English Language & Literature Teaching, Vol. 18, No.1 Spring 2012

Vocabulary Size of Korean EFL University Learners: Using an Item Response Theory Model*

Yongsang Lee**
(KICE)
Yuah V. Chon***
(Hanyang University)
Dongkwang Shin
(KICE)

Lee, Yongsang, Chon, Yuah V., & Shin, Dongkwang. (2012). Vocabulary size of Korean EFL university learners: Using an item response theory model. *English Language & Literature Teaching*, *18*(1), 171-195.

While noticing that there is insufficient interest in the assessment of EFL learners' vocabulary levels or sizes, the researchers developed two tests identical in form (Forms A and B) to assess the lexical knowledge of Korean university learners at the 1st ~ 10th 1,000 word bands by adapting a pre-established vocabulary levels test (VLT). Of equal concern was to investigate if the VLT was equally a valid and reliable instrument to be used on measuring the lexical knowledge of EFL learners. The participants were 804 university freshmen enrolled in a General Education English Course from four different colleges. The learners were asked to respond to either Form A or B. While scores generally fell towards the lower frequency bands, multiple regression found the Korean College Scholastic Ability Test (CSAT) to be a significant variable for predicting the learners' vocabulary sizes. From a methodological perspective, however, noticeable differences between Forms A and B could be found with item response theory analysis. The findings of the study provide suggestions on how future VLT for testing EFL learners may have to be redesigned.

[vocabulary size/vocabulary levels test/background variable/item difficulty/item response theory/Rasch model]

* This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government. (NRF-2010-332-B00424)
** First author
*** Corresponding Author

교보문고 KYOBO Book Centre

I. INTRODUCTION

Vocabulary knowledge has long been recognized as central to language acquisition and use (Alderson, 2000; Schmitt, 2000). For L2 learners who wish to express themselves in their target language, a size of 2,000 words is said to be a realistic goal, (Schmitt, 2000). For those intending to read authentic texts, a vocabulary threshold of 3,000 – 5,000 word families is needed (Nation & Waring, 1997). For more challenging materials such as university textbooks that attend to specialized vocabulary, learners would require knowledge of 10,000 word families (Hazenberg & Hulstijn, 1996). Incidentally, having a large vocabulary size is useful because the learner will be able to cope with a variety of receptive and productive tasks needed for academic and vocational purposes.

Regarding the assessment of vocabulary knowledge, Wesche and Paribakht (1996) provide a detailed analysis of techniques used for second language vocabulary assessment and the implied assumptions about the underlying construct of vocabulary knowledge. Read (2000), who has worked on developing vocabulary depth tests (that usually involve testing a limited number of words), has pointed out that the vocabulary size or levels test can be a more representative picture of the overall state of the learner's vocabulary in spite of its criticisms on superficiality. Another observation that can be made about the current status of research on vocabulary assessment is that for measuring vocabulary size and depth, vocabulary tests have been widely used as research tools rather than as assessment instruments. As such, this study takes interest in the use of the vocabulary size tests *per se* to measure the vocabulary size of L2 learners, which as a result is expected to provide criteria for the placement or admission of learners in language teaching programs. At the same time, a secondary interest of the present study was to see if any of the learner difference variables (i.e., gender, overseas experience, major, Korean College Scholastic Ability levels, learning styles) was significant for explaining the learners' vocabulary size.

The present study is, in fact, a follow-up from Chon and Shin's (2012) study that measured the vocabulary size of 403 Korean EFL university learners at 10 word bands of the British National Corpus (BNC) so that the test, in effect, is more appropriately labeled the vocabulary levels test (VLT). In addition to the data collected in the previous study, we continued to collect information on the L2 university learners' vocabulary size, however, this time additionally with another VLT which had been developed in the same way as the previous VLT. Therefore, a third interest of the present study was to find out if the newly developed VLT would produce similar results in subsequent tests. The information gained from this kind of validation process would help practitioners and researchers to be able to develop different versions of the VLT that are valid and

교보문고 KYOBO Book Centre

appropriate to different educational contexts or vocabulary learning goals. In sum, the VLT was equally a valid and reliable instrument to be used on measuring the lexical knowledge of EFL learners

II. BACKGROUND

1. Measurement on Vocabulary Size of ESL/EFL Learners

There has been much and various research on the measurement of learners' vocabulary levels or size (Fan, 2000; Goulden, Nation & Read, 1990; Gui, 1982; Laufer & Nation, 1999; Laufer & Paribakht, 1998; Morgan & Oberdeck, 1930; Nation, 1990; Nation & Belgar, 2007; Schmitt, Schmitt & Clapham, 2001; Waring, 1997; Webb, 2008). Many of those studies, however, have dealt with English native speakers or ESL learners rather than EFL learners (e.g., Chon, 2011; Son, 2007). While there is focus on receptive vocabulary tests, there are few studies that compare the receptive and productive vocabulary sizes of EFL learners, and those studies are restricted to small numbers of participants.

Laufer and Paribakht (1998) compared the vocabulary size of EFL learners in Israel to ESL learners in Canada at different levels of English proficiency. They found that the passive (i.e., receptive) vocabulary size was always significantly larger than the active (i.e., productive) vocabulary size, and these results agree with similar studies (Morgan & Oberdeck, 1930; Shin, Chon, & Kim, 2011; Waring, 1997; Webb, 2008). The reason for the passive-active vocabulary gap was believed to be due to the EFL learners' smaller passive vocabulary size consisting largely of high frequency words (i.e., within the first 2,000 words). The EFL learners invested more time and effort than the ESL learners to acquire a similar amount of passive vocabulary through deliberate learning (e.g., using word cards) rather than through incidental learning (e.g., guessing from context). However, such inferences proposed by Laufer and Paribakht (1998) may fail to explain the reason for the passive-active vocabulary gap with respect to EFL learners from differing educational backgrounds which would quite probably have a significant influence on their vocabulary size. As such, consideration of EFL learners' educational and personal variables need to be included in the interpretation of vocabulary sizes, and also even in the process of adapting pre-existing instruments that intend to assess some aspect of EFL learners' lexical knowledge. For instance, in EFL contexts where English is learnt as a foreign language by following a national curriculum, there would be differing focuses for vocabulary learning depending on the differing educational goals and environments.

