengthy definition. Hence, the test is able to assess the basic and initial stages of form-meaning mapping, which makes it feasible to administer with beginner and intermediate level learners (Kremmel and Schmitt 2018).

Nevertheless, despite its widespread use in both pedagogy and research, certain

limitations are associated with VLT. First, it should be noted that it is quite challenging to test the most frequent words since their meanings may not be expressed in the words of the same or higher frequency level. It is, therefore, possible that learners of the beginning level might not get the item correct despite their knowledge on that particular word. In addition, Meara (1992) points out the impracticality of VLT, since it allows only a confined number of test items within the given time due to its complexity. Furthermore, Kamimoto (2004) specified that the treatment for the guesswork has not been yet taken into consideration in VLT. Lastly, VLT consists of 10,000 frequency level words. Kremmel and Schmitt (2018) indicated that administering the test with 10,000 frequency level to novice L2 learner is redundant as it does not provide useful information.

Alternatively, there is another well-established VST—X-Lex (Meara and Milton 2003)—which is a checklist type test where the learners are asked to signify whether they know the word or not. As in the case of VLT, X-Lex also makes use of frequency information but comprises only 1,000-5,000 level, which has been found to be more strongly related to general language exams (Milton and Hopkins 2006). Several advantages exist for using X-Lex: First, the test has been proved to be of high practicality in the sense that when it is compared to VLT, test takers would require much less time to demonstrate whether they know the words. An average test taker would complete a series of the tests varying their frequency level in less than 30 minutes (Meara 1992). Second, to control the false answer (i.e., automatic check for 'yes') and to ensure the reliability, the test includes a fixed ratio of non-words (i.e., one-third of the test items). Hence, a test taker who says 'yes' to a large number of non-words will be regarded unreliable, and accordingly, the score would be penalized.

Admittedly, some doubts have been raised on the validity of the checklist format of X-Lex. For instance, Nation (2007) argued that a checklist type test requires the test takers to be interested enough to make a thoughtful and reliable judgment to test items. Another problem proposed is that test takers tend to answer 'yes' when facing the questions (Milton 2009). In addition, this tendency of over-estimation could be affected by individual differences or cultural differences (see Shillaw 1999; Vassiliu 1994). Despite these limitations, previous studies using X-Lex have reported that the reliability of the tests was kept intact even with the tendency of the over-estimation (Al-Hazemi 2000; Vassiliu 1994). Moreover, the advantage of the X-Lex in terms of practicality overrides its limitations given that the target population of this study is young EFL learners who may not sustain their effort and concentration

for a long time. To this end, the present study will employ the checklist format due to its high practicality and the adequate level of reliability and validity (see Milton 2007).

2.2. Phonological vocabulary size and phonological awareness

The two most popular VSTs mentioned in the previous section are in common in that they provide information exclusively on receptive orthographic vocabulary knowledge. In assessing receptive vocabulary knowledge, however, there is another notable dimension which has been received much less attention—phonological vocabulary knowledge. As Nation's (2001) taxonomy indicates, "knowledge on a word form" indicates not only what the word looks like but also what it sounds like. In addition, it has been generally acknowledged that phonological awareness plays a crucial role in word identification (Stonavich 2000). Phonological awareness, as defined by Yopp (1992), refers to the ability to hear and manipulate the sounds in the spoken form of a word and the understanding that spoken words and syllables are made up of sequences of phonemes. As phonological awareness, along with orthographic knowledge, have been regarded as the two most critical skills underlying both L1 and L2 literacy, it should be investigated whether students' phonological and orthographic knowledge progresses in a parallel manner.

With this consideration in mind, there have been some recent attempts to suggest that learners' phonological vocabulary knowledge is distinct from their orthographic knowledge (Field 2008; Milton and Hopkins 2006; Milton, Wade and Hopkins 2010; Mizumoto and Shimamoto 2008). For instance, Mizumoto and Shimamoto (2008) observed the difference between phonological and orthographic vocabulary size of Japanese university students and proposed that the mapping between the spoken form and its meaning further needs to be incorporated in VSTs. Milton and Hopkins (2006) tested both Arabic and Greek English learners and suggested that the phonological vocabulary knowledge has relevance to their L1 background and their L2 proficiency. That is to say, the orthographic similarity between L1 and L2 affects their reliance on phonological vocabulary knowledge; advanced learners showed a greater difference between orthographic and phonological vocabulary knowledge.

Although attempts to measure phonological vocabulary knowledge is still in its early stage, a few researchers propitiously have designed phonological tests of vocabulary breadth up to now (Fountain and Nation 2000; Matthews and Cheng 2015; Mclean, Kramer and Beglar 2015; Milton and Hopkins 2006). Among those

tests, Aural-Lex (Milton and Hopkins 2006) was selected as an ideal measurement for learners' aural and written vocabulary size in the present study. This test includes precisely equivalent items to the orthographic X-Lex, bearing the only difference in the modality of delivery. To be more specific, the test takers are presented with the test words in spoken form and asked to indicate whether they know the auditorily perceived word. Due to highly shared characteristics between X-Lex and Aural-Lex, the different performance found in the tests might be mainly due to the different delivery of the mode, not other characteristics of the tests.

