Modern English Education, Vol. 18, No. 4, Winter 2017
http://dx.doi.org/10.18095/meeso.2017.18.4.02

Predictability of the Cloze Test as a Measure of Written Productive Vocabulary*

Young-sook Ryoo
University of Seoul

Ryoo, Young-sook. (2017). Predictability of the cloze test as a measure of written productive vocabulary. *Modern English Education*, *18*(4), 25-45.

The present study explores whether cloze test scores under two different scoring methods and English test scores on TOEIC can predict written productive vocabulary of Korean EFL learners. One hundred and fourteen students' written productive vocabulary in summary writing was analyzed using *VocabProfile,* which provides four frequency-based vocabulary level lists: the most frequent 1,000 words (K1: function words and content words), second most frequent 1,000 words (K2), academic word lists (AWL), and off the list words (OLW). The study found that the exact-word method led to significant differences in K1, function words, K2, AWL, and OLW, and the acceptable-word method resulted in significant differences in K1, function words, K2, and OLW, indicating only one difference in AWL. On the other hand, the English proficiency test yielded significant differences in K1 and function words alone. These study results suggest that the cloze test under the two different scoring methods did not make a remarkable difference in the predictability of Korean EFL learners' written productive vocabulary. In addition, both cloze test scores and general English test scores showed they could serve as predictors of Korean EFL learners' written productive vocabulary, but cloze test scores led to more significant differences in vocabulary use in writing.

**[cloze test/written productive vocabulary/English proficiency/
클로즈 테스트/문자로 쓰인 생산적 어휘/영어 능숙도]**

---

* Part of this paper was presented at 2013 ALAK (The Applied Linguistics Association of Korea) International Conference.

I. INTRODUCTION

The cloze test was originally developed by Taylor (1953) as a measure of readability of texts and as an alternative to conventional standardized tests. Taylor studied the value of the cloze test to estimate the readability of reading materials. Since then, the cloze test has been used in classes for language proficiency tests, and its reliability has been explored in many studies. Also, different varieties of cloze tests such as the fixed ratio cloze, rational cloze, multiple choice cloze, and C-test have been explored to see the relationships with various language proficiency areas. Research has shown that the cloze test is highly correlated with student performance in reading (Bachman, 1985; Gellert & Elbro, 2013; Ito, 2004), writing (Hanania & Shikhani, 1986; S. Lee 1997), vocabulary knowledge (Harsch & Hartig, 2016; Keshavarz & Salimi, 2007), and overall language proficiency (Aiken, 1977; Saeedi, Tavakoli, Kazerooni, & Parvaresh, 2011). Notably, Brown (2001) argue that successful completion of cloze test items reflects L2 learners' knowledge of vocabulary, grammar, and discourse structure as well as their reading skills. In addition, Keshavarz and Salimi (2007) maintain that cloze tests are different from the general fill-in exercises using isolated sentences because cloze tests are applied to a longer passage and are therefore contextualized.

Although the cloze test has been widely used to measure L2 learners' language proficiency, it has also been criticized for its limitations. The most frequently mentioned issue in regard to the cloze test is what it actually measures in terms of L2 learners' language proficiency. In particular, it has been questioned whether it actually measures what it purports to be measuring. Brown and Hudson (1998) mentioned the limitations of the fill-in assessment by stating that it generally focuses on testing a single word or short phrase at most. If cloze test scores can predict the general English proficiency of L2 learners, this issue will be clearer.

The scoring method is another issue regarding the cloze test. Under the exact-word method, only an original word from the passage is counted correct. On the other hand, all the possible alternatives for each cloze item are counted correct under the acceptable-word method. Oller (1972) said, "The task of guessing the exact-word is not necessarily a language skill in the ordinary sense of the term" (p. 152). In addition, Brown and Hudson (1998) state that a blank to fill in may have a number of possible answers. This means that there is a possibility that each respondent may produce completely different, possible alternatives for a cloze test item. They illustrated that Brown (1980) found as many as 28 possible answers for one particular cloze test blank, which is worthy to note in that it may be needed to determine the effects of different cloze scoring methods.

It could be assumed that replacing vocabulary for cloze test items is very similar to actual writing where writers retrieve vocabulary according to contexts (i.e., meaning and

grammar). In this vein, we might be able to say that the cloze test is contextualized written productive language. This feature of the cloze test also prompted the researcher to further investigate whether the acceptable-word method can better predict students' written productive vocabulary than the exact-word method. To the researcher's knowledge, there have been few attempts to examine cloze test scores under the two different scoring methods although many studies have reported that the cloze test can function as a measure for various English proficiency areas.