교보문고 KYOBO Book Centre

Shin, Chon, and Kim (2011) within an EFL context measured receptive and productive vocabulary sizes of Korean high school students. They tested three different English proficiency groups (i.e., high, middle, and low) and found that the receptive vocabulary size (henceforth RVS) was significantly larger than productive vocabulary size (henceforth PVS). The authors found that particular words were familiar to the learners regardless of proficiency groups. One of the explanations was identified to be the learners' knowledge of loanwords, and the test-oriented characteristic of Korean students. The study implies how various personal background variables may influence the formation of lexical knowledge and the need to interpret scores of vocabulary size or levels based on consideration of those variables. Concurrently, we became doubtful as to if the pre-existing instruments based on word frequency information of native-speaker corpus was equally valid for measuring EFL learners' vocabulary sizes. This study, as mentioned previously, pursues to test this out by comparing two vocabulary levels test identical in form to make note of findings for the future development of instruments for assessing vocabulary knowledge, particularly for those in the EFL contexts. For this purpose, this study adopts the item response theory (IRT) to compare EFL learners' vocabulary levels using two equivalent test forms. The IRT would enable us to investigate EFL learners' vocabulary levels more precisely by taking into account the measurement errors that may occur across test scores.

Since the focus of our study is to ultimately examine the validity and reliability of the vocabulary levels test that are based on the word level of native speakers, we review, in the following, literature on the different aspects of vocabulary tests and studies that have dealt with measuring vocabulary knowledge. Also, we provide an overview of item response theory in connection to our interest in analyzing the differences in item difficulty of two vocabulary levels tests.

2. Vocabulary Assessment

To understand the assessment of vocabulary sizes, it is necessary to examine vocabulary testing formats in which various aspects of words are measured. According to Read's (2000) model, vocabulary testing formats can be broadly classified into three dichotomous dimensions as shown in Figure 1.

A *discrete* test takes vocabulary knowledge as a distinct construct, separated from other components of language competence. In fact, vocabulary testing can be discrete when the target words are highlighted within a given sentence. In contrast, vocabulary can be measured as part of language competence; in this case, an *embedded* vocabulary measure contributes to the assessment of a larger construct. An embedded measure can

교보문고 KYOBO Book Centre

be found, for instance, in a reading task, which is followed by a number of items assessing the learners' understanding of particular words or phrases.

FIGURE 1

Dimensions of Vocabulary Assessment (Read, 2000, p. 9)

Discrete A measure of vocabulary knowledge or use as an independent construct	↔	**Embedded** A measure of vocabulary which forms part of the assessment of some other, larger construct
Selective A measure in which specific vocabulary items are the focus of the assessment	↔	**Comprehensive** A measure of the vocabulary content of the input material (reading/listening tasks) or test-takers' responses (writing/speaking tasks)
Context-independent A vocabulary measure in which the test-taker can produce the expected response without referring to any context	↔	**Context-dependent** A vocabulary measure which assesses the test-taker's ability to take into account contextual information in order to produce the expected response

The second dimension concerns the range of vocabulary to be included in assessment. Most receptive vocabulary tests are *selective.* The term does not imply a multiple-choice test, but that a vocabulary test is based on a set of target words selected by the test-writer while considering some characteristics of the words. The target words may either be selected as individual words and then incorporated into separate test items, or alternatively the test-writer can first choose a suitable text and then use certain words from it as the basis for vocabulary assessment.

In comparison, a *comprehensive* measure takes account of all the vocabulary content of spoken or written text. It is an indirect measure of vocabulary through the analysis of language produced in test-takers' responses, such as in speaking tasks. In this way, it may be difficult to precisely predict the kinds of vocabulary the test-takers produce so that controlling or selecting the target words may not be an easy task *in comprehensive* measures. A comprehensive measure can also be applied to the input material for reading or listening tests, for instance, when test-writers use a readability measure as one way of judging the suitability of a text for the assessment of a particular group of test-takers (see Read, 2000 for details).

The third dimension concerns the role of context. In the traditional sense, contextualization has meant that a word is presented to test-takers in a sentence rather than as an isolated element. An L2-L1 translation task, for example, is completely *context-independent*. In comparison, a test is *context-dependent* when test-takers are prompted to define the target word in a passage for an exact sense of the meaning.

교보문고 KYOBO Book Centre

Context, then, may be more appropriate for measuring real vocabulary knowledge. In sum, the format utilized in the present study to measure the vocabulary size of EFL learners is based on Nation (2010), and Laufer and Nation (1999) which tests *discrete*, *selective*, and *context-dependent* vocabulary knowledge. The following presents the vocabulary format used for assessing receptive and productive vocabulary knowledge:

Receptive Test

• Fourth 1000 Level

2. latter: I agree with the **latter**. (Answer: c)

a. 교회에서 온 남자 (= man from church)
b. 주어진 이유 (= reason given)
c. 마지막 것 (후자) (= the latter)
d. 대답 (= response)

(Nation, 2010)

Productive Test

• The 5,000 Word Level

6. The thieves have forced an ent_____ to the building.

(Laufer & Nation, 1999)

Nation's (2009) vocabulary size test, also broadly used, is one of the latest versions of the vocabulary tests that exist. However, as will be explained later (see later *Procedure*), we modified Nation's vocabulary testing format by considering some of the characteristics of EFL learners. For measuring receptive vocabulary size (RVS), the L2 target words were randomly selected in advance and highlighted in the sample sentences. For the measurement of learners' productive vocabulary size (PVS), we newly developed them in the form of C-tests (filling-in-the-blank) by also providing an L1 translated sentence where the L1 equivalent of the L2 target words were highlighted (See later *Instruments* for details).

3. Previous Vocabulary Test Formats for Receptive and Productive Vocabulary Sizes

One widely used measure to assess the size of vocabulary knowledge in previous studies is the Vocabulary Levels Test (henceforth VLT). The VLT has a word-meaning matching format consisting of words representing different word levels (e.g., 2000, 3000, 5000 word levels). There have also been some trials attempting to measure the gap

교보문고 KYOBO Book Centre

between the receptive and the productive vocabulary size by use of the VLT. In those studies, however, several problems have been overlooked.

In Waring's (1997) study, for instance, the receptive test - a meaning recognition test - may be considered to have problems with construct validity. In the receptive test format, which is based on one of Nation's test formats (1990), an item tries to test six words at the same time. This condition results in items being excessively difficult and possibly leading to an underestimation of the learners' actual vocabulary size (Webb, 2008). In addition, the productive version, which has a C-test format, sometimes requires test-takers to recognize the collocational relationships of words (e.g., *earn a salary, wind roar*); this is beyond single-word-level knowledge. Keshavarz & Salimi (2007) demonstrated the strong relationship between a learner's performance on a cloze test and his or her collocational competence; they found that the cloze test format is more difficult for EFL learners. A problem of test reliability may be related to test administration.

Webb (2008) also investigated the relationship between the receptive and productive vocabulary sizes among EFL university learners. Webb adapted equivalent receptive and productive test formats in the form of translation tests with different receptive and productive target words in order to provide more reliable results by overcoming the flaws of previous VLTs. In his study, Webb attributed the differences between the receptive and productive vocabulary sizes to the smaller vocabulary size of EFL learners than those of ESL learners. Webb interprets that EFL learners are likely to learn more words through explicit instruction than are ESL learners (Laufer & Paribakht, 1997). The explanation made by Webb is nevertheless not completely valid since there are variables in different EFL educational contexts that need to be considered when selecting target words for a vocabulary test.