Meanwhile, the previous studies employing phonological size tests have generally shown the tendency for L2 learners' orthographic vocabulary size to surpass phonological vocabulary size (e.g., Cheng and Matthew 2018; Mizumoto and Shimamoto 2008). This discrepancy between written and aural vocabulary knowledge could be derived from the prevalent vocabulary learning practice where learners, especially who are in an EFL setting, memorize only the spelling of the words. Indeed, Goh (2000) conducted an interview with Chinese EFL learners and observed that the majority of students did not attempt to learn the pronunciation of new words, and thus, they had not fully developed the automaticity with sound-to-script relationships. In this regard, it is not by chance to find that L2 learners' aural vocabulary size is smaller than written vocabulary size.

Yet, within the research focusing on differences between phonological and orthographic vocabulary size, there remains an unexplained aspect. For instance, it is uncertain whether the differences between the two knowledge show a similar pattern regardless of the characteristics of learners, such as their age, proficiency, or L1 background. Milton and Hopkins (2006) revealed that Arabic students did not show statistically significant differences between orthographic and phonological vocabulary size and the descriptive result indicated that their phonological vocabulary size was even slightly larger. In relation to the age difference, it is suspected that younger L2 learners might exhibit more balanced vocabulary knowledge since their literacy level is at its early stage. Milton et al. (2010) mentioned that beginning level learners have vocabulary knowledge predominantly in spoken form, while more advanced learners grow their lexicons through the written form. Nonetheless, recent discussions on vocabulary size tests with different modality tend to focus exclusively on college level students (Cheng and Matthew 2018; Mizumoto and Shimamoto 2008). In this regard, the focus of the present study is to assess phonological vocabulary size and orthographic vocabulary size of Korean EFL students in middle school. Having discussed the two different kinds of vocabulary size, it is now

necessary to explain how these types of vocabulary knowledge are related to other language skills, particularly reading and listening.

2.3. The relationship between L2 vocabulary size and L2 reading and listening

As was mentioned in Introduction, previous research has recognized the critical role played by vocabulary knowledge in successful language achievement (e.g., Chang 2007; Coxhead and Byrd 2007; Koda 1989). In this regard, the relationship between vocabulary knowledge and the four basic language skills has been the major area of interest within the field. Specifically, a considerable amount of literature has been published on the positive relationship between vocabulary knowledge and reading comprehension (Koda 1989; Laufer and Revenhorst-Kalovski 2010; Li and Kirby 2015; Schmitt, Jiang and Grabe 2011; Stæhr 2008; Qian 1999; 2002). For instance, Qian (2002) examined the relationship between reading and vocabulary knowledge and revealed that both vocabulary depth and breadth showed a significant correlation with reading comprehension (for depth: r = .77; for breadth: r = .74). Schmitt et al. (2011) also found a relatively linear relationship between the percentage of vocabulary known and the degree of reading comprehension. Li and Kirby (2015) further confirmed that vocabulary size has a great predictive value on reading comprehension test, especially with the multiple choice format. In addition, a considerable amount of literature has been published on the relationship between vocabulary knowledge and reading skill in Korean context. For recent example, Lee (2018) found a significant correlation between vocabulary size and reading proficiency, whereas the relationship between syntactic awareness and reading was not found.

As there have been assumptions that similar processes underlie reading and listening comprehension (e.g., Bejar, Douglas, Jamieson and Turner 2000), a handful of researchers attempt to investigate whether findings from reading research, which indicates the positive correlation with vocabulary, can also be transferred to in the case of listening comprehension (Bonk 2000; Kelly 1991; Stæhr 2009; Van Zeeland and Schmitt 2012). Bonk (2000) investigated Japanese EFL students in university and revealed that learners of all proficiency levels required considerably high lexical familiarity for competent listening comprehension. In a similar vein, Stæhr (2009) confirmed a significant correlation between listening and vocabulary knowledge and suggested that a lexical coverage of 98% appears to be essential for dealing with the spoken texts in the listening test. More recently, Van Zeeland and Schmitt (2012) proposed that 95% lexical coverage is sufficient to achieve adequate listening

comprehension. Overall, it seems reasonable to conclude that obtaining a desirable vocabulary size is great beneficial to one's listening proficiency.

Nevertheless, much of the reviewed studies on vocabulary and language skills tend to neglect phonological vocabulary knowledge in assessing learners' vocabulary size. To put it differently, the VST employed in their research has been limited to assess orthographic knowledge. As has been discussed, there is an agreement with the presumption that lexicon has two halves, an orthographic and a phonological half (Milton et al. 2010). Thus, the research investigating the relationship between vocabulary knowledge and language sub-skills should necessarily take into account both orthographic and phonological aspects of vocabulary in the measurement of lexical knowledge.

Drawing upon this major limitation, several studies employing phonological vocabulary test has been carried out to see the relationship between aural vocabulary knowledge and reading/listening (Cheng and Matthews 2018; Matthew and Cheng 2015; Milton et al. 2010; Mizumoto and Shimamoto 2008). In particular, the studies have highlighted the relationship between phonological vocabulary knowledge and listening skills based on the presumption that L2 listening necessarily requires an adequate level of spoken word recognition (Tsui and Fulliove 1998; Yi'an 1998). For instance, Matthew and Cheng (2015) proved that the phonological knowledge on the words strongly correlated with International English Language Testing System (IELTS) listening test (r = .73) and it was able to explain 54% of the variance found in the listening scores. Similarly, employing IELTS test, Milton et al. (2010) found that phonological vocabulary knowledge does strongly correlate with listening scores (r = .67).