Considering all the issues mentioned above, the present study explores whether the cloze test under the two different scoring methods and a general English proficiency test can predict Korean EFL learners' written productive vocabulary in writing. First, the study examines Korean EFL learners' written productive vocabulary according to their cloze test scores under the exact-word method and then investigates again under the acceptable-word method to see which method is fairer with Korean EFL learners as a predictor of written productive vocabulary. Next, the study examines whether scores on a general English language proficiency test can also predict written productive vocabulary because many previous studies have reported the cloze test as a comprehensive measure of general English proficiency. Thus, the following research questions were addressed for the study:

1) Does students' written productive vocabulary show significant differences according to cloze test scores? If so, which method better predicts students' written productive vocabulary?
2) Does students' written productive vocabulary show significant differences according to general English proficiency scores?

II. LITERATURE REVIEW

1. Research on the Cloze Test Under Two Different Scoring Methods

When one considers constructing cloze tests, there are many things to decide in relation to test construction and interpretation. The variables involved in cloze tests are as follows: scoring methods (exact-word and acceptable-word); deletion patterns (rational deletion and fixed ratio deletion); response types (open ended, multiple choice, and C-test); item types (content words, function words, collocations, and other particular parts of a text); and text difficulty. As Watanabe and Koyama (2008) suggested, these test design features are important factors that affect the respondents' performance on the cloze tests and the interpretation of the test results. Previous research shows that these variables are needed to be isolated or merged together to gain a deeper understanding of the L2 learners' language

proficiency in terms of language production and the relationship with the other language proficiency tests.

Among those variables mentioned above, the scoring method has been one of the most frequently studied areas in the field. Some studies supported the acceptable-word method in the light of a common concern that the exact-word method will be unfair for the less proficient L2 language learners. For example, Oller (1972) compared five different scoring methods. The five categories of responses were differentiated as follows: restorations of the original words (exact-word method); entirely acceptable fill-ins (acceptable-word method); responses that violate long-range constraints (e.g., "the man" for "a man" when no man has been mentioned previously); responses which violate short-range constraints (e.g., "I goes" for "I go"); and entirely incorrect fill-ins or items left blank indicating complete lack of comprehension. The study revealed that the best of the five scoring methods is the acceptable-word method that counts any contextually acceptable responses as correct because of more significant correlations with existing language tests. In other words, the study reported that the acceptable-word method is superior to the exact-word method.

In line with Oller (1972), Brown (1980) recommends the acceptable-word method compared with the exact-answer, acceptable-answer, multiple choice, and clozentropy. Clozentropy is a test which logarithmically weights the acceptable answers based on their frequency in a pretest written by a native speaker. These four scoring methods were examined in terms of reliability, validity, mean item facility, mean item discrimination, and usability. Brown reveled that there are differences in all those aspects except for validity and suggested that there are relative merits of the four scoring methods explored in the study. Based on the study results, he concluded that the best overall coring method is the acceptable-word method. He also wrote that the acceptable-word method appears much fairer to students than the exact-word method. Notably, he stated that counting acceptable answers wrong is not fair. He further added that counting a possible answer wrong which is actually correct, simply because the author did not choose the word, is not right.

Similarly, Brown, Yamashiro, and Ogane (2001) administered the revised version of the test from Bachman (1985) to Japanese students, and they revealed the acceptable-word method is superior to the exact-word method as a predictor of language proficiency for EFL learners. Kobayashi (2002a) also looked at the effects of different scoring methods and concluded that the acceptable-word method may be more effective as a predictor of learner proficiency levels. He employed three scoring methods: the exact-word scoring; the semantically and syntactically acceptable scoring method; and the semantically acceptable, but syntactically unacceptable-word scoring method. Noticeably, Kobayashi (2002a) used two types of the acceptable scoring methods to further examine the interactions between the scoring methods and item characteristics. The study findings suggested that the acceptable-word methods may be more effective than the exact-word method. The cloze test scores of low-proficiency learners

increased when switching from the exact-word method to the acceptable-word method than those of high proficiency learners. Evidence from previous studies suggests that the acceptable-word method is more effective and fairer to assess the English language proficiency of L2 learners.

In this vein, Keshavarz and Salimi (2007) chose the acceptable-word method for the open-ended cloze test in their study of the relationship between Iranian EFL learners' overall collocational competence and their performance on two types of cloze tests (open-ended and multiple-choice). Their study focused on the cloze tests as an overall language proficiency measure and revealed that there was a significant relationship between performance on cloze tests and collocational competence.