4. Item Response Theory (IRT) Model

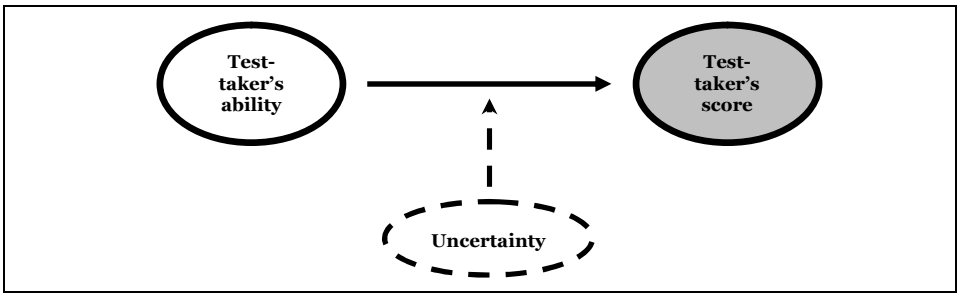
Item response theory (IRT) arose from a psychometric need to provide test designers with valid and reliable testing tools since classical test theories did not consider the measurement error across test scores (Henard, 1998). IRT contains a large family of models and the Rasch model is one of these models. The Rasch model is a measurement model which shows the probability of the correct answer by considering the learner's ability and the item difficulty. This Rasch model is processed under two key concepts: local independence and unidimensionality. Unidimensionality demands that only one ability or latent trait be measured by the various items that make up the test form. This

교보문고 KYOBO Book Centre

unidimensionality is very commonly adopted by test constructors since they usually attempt to construct unidimensional tests in which test scores are easier to interpret (Hambleton & Swaminathan, 1985). The other concept, local independence, demands that test-takers' responses to various items in a test are statistically independent of each other (Hambleton & Swaminathan, 1985). With respect to these two concepts, Hambleton and Swaminathan pointed out some of the advantages of the Rasch model. First, the estimate of the test-taker's ability is independent of the particular item in the test form. Second, the descriptor of each test item is also independent of the particular item which is used for the purpose of item calibration. Essentially, the test-taker's language ability is a latent trait which cannot be directly measured. Therefore, the test-taker's score does not directly represent his or her ability. This means that the test-taker's score should be interpreted with measurement tools, but the score may not correspond exactly to the test-taker's actual ability. Figure 2 represents out illustration of how uncertainty exists between the test-taker's ability and score.

FIGURE 2

The relationship between test-takers' abilities and scores



This uncertainty influences the discrimination of the test-takers' abilities; it could be indicated by a probability. Test-takers' responses could also be anticipated with this probability and could be estimated by an IRT model. In the present study, we will examine whether the native speaker's version of the VLT is applicable to the EFL context by analyzing the differences in item difficulty of two VLT forms with an IRT model.

5. Statement of the Problem

As suggested in the related literature, existing studies have limitations, such as, in lacking construct validity, or in neglecting to consider EFL learners' unique educational

교보문고 KYOBO Book Centre

background variables when measuring vocabulary knowledge. The present study, therefore, attempts to overcome these flaws, and the three issues that are raised are: First, what are the receptive and productive vocabulary sizes of Korean EFL university learners? Second, what are the factors that may affect the receptive and productive vocabulary sizes of Korean EFL learners? Third, is it valid and reliable to apply the native speaker's version of the VLT to Korean EFL learners? In the present study, these issues were investigated by proposing valid and reliable measurement tools such as with IRT.

III. METHOD

1. Context and Participants

We measured the vocabulary size of EFL university participants in Seoul, Korea. There were 804 participants in the study, and all the participants were in their first year with an average of ten years instruction in English since elementary school. At the time of the study, the students were enrolled in Freshmen English as a requirement of their English language course in the General Education English Program. The students in the General Education English Program were deemed appropriate for measuring the vocabulary size of university learners since the students in the program were not restricted to certain majors, and their age and amount of instruction for English could be controlled in spite of the large size of the participants. The 804 participants were randomly assigned to take either form A or B of the vocabulary levels tests (See later *Procedure* for details). As shown in Table 1, the participants were from 20 different departments and they could be classified into four sub-groups according to the schools they were attending: (1) School of Humanities, (2) School of Social Sciences, (3) School of Natural Sciences, and (4) School of Engineering. This was in accordance with how the university recognized the departmental categories.

Sub-group 1 consisted of the students in the School of Humanities which includes departments, such as, Korean Language and Literature, History, and Philosophy. Of the 403 participants who were tested with form A, 104 were included in sub-group 1, and 103 of the 401 participants who took form B were also a part of sub-group 1. Although there are slight differences in the number and departmental categories in the composition of the participants between those who took form A and B, in the present study we assume the two groups to be homogeneous because the grouping was based on random selection.

교보문고 KYOBO Book Centre

TABLE 1

Composition of Participants

Group	Sub-group 1	Sub-group 2	Sub-group 3	Sub-group 4
No. of Participants	N=104 (A) N=103 (B)	N=74 (A) N=52 (B)	N=97 (A) N=101 (B)	N=128 (A) N=145 (B)
Schools	Humanities	Social Sciences	Natural Sciences	Engineering
Departments	• Korean Language & Literature • History • Philosophy • Religious Studies • English Literature & Linguistics • American Culture • German Culture • French Culture • Chinese Culture	• Sociology • Political Science • Psychology	• Mathematics • Physics • Chemistry • Life Science	• Electronic Engineering • Computer Science & Engineering • Chemical and Biomolecular Engineering • Mechanical Engineering

* *Note*: (A) = Form A; (B) = Form B

2. Instruments

1) Background Variables

In connection to our interest in investigating the relationship between personal background variables and the EFL learners' vocabulary sizes, we included five variables. They were (1) gender, (2) overseas experience, (3) major (school), (4) College Scholastic Ability Test (CSAT) level, and (5) vocabulary learning style. Learners were asked to mark the personal variables before solving the items on the vocabulary levels test. If the learners had had overseas experience, the learners were asked to mark so. The learners were also asked to state their area of study, which could be marked within one of the four categories: (1) School of Humanities, (2) School of Social Sciences, (3) School of Natural Sciences, and (4) School of Engineering. For the marking of CSAT levels (i.e., nine levels), learners were asked to state the level they had obtained to enter the university. The learners were also asked to mark the method they had used for vocabulary learning, which were studying vocabulary (1) during class, (2) through vocabulary exercise books, (3) through lectures at private institutes (e.g., cram schools), (4) miscellaneous categories, and (5) through incidental acquisition (e.g., reading).