Despite the observed correlation between phonological vocabulary knowledge and listening, the role of phonological knowledge in the performance of listening comprehension remains unclear. Although Cheng and Matthew (2018) suggested that phonological vocabulary knowledge per se can explain 51% of the variances of the listening score, the test format they employed simultaneously assessed orthographic knowledge and the test item was given in a sentence, which might have required a different type of processing. In addition, in Stæhr (2008), it was revealed that vocabulary size correlated more strongly with reading (r = .83) than with listening (r = .69). Likewise, Van Zeeland and Schmitt (2012) indicated that the role of vocabulary knowledge seems to be smaller in L2 listening comprehension as they found the smaller lexical coverage is required for listening than for reading.

Meanwhile, with regards to reading skills, it may intuitively sound unreasonable

on the surface to postulate the relationship between reading and phonological vocabulary knowledge. Nonetheless, as noted by Metsela and Ehri (1998), extracting phonological information from a word is the first and the most significant step in the reading comprehension process, and therefore, it stands to reason that accessing phonological information is essential, especially in reading a foreign text. In L2 research, Nassaji and Geva (1999) empirically proved that both orthographic and phonological processing contributed significantly to reading performance. Judging from these arguments, it seems highly probable to investigate whether phonological vocabulary knowledge is related to successful reading performance.

Given this research background, the present study purports to estimate the English vocabulary size of Korean EFL middle school learners both in terms of phonological and orthographic aspects, and to measure the extent to which this vocabulary size contributes to students' performance on listening and reading comprehension.

3. Method

3.1. Participants

33 Korean 3rd-grade middle school students (9th-grade in K-12 system) in Seoul participated in the study. The initially recruited participants were 38, but five students were excluded from the analyses due to the incompletion of the tasks. The participants were recruited from two different mixed-proficiency level classes and these two classes are regarded as homogeneous in relation to their English proficiency, judging from their English final-term scores. It needs to be specified first that there was no separate measure on students proficiency, but given their grade level and the informal comments from their English teacher, the students' levels were assumed as beginner to pre-intermediate in general. None of them have experience in abroad more than one year.

3.2. Materials

Orthographic and a phonological vocabulary size test (OVST and PVST, respectively) were created with 300 items (including 100 non-words items) were adopted from X-Lex (See Appendix 1). In this test, the students were required to indicate whether they know a specific word and/or whether they thought the word is a non-word.

The rationale of employing the format of X-Lex is due to its several advantages discussed in the literature review. In addition, X-Lex is more desirable to see the relationship between the role of vocabulary knowledge in reading and listening. Kremmel and Schmitt (2016) stated that fluent reading and listening requires rapid word recognition and automatic recall of the meaning. A matching or multiple-choice format, however, does not represent real-world reading, because the text would not provide explanations to choose from for unknown words in the actual course of reading or listening. Since one of the purposes of the study is to examine the relationship between vocabulary tests and reading comprehension (RC)/listening comprehension (LC), X-Lex seems more reasonable to be adopted in this research.

OVST and PVST differ only in terms of modality, that is, the test items in each test are delivered either auditorily or in written form. To put it differently, the same test item was presented aurally in one class and visually in another one. Each test comprises five frequency levels (i.e., 1000~5000) and hence, each of the levels consists of 30 items (including 10 non-words). Since the target words for the two tests were different to avoid priming effect, the major concern was to control the effect of the test items. In an attempt to deal with this issue, each VST was constructed with the same word category ratio, which is presented in Table 1.

Table 1. Distribution of word classes in the vocabulary size test employed in the study

Level	1,000	2,000	3,000	4,000	5,000
Nouns	5	5	5	5	5
Verbs	5	5	5	5	5
Nouns/Verbs	4	4	4	5	5
Adjectives	4	4	4	4	4
Adverbs	2	2	2	1	1
Total	20	20	20	20	20

Note. Non-words are excluded

The PVST was constructed using a voice from Google translator (female voice with an American accent). In the original version of Aural-Lex, U.K. received pronunciation (RP) is utilized for each item, but the present study employed American pronunciation as it is predominantly used in English education at schools throughout Korea (Yim 2007). The pauses between each word were timed evenly (i.e., 3 sec.) using Audacity, a sound editing software. The total length of time

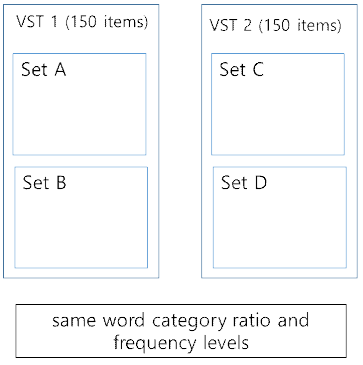
recorded in PVST was approximately 13 minutes but it was delivered through the two periods. To be more specific, half of the items were presented in the first period and the other half were presented in the second period). The major difference between PVST employed in this study and Aural-Lex is that the participants were not allowed to repeat the word they heard as in the case of LC test in order to make the test more similar to the real-world listening.