On the other hand, Sadeghi (2014) revealed that phrase cloze test developed as a new version of cloze testing was no better than a corresponding reading comprehension test when it was administered to 53 Iranian EFL learners. The phrase cloze test was constructed by leaving out two to four words from cloze passages used in the study, and the students were instructed to fill in the blanks with phrases rather than single words. They used two types of cloze texts with a different level of difficulty that had 40 cloze items. The cloze tests were scored with both the exact-word scoring method and acceptable-word scoring method respectively. Under the acceptable-word scoring method, the first time, all acceptable phrases were given credit, and then the second time, all acceptable words as well as all acceptable phrases were also given credit. Sadeghi compared the phrase cloze test scores with reading test scores from two conventional reading tests with different levels of difficulty. He concluded that there was no significant relationship between phrase cloze test scores and conventional reading test scores regardless of the cloze scoring method and reading passage difficulty.

Other studies supported and recommended the exact-word method because a high correlation was found between the exact-word method scores and the acceptable-word method scores. For example, Stubbs and Tucker (1974) proposed the exact-word method because it was as valid as the acceptable-word method. Similarly, Clausing and Senko (1978) examined the exact-word scoring and two different types of acceptable scoring: fully acceptable and partially acceptable. They found a high correlation between the two scoring methods and suggested that the quick scoring technique (exact-word scoring) be recommended. As for another study supporting the exact-word scoring, Watanabe and Koyama (2008) provided comprehensive profiles of cloze test research. They also stated that the exact-word method is preferred because of its practicality involved in the ease of test construction and scoring. They added choosing which scoring method to adopt depends on the decision of the test administrators because they have to consider the purpose of the test, feasibility of creating and using acceptable answers, possible errors during the scoring procedures, and other practical issues. Thus, previous studies stated that the exact-word method is recommended because it is easier and quicker to develop and mark for English teachers and researchers.

Based on the existing research, it is uncertain whether the cloze test can predict L2 learners' written productive vocabulary. In particular, it may be worthy to investigate which type of scoring methods (exact and acceptable) can serve as a more adequate measure of L2 learners' written productive vocabulary. Therefore, the findings from the present study may provide insight into the relationship between Korean EFL learners' performance on the cloze test and their written productive vocabulary.

2. The Cloze Test as an English Proficiency Measure

The cloze test is based on the idea of removing words from a text and asking test takers to fill in the missing words that are semantically and syntactically appropriate, so previous studies have used it as a measure of L2 learners' English proficiency. Research has shown that cloze testing is highly correlated with L2 learners' English proficiency as mentioned earlier.

However, studies on the cloze test as a measure of reading comprehension have shown different study results even though it was originally invented to measure readability of texts. Some studies revealed that cloze testing can be used as a measure of L2 learners' reading comprehension, while others showed opposing results.

For example, Ito (2004) proved the possibility of using the cloze test as a measure of reading comprehension. Ito examined the reliability and validity of translation tests as a reading comprehension measure. In the study, two types of translation tests were taken: open-ended translation and multiple-choice translation. In addition, short-answer question and cloze testing were also used to explore the efficiency of the translation test as a reading comprehension measure. The study showed that open-ended translation might be a reliable and valid testing method, and multiple-choice translation also has a potential for measuring reading comprehension. The two types of translation tests showed a moderate correlation between them. Notably, the study revealed that the open-ended translation test shows a high correlation with the cloze test. Put another way, the cloze test was included in the study as a measure of reading comprehension and also proved the possibility as a reading comprehension measure.

In this line of research, Gellert and Elbro (2013) is also based on the idea that filling in cloze test gaps requires proper understanding of the text. The study found that carefully selected word gaps can measure L2 learners' performance on reading comprehension. Gellert and Elbro indicated that the comprehension-focused cloze test demonstrated its reliability and validity, which means the cloze test can serve as a measure of L2 learners' reading comprehension.

Other studies explored whether cloze testing can predict L2 learners' vocabulary knowledge. For instance, Keshavarz and Salimi (2007) explored the relationship between Iranian EFL learners' English collocational competence and their performance on two types of cloze test. The open-ended cloze test and multiple-choice cloze test were administered to the students. The open-ended cloze test was scored by using the acceptable-word method, but for the multiple-

choice cloze items only one answer was counted correct. The study found that the correlations between students' collocational competence and two types of cloze tests were significant. The relatively high and positive correlations among the scores on the collocation tests and the scores on the cloze tests strongly suggest that the two types of cloze test did not lead to any significant differences. In other words, the acceptable-word method was not more predictive of Iranian EFL learners' vocabulary knowledge. This finding also suggests that cloze testing can predict EFL learners' vocabulary knowledge with a particular focus on their collocational competence.