2) The Receptive Vocabulary Size Test

In the study, we adapted the bilingual version of Nation's (2010) vocabulary levels

교보문고 KYOBO Book Centre

test[1] (VLT) to measure the receptive vocabulary size of Korean university students. That is, since an English (monolingual) version of the test, and the English words in the multiple choices would have become another interfering variable for measuring the learners' vocabulary size, the bilingual version was chosen for the study. The VLT was developed by selecting 10 word items at each band from the 1st to the 10th 1,000 word families which is based on the fourteen 1,000 British National Corpus (BNC) word lists by Nation and Beglar (2007). In the list of words, each word would roughly represent a sample of 100 words at each level, and may be used for estimating the total vocabulary size of the learners (i.e., by multiplying a 100 to the learner's total score). Ten of the fourteen bands were used in our study to represent 10,000 words, which was considered sufficient for measuring the receptive vocabulary size of EFL learners; according to Nation (2006), non-native-speaking PhD students have approximately a 9,000 word family vocabulary so that our selection of target vocabulary size was deemed appropriate. In the final stages of constructing the vocabulary level test, we included the 100 words randomly selected from 10 frequency bands. For each frequency band, we randomly reselected words with *Random Item Generator* v.1 available at *Compleat Lexical Tutor* (http://www.lextutor.ca/), which draws its randomizations from the 14 BNC lists. In this way, forms A and B were newly developed. A sample item of the receptive vocabulary levels test is presented as follows:

30. lonesome: He felt **lonesome**. (Answer: ③ 외로운)
① 은혜를 모르는 ② 피곤한 ③ 외로운 ④ 힘이 넘치는 ⑤ 잘 모르겠음
(① Not thankful, ② tired, ③ lonely, ④ overly energetic, ⑤ Do not know)

In the study, the choice ⑤ '잘 모르겠음' (i.e., *Do not know*) among the options was also added to provide a more reliable measure by reducing the possibility of guessing by the learners. The test not only provides the total vocabulary size of the learners, but also the vocabulary size at each word band. Thus, the diagnostic trait of the test is reinforced.

3) The Productive Vocabulary Size Test

The adapted version of the productive vocabulary test was based on the format of Laufer and Nation's (1999) productive vocabulary test. There are few testing formats for measuring controlled productive vocabulary size, and for the same reason has been broadly used. The original test assesses the 2,000, 3000, 5000, University Word List, and

[1] Nation's (2010) vocabulary size test is downloadable from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

교보문고 KYOBO Book Centre

10,000 word levels. Each level consists of 18 items in the type of C-tests by providing 2-4 of the initial letters of the target word. The test may be criticized, however, for not just assessing knowledge of single word items, but also collocational or structural knowledge. In our study, to overcome such shortcoming of the productive levels test, a new productive vocabulary size test was developed for the 1st to the 10th 1,000 levels with 10 items at each level totaling 100 items. The 100 lexical items were randomly selected by use of *Random Item Generator* v.1 in the same way as in the receptive VLT. The following is a sample item of the productive vocabulary size test:

42. 우리의 중요한 **미덕**은 손윗사람을 공경하는 것입니다. (Answer: virtue)
☛Our main vi_ _ _ _ is respect for elders.

Our productive vocabulary size test was kept in the C-test format as in the original version of Laufer and Nation's productive vocabulary levels test; however, we added L1 sentence equivalents to L2 sentences, which enabled learners to more easily retrieve the L2 target word by referring to the accompanying L1 translation. This was also expected to reduce the burden that a word item may require of the learners' collocational knowledge. In addition, to make the L2 target word salient, the L1 target word was highlighted in boldface, and blank spaces were provided with the first one or two letters of the target word to avoid the probability of use of a synonym in relation to the L1 equivalent presented. One letter was provided for a one syllable word, while two letters were provided for words with two or more syllables. To control the words making up the prompts in the tests (e.g., the words 'look', 'at', and 'the' from *Look at the b_ _ _*.), we restricted those words to the 1st 1,000 word level. Thus, Forms A and B of the productive vocabulary size test were developed as in the receptive test.

4) The Item Response Theory (IRT) Analysis of Vocabulary Test Items

Since an area of interest for us was to investigate whether item difficulties differ between forms A and B for both the receptive and productive vocabulary size tests, the item response theory (IRT) model was applied to the data from these two forms. For the data analyses, the IRT analysis program of ConQuest (Wu, Adams, & Wilson, 1988) was used. As mentioned previously, the advantage of the IRT model is in the consideration of the measurement error across scores. According to the previous classical test theory (CTT), it is assumed that observed test scores represent EFL learners' vocabulary levels, but these scores are often affected by measurement situations. In comparison, the type of assessment used in our study may be difficult or challenging enough to discriminate high vocabulary level of EFL learners, whereas it

교보문고 KYOBO Book Centre

may not discriminate low vocabulary level of EFL learners. As a result, the low proficiency EFL learners, in spite of belonging to the same vocabulary level, may obtain different test scores. The observed score, in fact, are expected to have more errors (i.e., measurement errors) in the low range of observed scores, but the CTT assumes that this error is uniform across the entire range of observed scores. The IRT model, however, makes it clear that the error can vary based on the test information function. The IRT model estimates EFL learners' latent trait, which is the vocabulary level in this study, based on the probability of a correct response to an item. This probability is used to obtain the item information function, which can be calculated to provide the size of measurement error at a given vocabulary level. As such, we were able to obtain unbiased estimates of EFL learners' vocabulary level across observed scores and the item difficulties. In addition, since we sampled different groups of students with different test forms, the IRT model allowed us to compare items from the different forms under the assumption that the two groups have essentially the same abilities. With the use of such sampling method and sample size, the assumption could be easily satisfied.

3. Procedures

For the administration of the vocabulary levels test (VLT), a majority of the faculty members in the General Education English Program (GEEP) at the university were recruited to help in disseminating the test papers for the vocabulary levels test. Permission was granted by the Head of GEEP, and in total eighteen instructors agreed to cooperate in the study. Before the actual administration of the VLT, the researcher prepared sufficient copies of both Forms A and B respectively in equal numbers so that the instructors only had to pick up enough number of copies to their classes. VLT were administered in the first and second weeks of the Spring Semester of 2011 during the learners' regular class periods. The learners were provided adequate time for them to complete the test papers. By the end of the data collection period, 804 test papers were returned to the researchers.