With regard to the RC test, the instrument consists of 10 written paragraphs (Mean word count=95.6) each of which associates one comprehension question. The test items were adopted from the 2017 National Assessment of Educational Achievement in Korea. In relation to the LC test, the students took the test one month prior to the vocabulary and reading tests, as a part of the performance assessment in school. The instrument included 20 test items and was developed by the Seobu District Office of Education in Seoul.

3.3. Procedure

To describe the data collection process (see Figure 1), the two types of tests were administered throughout three 45-minute periods for each classroom. To rule out the possibility that students might lose their attention with 150 items in the same modality format, the students were asked to take both PVST and OVST in the same period. Nonetheless, due to the short given time, the participants took the first half of each VST in the first period, and the second half was administered in the second period. In addition, participants undertook the reading test in the third period. As briefly mentioned in the previous section, the same test item was delivered differently across the class. To put it in more detail, 300 test items were divided into four sets (A-D) and each set was delivered differently across the classes. For instance, students in class 1 undertook the spoken form of set A and the written form of set B during the first period, whereas participants in class 2 received the written form of set A and the spoken form of set B in the same period. During the second period, the spoken form of set C and the written form of set D were administered in class 1, and the written form of set C and the spoken form of set D were delivered in class 2. To avoid the effect of the order of administration, the test takers who received the written version first in the first session undertook the spoken version first in the next period, and vice versa. Regarding the VSTs, each test was scored under the scoring matrix presented by X-Lex (see Meara 1992, p. 13). To be more specific, the numbers of Hits (i.e., correct answers) and False Alarms (i.e., 'yes' response

for non-word) were counted, and their combination produces the final score out of 100 based on the scoring matrix.



Period (45 min.)	Class 1	Class 2
1	PVST 1 – set A	PVST 1 – set B
	OVST 1 – set B	OVST 1 – set A
2	OVST 2 – set C	OVST 2 – set D
	PVST 2 – set D	PVST 2 – set C
3	R/C Test	

Figure 1. The data collection procedure

3.4. Data analysis

Four steps of analyses were conducted. First, test scores from the two VSTs, RC, and LC tests were analyzed to yield a range of descriptive statistics including learners' vocabulary frequency profile. Second, in order to respond RQ1, separate paired sample *t*-tests were conducted to confirm whether there is a significant difference between the learners' OVS and PVS in general and in each frequency level. Then, to address RQ2, Pearson correlation analysis was undertaken to verify the strength of relationships among OVST, PVST, RC, and LC. Lastly, in response with RQ3, two hierarchical multiple regression analyses were performed to determine the predictive power of each VST on RC and LC.

4. Results and Discussion

4.1. The Comparison between OVST and PVST

The minimum, maximum, mean scores, standard deviations (SD), and the values of skewness and kurtosis for the four types of the test undertaken are presented in Table 2. Mean scores suggest that participants achieved higher scores for OVST than for PVST. Values of skewness show that scores from RC and LC are not normally

distributed, evidenced by values less than -1. In addition, although there is no absolute consensus as to the acceptable range of skewness, some researchers (e.g., George and Mallery 2010) indicated that the values between −2 and +2 are regarded as admissible. In this regard, it is not entirely unreasonable to conduct the statistical analyses with the given data.

Table 2. Descriptive statistics for scores derived from the test instruments (N=33)

	Mix (0)	Max (100)	Mean	SD	Skewness	Kurtosis
OVST	4.8	81	43.38	20.23	.09	−.87
PVST	.89	65.2	34.38	16.36	.00	−.95
RC	10	100	71.21	27.92	−1.06	.18
LC	30	100	79.24	20.12	−1.18	.24

Regarding the vocabulary profile of OVS and PVS, Figure 2 displays mean scores of each frequency level in OVST and PVST. What stands out in the figure is PVST and OVST generally support the frequency model of learning (see Meara 1992), which proposes that the higher frequency the word has, the more likely it is to be acquired. Although there is a slight increase between 2,000~3,000 level in OVST, the discrepancy is assumed to be only minimal. As can be seen, a distinctive downward slope is found between 1,000~2,000 level and 3,000~4,000 level. Second and more importantly, the figure shows mean scores for OVST were higher than those for PVST, except the 4,000 level.

As can be seen from these results, it appears that Korean EFL middle school students have achieved competent vocabulary knowledge for 1,000 frequency level, particularly in respect of orthographic vocabulary, while the mastery of 4,000~5,000 level words is far from being satisfactory in terms of both orthographic and phonological knowledge.

Next, separate paired-samples *t*-tests were conducted to examine if the difference between OVS and PVS is statistically significant. Table 3 presents the mean differences between OVST and PVST on the whole and between each level of OVST and that of PVST. It is apparent from this table that the EFL students possess larger orthographic vocabulary than phonological vocabulary in general. In specific, those differences were significant within 1,000~3,000 frequency level words; beyond 3,000 level, both types of vocabulary size seem to be considerably sparse.

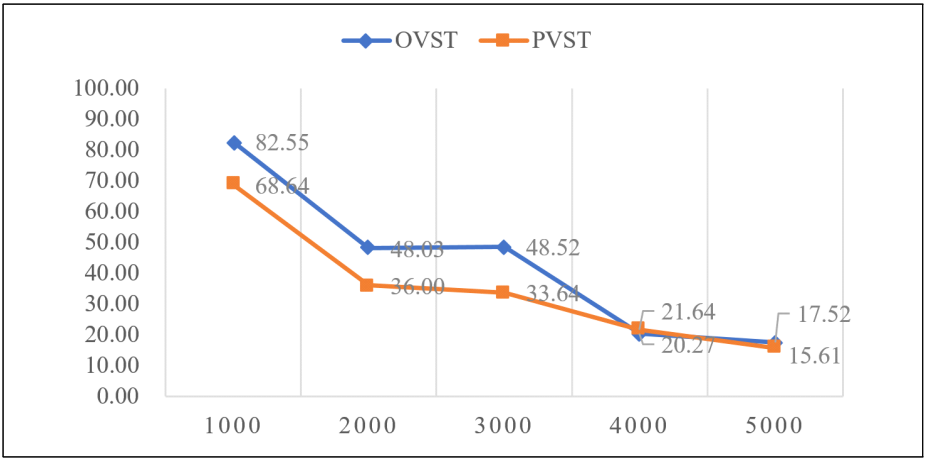


Figure 2. Mean scores from OVST and PVST in each frequency level

Table 3. Mean differences between OVST and PVST

	Mean Differences	SD Differences	*t*(32)	*p*
OVST-PVST	8.94	14.33	3.58	.001
OVST1-PVST1	13.90	18.67	4.28	.000
OVST2-PVST2	12.03	20.57	3.36	.002
OVST3-PVST3	14.88	18.41	4.64	.000
OVST4-PVST4	−1.36	21.70	−.36	.720
OVST5-OVST5	1.91	16.23	.68	.504

The result indicates the significant difference between the two types of VST. It was observed that Korean EFL middle school students appear to retain larger OVS than PVS, as the Japanese university students reported in Mizumoto and Shimamoto (2008). It is hypothesized that middle school students, particularly those in the 3rd grade, have already started a noticeable training on literacy skills and therefore, they have become more familiar with the orthographic form of the words. It remains to be seen if such pattern is consistently found among the students of different grades or school programs.

One interesting finding contributing to students' low performance in PVST is that the ratio of False Alarm (i.e., the ratio of non-words indicated by the participants that they know among the total non-word items) was considerably higher in PVST (18.42%) than OVST (5.39%). It seems that students were not able to delicately

detect the phonemic level sound, evidenced by an informal report by one participant who perceived "lorrid (non-word)" as "glory", and "antile (non-word)" as "entire". A possible explanation for this observation could be the lack of their L2 phonemic awareness. Although the participants seemed to succeed in recognizing its syllable structure in a quite similar way (e.g., perceiving a nonword "lorrid" as a real word "glory"), they were less successful in terms of accurate aural word recognition due to their deficiency in phonological awareness.

The fact that the students' phonological vocabulary knowledge has not sufficiently developed might be related to unbalanced attention for the written and spoken text in real classroom. Despite the fact that the textbook equally comprises all the basic sub-skills of English (i.e., reading, listening, writing and speaking), it is admitted that both teachers and students tend to aim their attention exclusively at receptive written skills possibly due to the washback effect of Korean College Scholastic Ability Test (CSAT). Consequently, the objective of vocabulary learning has also drawn upon expanding their receptive orthographic knowledge only. Considering numerous previous studies suggesting the importance of phonological awareness in the four skills (e.g., For reading, Carroll, Snowling, Hulme and Stevenson 2003; Stahl and Murray 1994; For writing, Abbott and Berninger 1993; Allor 2002; Diamond, Gerde and Powell 2008; For listening, Cheung,2007; Li, Cheng and Kirby 2012; For speaking, Giambo and McKinney 2004), an equal amount of attention should be paid to the spoken input, with strong emphasis on low-level skills.

4.2. The relationship between the two types of VSTs and RC/LC

In order to examine the relationships among the four types of the test, pearson correlation analysis was performed. Table 4 confirms that all the test instruments are significantly correlated with each other (ranging from .50-.72), which ties well with the previous studies (e.g., Cheng and Matthew 2018). With regard to the question,

Table 4. Correlation between scores derived from the test instruments

	OVST	PVST	RC	LC
OVST	1	-	-	-
PVST	.72**	1	-	-
RC	.57**	.50**	1	-
LC	.67**	.55**	.80**	1

** $p<.01$

which type of VST is more closely related to RC and LC, the analysis shows that OVST has a stronger relationship with both RC and LC than PVST does.

Furthermore, closer inspection of the data revealed that both RC and LC have a relationship with 1,000-3,000 vocabulary level with a higher significance, regardless of the types of vocabulary test (see Table 5).