4. Data Analysis

The responses collected from the learners via the VLT were coded and analyzed with SPSS (Statistical Package for Social Sciences) 17.0. The frequencies and means were calculated at each of the 10 word bands for both the receptive and productive tests. For the analyses with learner's background variables, multiple regression with the enter method was conducted to examine the relationship between learners' background variables and vocabulary level of the learners. In the next step of the analysis, the Rasch model, which is

교보문고 KYOBO Book Centre

one of the IRT models, was applied to EFL learners' response data to estimate item difficulties. In the Rasch model analyses, item difficulties in two forms were estimated assuming that the distributions of EFL learners' proficiency for the two test forms are the same, and then compared to investigate whether two test forms are equivalent.

IV. RESULTS AND DISCUSSION

1. Receptive and Productive Vocabulary Sizes of EFL Learners

The receptive and productive vocabulary sizes of Korean EFL university learners, investigated via Forms A and B of the vocabulary levels test yielded different receptive and productive vocabulary sizes at the 1st to the 10th word frequency bands as shown in Table 2.

TABLE 2
Receptive and Productive Vocabulary Sizes at Different Word Frequency Bands

Word Level	Form A (N=403)			Form B (N=401)		
	RVS	PVS	PVS/RVS ratio	RVS	PVS	PVS/RVS ratio
1st 1000	9.87	9.44	95.65	9.38	8.86	94.39
2nd 1000	9.84	8.29	84.19	9.31	5.96	64.02
3rd 1000	9.46	7.18	75.85	7.81	4.10	52.54
4th 1000	8.60	5.24	60.88	7.31	4.51	61.69
5th 1000	7.20	4.97	69.01	6.41	1.79	27.93
6th 1000	8.08	4.55	56.33	3.93	3.33	84.94
7th 1000	6.96	2.60	37.29	4.68	0.90	19.20
8th 1000	5.06	2.40	47.35	4.67	1.00	21.52
9th 1000	4.42	1.69	38.29	2.55	0.09	3.53
10th 1000	4.74	1.22	25.81	3.72	1.04	27.93
Total	74.23	47.57	64.08	59.77	31.59	52.85

Although we randomized target words from the 10 different frequency word bands to develop two homogeneous test forms, a noticeable gap, as shown in Table 2, was found between the results of the two forms (except at the three word bands: 1st 1,000, 4th 1,000, and 10th 1,000). When the participants' total scores were calculated for the vocabulary sizes, the RVS for learners who took Form A was almost 7,400 and about 4,800 for PVS. In comparison, the RVS of those learners who took form B was almost 6,000 and 3,200 for PVS. In addition, the calculation with Forms A and B respectively resulted in 64% and 53% in PVS/RVS ratios. In spite of our effort to present target words by randomly having selected words for the respective word bands (i.e., 1st ~ 10th 1,000) (Nation & Beglar, 2007), we attribute the difference between Forms A and B to variables that we may have

교보문고 KYOBO Book Centre

not been able to control (e.g., test takers' cultural or educational backgrounds).

In addition, sudden rises in low frequency word bands (i.e., Form A: 6th 1,000; Form B: 7th & 8th 1,000 word levels) were also observed. As in a previous study (Shin, Chon, & Kim, 2011), we may be attribute this to pressure on students to learn and memorize relatively lower frequency (or academic words) rather than the higher frequency words that are likely to be needed more for the reading and writing of English general texts. Nevertheless, when we consider that Japanese EFL women university learners' vocabulary sizes - which were measured for 1st, 2nd, 3rd and 5th 1,000 word bands in Waring's (1997) study - demonstrated a mean PVS/RVS ratio of 44.8% for the four word bands, the Korean EFL university learners seem to have higher PVS/RVS ratios: 64% and 53%. In the following section, the test takers' personal backgrounds variables, which may have affected the results of the two test forms, will be examined in more detail.

2. Personal Backgrounds Affecting the Receptive and Productive Vocabulary Sizes

In this study, we examined five variables which may have affected the receptive and productive vocabulary sizes. The five factors were (1) gender, (2) overseas experience, (3) major (school), (4) CSAT level, and (5) learning style.

First, as mentioned previously, the 804 participants who took Form A consisted of 403 university students (male: 225; female: 178). Participants who took Form B consisted of 401 students (male: 207; female: 194). Second, 76 of the 403 students who had taken Form A had lived in English speaking countries for six or more months, and 88 of the 401 students who had taken Form B also had overseas experience. Third, as mentioned previously (see previous Table 1), the participants were classified into four sub-groups according to the schools they were attending: (1) School of Humanities, (2) School of Social Sciences, (3) School of Natural Sciences, and (4) School of Engineering. Table 3 shows the numbers of students based on the two categories: CSAT level and learning style.

TABLE 3
No. of Participants Classified by CSAT Level and Learning Style

Form A				Form B			
CSAT	Freq	Learning Style	Freq	CSAT	Freq	Learning Style	Freq
Level 1	225	Style 1	113	Level 1	225	Style 1	113
Level 2	113	Style 2	112	Level 2	123	Style 2	117
Level 3	47	Style 3	83	Level 3	32	Style 3	67
Level 4	11	Style 4	25	Level 4	17	Style 4	24
Level 5	7	Style 5	70	Level 5	4	Style 5	80
Total		403		Total		401	

* *Note*: Style 1 = during class, Style 2 = through vocabulary exercise books, Style 3 = through lectures at private institutes (e.g., cram schools), Styles 4 = miscellaneous categories, and Style 5 = through incidental acquisition

교보문고 KYOBO Book Centre

In the study, proficiency levels from the English section of the College Scholastic Ability Test[2] (CSAT) administered in November 2010 were used as a measure of the students' overall English proficiency. All the participants scored within one of the 5 levels of the CSAT, which indicates that the participants were from high-ranking high schools in Korea. The variable, vocabulary learning style, contained five different vocabulary learning styles when the learners were asked how they had been studying vocabulary. Table 4 provides the results of the data analyses based on the five variables mentioned above.

In order to investigate the effect of the five independent variables on the students' receptive vocabulary test scores, the regression model was applied to the data. For the analysis, Forms A and B were analyzed separately and nominal variables were dummy coded; that is, the reference groups were coded as 1 and the non-reference groups were coded as 0 (i.e., For each of the nominal variables, 'College of Humanities', 'male', 'no abroad experience', 'CSAT level 1', 'study words by a separate vocabulary book' were all coded as reference groups).