Table 5. Correlation between a mean score of 1000~5000 frequency levels and that of RC and LC

	RC	LC	OVST 1	OVST 2	OVST 3	OVST 4	OVST 5	PVST 1	PVST 2	PVST 3	PVST 4	PVST 5
RC	1	.80**	.62**	.50**	.59**	.35*	.38*	.51**	.49**	.61**	.16	.19
LC		1	.71**	.65**	.73**	.43*	.24	.50**	.60**	.68**	.24	−.07

*p<.05, **p<.01

Next, in order to investigate the predictive power of the two VSTs on RC and LC, two hierarchical multiple regression analyses were employed. Regarding the logic for the entry order of predictive variables, the one with the strongest magnitude of correlation (i.e., OVST) with the outcome variable was added in the first step.

The result of hierarchical regression analysis is presented in Table 6, which displays the relative predictive value of OVST and PVST on L2 reading scores. The

Table 6. Hierarchical regression analysis addressing the predictive value of vocabulary knowledge on RC

	R	R^2	ΔR^2	Unstandardized		Standardized
				B	SE B	β
Step 1	.57	.33				
Constant				36.87	9.96	
OVST				.79	.20	.50**
Step 2	.59	.35	.02			
Constant				34.56	10.09	
OVST				.61	.29	.44*
PVST				.30	.34	.18

*p<.05, **p<.01

first step explained 33% of the variance in reading scores. After entry of PVST in the next step, the variance explained in total was 35%. To put it differently, OVST provided a great and unique contribution in predicting L2 reading performance. These results further support evidence from previous observations (e.g., Cheng and Matthew 2018) which shows a strong relationship between OVST and RC. This result was also intuitively not surprising that the written text is necessarily composed of orthographic vocabulary. PVST shows a moderate relationship with RC (r = .50) as well, presumably because written word recognition involves activation of phonological information as a prerequisite of meaning access. PVST, nevertheless, was found to have a lower predictive value on L2 reading performance and therefore, OVST alone seems sufficient to explain the variances in reading scores.

In relation to the explanatory power of OVST and PVST on L2 listening, another hierarchical multiple regression analysis was conducted, and the results presented in Table 7. It was revealed that OVST can solely explain 45% of the variance in L2 listening, whereas only minimal contribution (ΔR^2 = .01) was made by adding PVST in the second step.

Table 7. Hierarchical regression analysis addressing the predictive value of vocabulary knowledge on LC

	R	R^2	ΔR^2	Unstandardized		Standardized
				B	SE B	β
Step 1	.67	.45				
Constant				50.44	6.35	
OVST				.66	.13	.67***
Step 2	.68	.46	.01			
Constant				49.08	6.62	
OVST				.56	.19	.56**
PVST				.18	.23	.15

**p<.01, ***p<.001

As can be seen above, the present research was unable to demonstrate the strong relationship between PVST and LC, which is counterintuitive and also contrary to the previous research (Cheng and Matthew 2018; Matthew and Cheng 2015; Mizumoto and Shimamoto 2008). Although students' score on PVST was lower,

their listening performance does not stand in sharp contrast to reading performance. This outcome might be attributed to the possibility that the phonological awrareness might play a less significant role in listening, at least as measured by the currently employed listening comprehension test. It is apparently difficult to conceive of successful listening comprehension without skilled aural word recognition. Nevertheless, it has been suggested that learners can compensate for their insufficient decoding skill through employing high-level skills (Vandergrift 2004). Indeed, researchers have empirically proved that listeners pay less attention to vocabulary from the text that readers do (Reves and Levine 1988). Although this issue is beyond the scope of the current research, it is possible that the learners in the study might have taken advantage of listening strategies (e.g., relying on rhyme, intonation or context) in order to overcome the shortcoming in their phonological vocabulary. Seen from this light, the contrasting finding from Cheng and Matthew (2018) could be connected to the PVST in their study in which the word is given in a sentence context, and therefore it more resembles the processing accompanied by a listening comprehension test.

In conjunction with this explanation, it should be noted that the difficulty level of the listening test is assumed to be quite low, evidenced by a negatively skewed distribution. Possibly, little confusion is likely to occur for learners with the input of dialogue format that was recorded with clear pronunciation and adequate level of speed. Moreover, the students were quite familiar with the test format, which is nation-wide and also employed in CSAT, since they take the same multiple-choice listening test on a semester basis throughout their school program. It is speculated that, therefore, the students could have relied not only on the listening strategies mentioned previously but also on testwise strategies, such as making use of the words in questions or listening to the specific part only. It is highly probable that the test of a more demanding level, in terms of the text difficulty or the test format, would require students to prove a higher level of phonological awareness. At this stage of understanding, it remains unclear whether phonological vocabulary size is indeed not highly predictive of the listening performance.

Lastly, this research found that OVST was better able to explain the variances in both RC and LC scores. Some researchers maintain that measurements for different aspects of vocabulary knowledge should be incorporated into VST in order to enhance the explanatory power of vocabulary in the four skills (Milton 2013). Nonetheless, the result of this study implies that OVST itself appears be a reliable measurement for predicting RC and LC. It strongly correlates with PVST and has

a great explanatory power on both skills. It is admitted that measuring one's vocabulary with a range of aspect in mind is entirely advisable and therefore, students should be encouraged to develop their vocabulary knowledge in a more comprehensive way. When it comes to predicting one's reading or listening proficiency with vocabulary knowledge, however, administering OVST itself seems to yield sufficient information on the proficiency level of L2 reading and listening, at least as measured by the assessments administered in nationwide.