As presented in Table 4, results of the regression analysis reveal that CSAT level is the most critical factor for anticipating Korean students' receptive vocabulary size. The effect of CSAT level on the receptive vocabulary size is statistically significant, and its effect size is -5.22 and -5.77 for each test form. In other words, Korean students' receptive vocabulary size is expected to increase by 5.22 or 5.77 when the CSAT level goes up by one level, if other conditions are controlled. In addition, the learning styles were also dummy coded, and a group of students who adopted learning style 2 were the reference group. This dummy coding allowed us to compare the effects of learning styles among groups and consider the interactive effect of the other variables. The results show that the effects of learning style 1 and 5 are statistically significant. The coefficient of learning style 1 (at -2.31 for Form A and -3.18 for Form B) indicate that students who adopted learning style 1 are expected to have a 2.31(i.e., 230 words) or 3.18 (i.e., 318 words) smaller receptive vocabulary size than those of the reference group. In the same way, students who adopted learning style 5 are expected to have a 3.50 or 3.92 smaller receptive vocabulary size compared with the reference group. This indicates that learning words through vocabulary exercise books is the most effective way for Korean students.

[2] CSAT consists of 9 levels (levels 1~9) where Level 1 is the highest.

교보문고 KYOBO Book Centre

TABLE 4

Receptive Vocabulary Test Scores and Background Variables

Receptive	Form A				Form B			
	Unstandardized coefficient		t	Sig.	Unstandardized coefficient		t	Sig.
	B	Std. error			B	Std. error		
(constant)	84.50	1.58	53.51	0.00	72.18	1.83	39.50	0.00
School 2	-0.19	1.20	-0.15	0.88	-1.41	1.56	-0.90	0.37
School 3	-0.53	1.15	-0.46	0.65	-0.61	1.33	-0.46	0.65
School 4	-0.28	1.07	-0.26	0.80	-2.18	1.20	-1.82	0.07*
Female	1.24	0.82	1.51	0.13	1.17	0.94	1.25	0.21
Overseas Experience	-0.49	1.01	-0.48	0.63	-0.68	1.12	-0.60	0.55
CSAT Level	-5.22	0.45	-11.59	0.00**	-5.77	0.55	-10.45	0.00**
Learning Style 1	-2.31	1.06	-2.17	0.03*	-3.18	1.22	-2.61	0.01*
Learning Style 3	-0.78	1.15	-0.68	0.50*	-1.38	1.40	-0.99	0.33
Learning Style 4	-1.07	1.76	-0.61	0.54	-0.01	2.07	0.00	1.00
Learning Style 5	-3.50	1.21	-2.90	0.00**	-3.92	1.33	-2.95	0.00**

** $p<0.01$, * $p<0.05$

. *Note*: Learning style 1 = during class, Learning style 2 = through vocabulary exercise books, Learning style 3 = through lectures at private institutes (e.g., cram schools), Learning styles 4 = miscellaneous categories, and Learning style 5 = through incidental acquisition

Table 5 shows the results of regression analysis based on the productive vocabulary data. The results show that CSAT level is again the most critical variable for predicting learners' productive vocabulary size as was the case for the students' receptive vocabulary size. As shown, the effect of the CSAT level on the productive vocabulary size is statistically significant at -7.11 and -5.85 for Forms A and B. These coefficients indicate that the students' productive vocabulary size is expected to increase by 7.11 (i.e., 711 words) or 5.85 (i.e., 585 words) as their CSAT level goes up by one level. The analysis of results from Form A also indicates that overseas experiences may be a significant variable affecting students' productive vocabulary size. Its coefficient of -2.70 indicates that students who do not have any experience abroad are expected to have a 2.70 (i.e., 270 words) smaller productive vocabulary size compared with students who have had such experience. On the other hand, the analysis of results from Form B shows that the group who adopted learning style 1 differed significantly when compared to the group with learning style 2 (the reference group). That is, the learners who used separate vocabulary exercise books (reference group) are likely to score 2.51 (i.e., 251 words) higher than the group of learners who rely on classroom lessons (learning style 1) for the improvement of productive vocabulary size.

교보문고 KYOBO Book Centre

TABLE 5

Productive Vocabulary Test Scores and Background Variables

Productive	Form A				Form B			
	Unstandardized coefficient		t	Sig.	Unstandardized coefficient		t	Sig.
	B	Std. error			B	Std. error		
(constant)	63.07	1.96	32.12	0.00	41.87	1.72	24.35	0.00
School 2	0.42	1.50	0.28	0.78	-1.59	1.47	-1.08	0.28
School 3	-0.52	1.43	-0.36	0.72	-0.84	1.25	-0.67	0.50*
School 4	-2.45	1.33	-1.84	0.07*	-0.03	1.13	-0.02	0.98
Female	-1.39	1.02	-1.37	0.17	-0.65	0.88	-0.74	0.46
Overseas Experience	-2.70	1.26	-2.14	0.03*	0.62	1.05	0.59	0.56
CSAT Level	-7.11	0.56	-12.70	0.00**	-5.85	0.52	-11.28	0.00**
Learning Style 1	-0.64	1.32	-0.49	0.63	-2.51	1.15	-2.19	0.03*
Learning Style 3	1.06	1.43	0.75	0.46	-0.48	1.32	-0.36	0.72
Learning Style 4	2.10	2.19	0.96	0.34	2.16	1.95	1.11	0.27
Learning Style 5	-1.24	1.50	-0.82	0.41	0.98	1.25	0.78	0.43

** $p<0.01$, * $p<0.05$

. *Note*: Learning style 1 = during class, Learning style 2 = through vocabulary exercise books, Learning style 3 = through lectures at private institutes (e.g., cram schools), Learning styles 4 = miscellaneous categories, and Learning style 5 = through incidental acquisition

These results clearly support the prediction that Korean students' vocabulary size may become larger as their CSAT level goes up. In addition, learning style and overseas experience revealed to be the significant variables affecting vocabulary sizes of Korean students. With respect to receptive vocabulary size, it appears that learning style is consistently significant, although not for the productive vocabulary size. Instead, overseas experience was more significant in Form A of the productive vocabulary size test; this may be due to the amount of input from native speakers to which they were exposed to when abroad. The variables major and gender, however, did not show any significant effect on vocabulary size.

3. IRT Analysis of the Item Difficulty of the Receptive and Productive Vocabulary Size Tests

In this study, the test scores of Forms A and B were examined using IRT analysis in order to compare the item difficulty of the two forms while controlling the measurement error across the test scores. Table 6 shows the average of the item difficulty difference between Forms A and B by the ten word-frequency bands.

교보문고 KYOBO Book Centre

TABLE 6

Item Difficulty Differences between Forms A and B by Word Bands

Receptive word bands	Average Difference	Productive word bands	Average Difference
1st 1000	1.52	1st 1000	2.48
2nd 1000	1.66	2nd 1000	1.92
3rd 1000	2.83	3rd 1000	2.01
4th 1000	1.56	4th 1000	1.22
5th 1000	2.61	5th 1000	2.71
6th 1000	3.68	6th 1000	3.35
7th 1000	1.75	7th 1000	3.23
8th 1000	1.76	8th 1000	2.26
9th 1000	1.86	9th 1000	2.23
10th 1000	1.68	10th 1000	1.79

A higher number in the average of the item difficulty difference indicates a greater difference in terms of the item difficulty. As indicated by Table 6, for receptive vocabulary, Forms A and B are most different in bands 3 (2.83), 5 (2.61), and 6 (3.68); and bands 5 (2.71), 6 (3.35), and 7 (3.23) are the most different for productive vocabulary.

Figure 3 shows the relationship between Forms A and B in terms of item difficulty for the receptive vocabulary and also indicates the correlation between item difficulty in Forms A and B.

FIGURE 3

Comparison of Forms A and B in terms of Receptive Item Difficulty

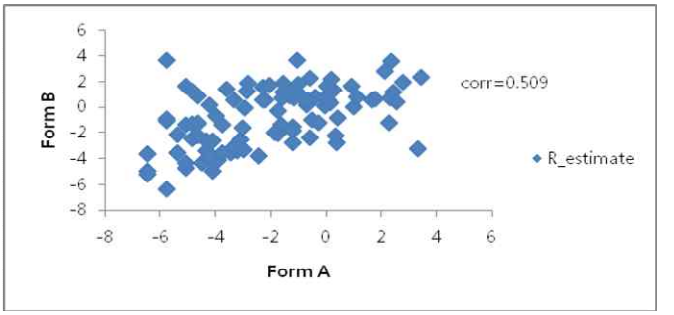
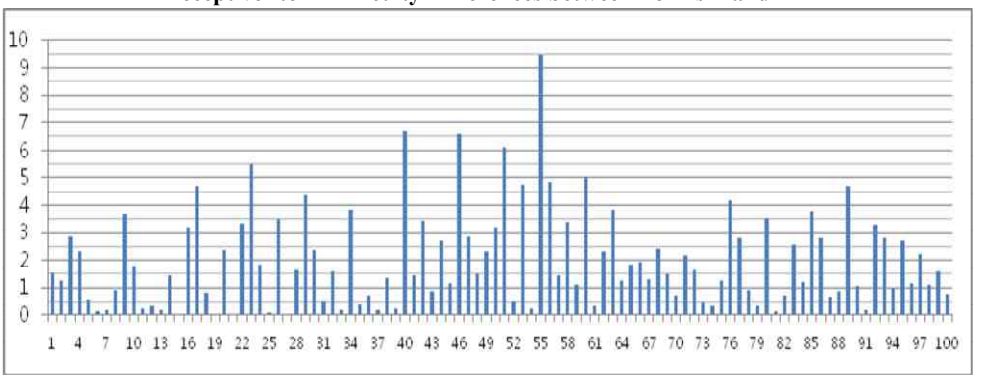


Figure 3 indicates that Forms A and B do not have a linear relationship, which the correlation coefficient (0.509) also supports. This means Forms A and B are not homogeneous in terms of the item difficulty. Figure 4 shows the distribution of the receptive item difficulty difference between the two forms by each item.

교보문고 KYOBO Book Centre

FIGURE 4

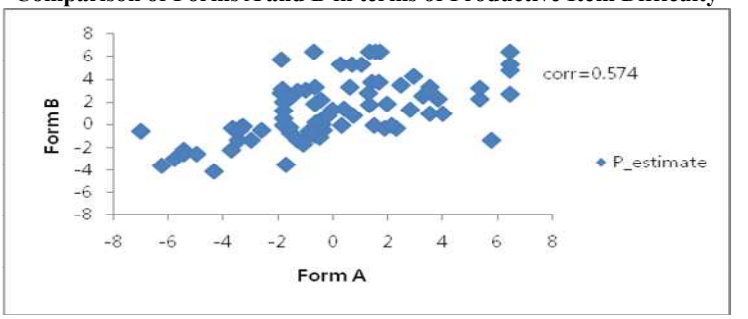
Receptive Item Difficulty Differences between Forms A and B



When we examined some items, the reasons for the large gap in item difficulty difference between the two forms were found to stem from the learners' Korean cultural background. The three items 55. *genius-threshold* (9.441), 40. *worship-allege* (6.704), and 46. *dragon-majesty* (6.575), which showed the biggest gap between Forms A and B, suggest an implication for constructing a vocabulary size test in the EFL context. The three word items *genius*, *worship*, and *dragon* used in Form A are very familiar words to Korean learners. According to the Basic Word List of the National Curriculum of English, those words are taught at the elementary school level in Korea. On the other hand, although *threshold, allege,* and *majesty* used in Form B were retrieved in the matching word frequency bands as the former three words, these three words seemed to be relatively more difficult for Korean learners, and this has been supported empirically by our test results. This indicates that randomizing for word selection may not guarantee equivalent test forms.

FIGURE 5

Comparison of Forms A and B in terms of Productive Item Difficulty



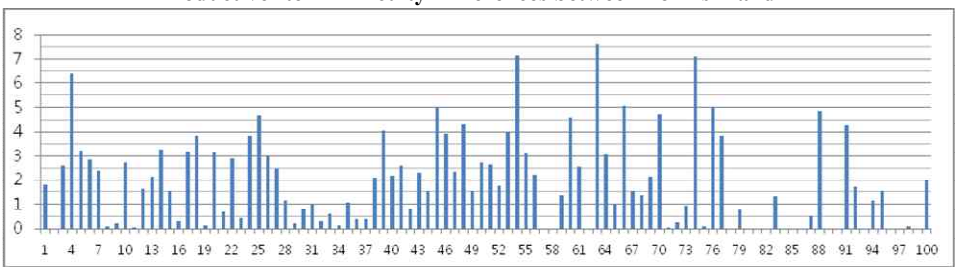
교보문고 KYOBO Book Centre

Figure 5 shows the relationship between Forms A and B in productive vocabulary item difficulty. As shown in Table 6, Forms A and B do not have a linear relationship, which Figure 5 also supports with the correlation coefficient (0.574).

Our analysis results clearly indicate that the participants' performances differ between Forms A and B, and thus the two forms are not equivalent as in the receptive test forms. The same results are found across frequency bands for both the receptive and productive test forms. In fact, for some productive items, item difficulties could not estimated because the students did not respond to those items correctly at all (e.g., 96. *livid-vertebrate*, 97. *subdue-embryonic*). Figure 6 shows the distribution of the productive item difficulty difference between the two forms by each item.