5. Conclusion

The present research sets out to compare the OVS and PVS of Korean EFL students in middle school and to address the relationship between the two different types of VST and L2 reading and listening comprehension. The students were found to have larger orthographic vocabulary knowledge than phonological knowledge; their vocabulary knowledge, regardless of its modality, drastically decreased from 1,000 to 2,000 level and from 3,000 to 4,000 level. Although both types of vocabulary knowledge exhibit a significant correlation with each other and with RC and LC, OVST was shown to be the most predictive in terms of the variance found in both reading and listening. The findings of the study contribute to existing research on L2 vocabulary knowledge by providing further evidence of non-parallel development of phonological and orthographic vocabulary knowledge by EFL middle school students and by suggesting the great predictive value of OVST on the performance in RC and LC tests employed in Korea.

Some limitations of this research are worth mentioning. First, the RC test consists of only a confined number of test items. In order to firmly support the relationship between vocabulary knowledge and RC, alternative comprehension tests with more items should be utilized in the future study. In a similar vein, another type of LC test should be further employed in order to investigate the role of phonological vocabulary knowledge in listening performance. Although the present study failed to find a predictive value of phonological vocabulary test on the listening comprehension test, it is still difficult to completely avoid the conclusions of previous studies that phonemic awareness does contribute to listening performance. One likely explanation could be the test materials employed in the present studies appear to have weakness in assessing one's low-level decoding skill since the students were able to perform the test without well-established phonological vocabulary. Alternatively, Korean

EFL students could be capable of compensating their lack of phonological vocabulary using other strategies. Nonetheless, those issues are beyond the scope of the current research.

Despite of the limitation, the results of the present study provide important pedagogical implication. The most evident outcome in the current study was that a considerable gap exists between OVS and PVS of middle school students. As noted in the national curriculum, the students who are to complete the middle school program should possess about 1,250 words. Nonetheless, it seemed that they achieve such level of vocabulary knowledge only in an orthographic mode. Given the importance of low-level phonological processing in a fluent command of language, both teachers and students should be aware the modality-specific aspects of vocabulary knowledge and accordingly encouraged to teach/learn and practice phonological aspects of lexical items. It is advisable that more spoken inputs and decoding practices are to be employed in classroom in order to raise phonological awareness by EFL students. Another implication is that OVST alone can be utilized to predict the students' performance in RC and LC test, specifically, those of which are administered in national achievement exams in Korea. It is to be hoped that this study lays the groundwork for future research into L2 learners' vocabulary size development and its relationship with L2 language skills.

References

Abbott, R. D. and Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary-and intermediate-grade writers. *Journal of Educational Psychology* 85(3), 478-508.

Al-Hazemi, H. (2000). Lexical attrition of some Arabic speakers of English as a foreign language: A study of word loss. *The Internet TESL Journal* 5(12). Retrieved from http://iteslj.org/Articles/Al-Hazemi-Attrition/

Allor, J. H. (2002). The relationships of phonemic awareness and rapid naming to reading development. *Learning Disability Quarterly* 25(1), 47-57.

Bejar, I., Douglas, D., Jamieson, J., Nissan, S. and Turner, J. (2000). *TOEFL 2000 Listening Framework*. Princeton, NJ: Educational Testing Service.

Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening* 14(1), 14-31.

Carroll, J. M., Snowling, M. J., Stevenson, J. and Hulme, C. (2003). The development of phonological awareness in preschool children. *Developmental Psychology* 39(5), 913-923.

Chang, A. (2007). The impact of vocabulary preparation on L2 listening comprehension, confidence and strategy use. *System* 35(4), 534-550.

Cheng, J. and Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing* 35(1), 3-25.

Cheung, H. (2007). The role of phonological awareness in mediating between reading and listening to speech. *Language and Cognitive Processes* 22(1), 130-154.

Coxhead, A. and Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing* 16, 129-147.

Diamond, K. E., Gerde, H. K. and Powell, D. R. (2008). Development in early literacy skills during the pre-kindergarten year in Head Start: Relations between growth in children's writing and understanding of letters. *Early Childhood Research Quarterly* 23(4), 467-478.

George, D. and Mallery, P. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update.* (10th ed.). Boston, MA: Allyn & Bacon.

Field, J. (2008). Bricks or discovering statistics using SPSS for Windows mortar: which parts of the input does a second language listener rely on? *TESOL Quarterly* 42(3), 411-432.

Fountain, R. L. and Nation, I. S. P. (2000). A vocabulary-based graded dictation test. *RELC Journal* 31(2), 29-44.

Giambo, D. A. and McKinney, J. D. (2004). The effects of a phonological awareness intervention on the oral English proficiency of Spanish-speaking kindergarten children. *TESOL Quarterly* 38(1), 95-117.

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System* 28(1), 55-75.

Kamimoto, T. (2005). The effect of guessing on vocabulary test scores: A qualitative analysis. *Paper presented at The European Second Language Association (EuroSLA) Conference 15*, Dubrovnik, Croatia.

Kelly, P. (1991). Lexical ignorance: The main obstacle to listening comprehension with advanced foreign language learners. *IRAL* 29(2), 135-149.

Koda, K. (1989). The effects of transferred vocabulary knowledge on the development of L2 reading proficiency. *Foreign Language Annals* 22, 529-540.

Kremmel, B. and Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly* 13(4), 377-392.

Kremmel, B. and Schmitt, N. (2018). Vocabulary levels test. In J. I. Liontas, M. DelliCarpini, & J. C. Riopel, eds., *The TESOL Encyclopedia of English Language Teaching*. Hoboken, NJ: John Wiley &Sons. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118784235.eelt0499

Laufer, B. and Nation, P. (1995). Vocabulary size and use lexical richness in L2 written production. *Applied Linguistics* 16, 307-322.

Laufer, B. and Revenhorst-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.

Lee, J. (2018). Relationship between vocabulary size, syntactic awareness and reading comprehension in low- and high-level Korean high school EFL students. *Secondary English Education (SEE)* 11(3), 3-23.

Li, M., Cheng, L. and Kirby, J. R. (2012). Phonological awareness and listening comprehension among Chinese English immersion students. *International Education* 41(2), 46-65.

Li, M. and Kirby, J. R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics* 36(5), 611-634.

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

McLean, S., Kramer, B. and Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research* 19(6), 741-760.

Meara, P. (1992). *EFL Vocabulary Tests*. New York: ERIC Clearinghouse.

Meara, P. and Milton, J. (2003). *X_Lex: The Swansea Levels Test*. Newbury: Express.

Meara, P. and Wolter, B. (2004). V_Links: Beyond vocabulary depth. *Angles on the English Speaking World* 4, 85-96.

Metsela, J. L. and Ehri, L. C. (1998). *Word Recognition in Beginning Literacy*. Mahwah, NJ: Lawrence Erlbaum Associates.

Milton, J. L. (2007). Lexical profiles, learning styles and the construct validity of lexical size tests. In H. Daller, J. Milton & J. Treffers-Daller, eds., *Modelling and Assessing Vocabulary Knowledge* (pp. 47-58). Cambridge: Cambridge University Press.

Milton, J. L. (2009). *Measuring second language vocabulary acquisition*. Bristol, UK: Multilingual Matters.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindquist, & B. Laufer, eds., *L2 Vocabulary Acquisition, Knowledge, and Use: New Perspectives on Assessment and cOrpus Analysis* (pp. 57-78). Amsterdam, the Netherlands: European Second Language Association.

Milton, J. and 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.

Milton, J., Wade, J. and Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, M. Torreblanca-López, & M. D. López-Jiménez, eds., *Further Insights into Nonnative Vocabulary Teaching and Learning* (pp. 83-97). Bristol: Multilingual Matters.

Mizumoto, A. and Shimamoto, T. (2008). A comparison of aural and written vocabulary size of Japanese EFL university learners. *Language Education and Technology* 45, 35-51.

Nassaji, H. and Geva, E. (1999). The contribution of phonological and orthographic processing skills to adult ESL reading: Evidence from native speakers of Farsi. *Applied Psycholinguistics* 20, 241-267.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. New York: Cambridge

University.
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. *International Journal of Innovation in Language Learning and Teaching* 1(1), 2-13.
Palmberg, R. (1987). Patterns of vocabulary development in foreign language learners. *Studies in Second Language Acquisition* 9, 201-220.
Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review* 56(2), 282-308.
Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning* 52(3), 513-536.
Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
Reves, T. and Levine, A. (1988). The FL receptive skills: Same or different? *System* 16(3), 327-336.
Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research* 12(3), 329-363.
Schmitt, N. (2010). *Researching vocabulary: A Vocabulary Research Manual*. Basingstoke, UK: Palgrave Macmillan.
Schmitt, N., Jiang, X. and Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal* 95(1), 26-43.
Shillaw, J. (1999). *The Application of the Rasch Model to Yes/No Vocabulary Tests*. (Unpublished doctoral dissertation). University of Wales, Swansea.
Staehr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal* 36(2), 139-152.
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.
Stahl, S. A. and Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology* 86(2), 221-234.
Tsui, A. B. and Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics* 19(4), 432-451.
Yi'an, W. (1998). What do tests of listening comprehension test?: A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing* 15(1), 21-44.
Yim, S. (2007). Globalization and language policy in South Korea. In A. Tsui, & J. Tollefson, eds., *Language Policy, Culture, and Identity in Asian Contexts*, pp. 37-53. Mahwah, NJ: Lawrence Erlbaum Associates.
Yopp, H. K. (1992). Developing phonemic awareness in young children. *Reading Teacher* 45(9), 696-703.
Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics* 24(1), 3-25.
Van Zeeland, H. and Schmitt, N. (2012). Lexical coverage in L1 and L2 listening

comprehension: The same or different from reading comprehension? *Applied Linguistics* 34(4), 457-479.
Vassiliu, P. (1994). *Vocabulary Profiles: A Necessary Prerequisite for Course Book Selection.* (Unpublished MA thesis), University of Wales, Cardiff.

Myoungjin Kim
PhD Candidate
Department of English Language Education
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
E-mail: lucy84@snu.ac.kr

Received: February 21, 2019
Revised version received: Apri 9, 2019
Accepted: Apri 9, 2019