FIGURE 6

Productive Item Difficulty Differences between Forms A and B



As shown in the results of the receptive test, productive items such as 63. *climax-grunt* (7.608), 54. *grubby-goldfish* (7.124), and 74. *humid-reek* (7.119) also indicate a large difference in item difficulty. *Climax, goldfish* and *humid* seem to be familiar words for Korean students compared with the words *grunt, grubby*, and *reek*. As a result of these analyses, such uneven distribution in item difficulty seems to have occurred due to the learners' test-preparation materials, knowledge of loanwords, or by incidental overlearning of specific items (e.g., via curriculum-based textbooks). In addition, the unit of 1,000 words per word band used for vocabulary assessment may have been too large a pool of words for EFL learners; when ten words from a 1,000 word band are randomly selected, the item difficulties could not be controlled. Thus, in future construction of vocabulary levels test for EFL learners, who are situated in unique educational contexts, it may seem necessary to divide the word frequency bands by a smaller unit of words rather than 1,000. It seems that testing EFL learners with on a narrower range of words for each band will produce a more valid estimate of the learners' vocabulary sizes.

교보문고 KYOBO Book Centre

V. CONCLUSION

Our initial attempt to construct vocabulary levels tests (VLT) was aimed to assess vocabulary knowledge of EFL Korean university learners at the different word frequency levels. The construction of the tests in two forms (i.e., Forms A and B) was based on the methodology Nation (2010) had used to construct his vocabulary levels test (i.e., use of randomization at each word level to select target words based on the fourteen word bands of the BNC). Both Forms A and B were constructed based on the same range of words (1st – 10th 1,000 word bands), and we expected gradual decrease in vocabulary sizes for both receptive and productive vocabulary sizes at the lower frequency word bands regardless of the type of forms used on the EFL learners. The EFL learners' vocabulary sizes for both forms *did* show general falls. In connection to our interest in personal background variables, our examination of the learners' vocabulary sizes with the use of the multiple regression model showed that the Korean learners' CSAT level was the most critical variable for anticipating Korean students' receptive and productive vocabulary sizes. Studying with a vocabulary exercise book seemed to be the most effective for the learners' improvement of receptive vocabulary. The results of the study also demonstrated that the students with overseas experience resulted in larger productive vocabulary sizes.

To see if Forms A and B produced similar results respectively at the ten different word bands (since theoretically the two forms were designed to measure vocabulary knowledge in the same way), IRT analysis, however, showed that the learners' performances differed between the two forms. We realized that even within the same word frequency bands, some of the words had been more easily learned through deliberate teaching in the early stages of learning English (e.g., elementary school) or due to their culture-specific content schema the learners held about certain English words.

As seen in how culture-specific or curriculum-based words may receive relatively more attention with regard to the development of EFL learners' vocabulary sizes, the results of the study, as a whole, led us to some skepticism about how vocabulary levels tests have previously been designed to measure L2 learners' vocabulary size. The results of the study lead us to propose some methodological implications for the design of future vocabulary levels test. Particularly for those learners in EFL contexts, we would need to construct VLT that takes the learners' cultural and individual background variables into consideration since it would be unrealistic to expect EFL learners to acquire the English vocabulary in the same way native speakers expand their vocabulary size. According to Nation (2001), native speakers on average are believed gain an increase of 1,000 word families in a year, but this kind of lexical development is not practically possible for EFL

교보문고 KYOBO Book Centre

learners who are limited to the amount of input they receive. As seen in the study, we tried devising vocabulary levels tests in two forms with use of the same methodological principle, but the results did not reflect this when we found that there were differences in the scores of vocabulary sizes at each word band between Forms A and B. We could no longer see the tests at similar difficulty levels. As such, we believe that the adoption of the VLT based on the word lists of native-speaker corpus is unlikely to offer valid target words for EFL learners. As such, we suggest that culture-specific or curriculum-bound words be excluded or reconsidered before selection. To improve construct validity, there would also be need to incorporate a narrower range of words (e.g., 500 word families rather than 1,000 words), and include information on the learners' extent of familiarity with the words. Since different test forms may differ in item difficulty, a statistical process, which is equating, should be used to adjust scores on test forms to compare EFL learners' vocabulary level. Since data analyses using an IRT model are more beneficial than those using CTT, equating based on IRT would result in more accurate research findings, and thus the evolving research question should be addressed in the future studies adopting equating methods.

REFERENCES

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Chon, Y. V., & Kim, Y. -H. (2011). Lexical discovery and consolidation strategies of proficient and less proficient EFL vocational high school learners. *English Language & Literature Teaching*, *27*(3), 27-56.

Chon, Y. V., & Shin, D. (2012). Lexical profiles and socioeducational variables of Korean EFL university learners. *Korean Journal of Applied Linguistics*, *28*(1), 115-146.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Goulden, R. P., Nation, I. S. P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11*, 341-363.

Gui, S. (1982). A survey of the size of vocabulary knowledge of Chinese students. *Language Learning and Communication, 1*(2), 163-178.

Hambleton, R. K., & Swanminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer.

교보문고 KYOBO Book Centre

Hazenberg, S., & Hulstijn, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics,* 17, 145-163.

Henard, D. H. (1998). *Using spreadsheets to implement the one-parameter item response theroy (IRT) model.* Paper presented at the annual meeting of the Southwester Psychological Association, New Orleans.

Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language Testing 16*(1), 33-51.

Laufer, B., & Paribakht, T. S. (1998). Relationship between passive and active vocabularies: Effects of language learning context. *Language Learning, 48*, 365-391.

Morgan, B. Q., & Oberdeck, L. M. (1930). Active and passive vocabulary. In E. W. Bagster-Collins (Ed.), *Studies in modern language teaching* (pp. 213-221). London: Macmillan.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Heinle and Heinle.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

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

Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt and M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.

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

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

Schmitt (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

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.

Shin, D., Chon, Y. V., & Kim, H. (2011). Receptive and productive vocabulary sizes of high school learners: What next for the basic word list? *English Teaching, 66*(3), 123-148.

Son, J. (2007). The effects of vocabulary exercises on EFL vocabulary learning and retention. *English Language & Literature Teaching, 14*(4), 167-192.

Waring, R. (1997). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata [Occasional papers of Notre Dame Seishin University, Okayama], 1*, 53-68.

교보문고 KYOBO Book Centre

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

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

Wu, M. L., Adams, R. J., & Wilson, M. (1988). ACERConquest [computer program]. Hawthorn, Australia: ACER

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

Yongsang Lee
National English Ability Test Division
Korea Institute for Curriculum and Evaluation
15-5 Jeong-dong Jung-gu Seoul 100-784
Email :yong21c@hotmail.com

Yuah V. Chon
Dept. of English Education, College of Education
Hanyang University
17 Haengdang-dong
Seongdong-gu Seoul 133-791
Tel: (02) 2220-1144
Email: vylee52@hanyang.ac.kr

Dongkwang Shin
National English Ability Test Division
Korea Institute for Curriculum and Evaluation
15-5 Jeong-dong Jung-gu Seoul 100-784
Email: sdhera@hotmail.com

Received in January, 2012
Reviewed in February, 2012
Revised version received in March, 2012

교보문고 KYOBO Book Centre