As another type of cloze test, the C-test has been examined as to whether it can predict L2 learners' vocabulary knowledge. The C-test provides blanks where the first half of a word is provided. Harsch and Hartig (2016) compared German EFL learners' performance on a C-test and Yes/No vocabulary test. They stated the C-test was embedded and contextualized in nature, while the Yes/No vocabulary test was inherently discrete and decontextualized. They reported that the C-test seems to be the more powerful predictor for L2 receptive language skills (listening and reading comprehension) than the Yes/No vocabulary test. They, finally, concluded that the C-test was a reliable, economical, and robust measure for placement and screening tests.

Compared to an abundance of cloze test studies, there have been fewer attempts to determine if the cloze test is significantly correlated with a general English language proficiency test such as Test of English as a Foreign Language (TOEFL), which is an established, standardized, and internationally administered test of general English proficiency. For example, Saeedi et al. (2011) investigated the validity of the C-test and cloze test as a general language proficiency measure and found the cloze test is significantly correlated with a general English proficiency test. As a criterion measure, the researchers used TOEFL to assess 90 Iranian English majors' general English language proficiency. The study used the exact-word method and the analysis results indicated that there was a statistically significant correlation among participants' scores on the cloze test, C-test, and the TOEFL.

In Korea, E. Y. Chae and J. A. Shin (2015) compared Korean EFL learners' performances on two types of cloze test: a timed cloze test (timed condition for each cloze test item response) and an untimed cloze test (typical cloze test condition with a pencil and paper mode). The students' responses to each cloze item were scored by using the exact-word method. They reported that there was a significant, positive correlation between TOEIC scores and the timed cloze test scores. This result indicates that cloze test can predict Korean EFL learners' general English proficiency.

Thus, the cloze test has been widely used as a measure of English proficiency in second or foreign language contexts. However, it seems that what the cloze test actually measures still remains uncertain. In addition, almost no major studies, to the researcher's knowledge, have explored the relationship between EFL learners' written productive vocabulary in writing and their performance on the cloze test and general English test.

III. METHOD

1. Participants

The students (n = 114) from four classes participated in the study. They were taking an English course required for first-year university students in Seoul. Their majors were various: Chinese language and literature, law, English language and literature, international relations, public administration, social welfare, and business administration. They were divided into a high proficiency group and a low proficiency group based on cloze test scores under the two scoring methods respectively and scores on Test of English for International Communication (TOEIC). The students who obtained average scores were excluded to lead to clearer proficiency differences between the two groups. More specifically, the two different proficiency groups were classified according to the average score of the two scoring methods respectively. The number of students assigned to the higher or lower proficiency groups according to cloze test scores and TOEIC scores will be presented in more detail in the results section.

2. Data Collection and Data Analysis

The students were asked to summarize a book they read for the writing task, and they were allowed to add their feelings at the end of the summary. They chose their own English books to read and summarize under the guidelines the researcher provided. They were supposed to choose an English book over 200 pages in length and produce summary writing with over 300 words. Summary writing from a variety of sources was taken for the current study to ensure that student writing carries various vocabulary from one text to another. Their writing was stored as text files for data analysis.

VocabProfile (Cobb, 2002) was employed to assess students' vocabulary in writing. This program analyzed each student's lexical profiles and it provided four frequency word lists. The analysis results are K1, K2, AWL, and OLW. K1 and K2 correspond to the most frequent 1000 words and the second most frequent 1000 words. In particular, the two subsets of K1, FW, or function words and CW, or content words were also analyzed to gain a deeper understanding of vocabulary use in student writing. Function words include articles, prepositions, conjunctions, and auxiliaries. Put differently, function words are grammatical items. Content words, on the other hand, include adverbs, nouns, adjective, and verbs. Content words represent vocabulary items with semantic content, and function words work syntactically alone. AWL is words from Coxhead's Academic Word List (Coxhead, 2002) and OLW is everything else not appearing in other vocabulary lists. AWL and OLW are considered more infrequent vocabulary, or higher level vocabulary than K1

and K2. It has been usually assumed that the more amounts of infrequent words a written text contains, the more advanced vocabulary the author has (e.g., Laufer, 1991; Laufer & Nation, 1995; Meara & Bell, 2001; Y. Ryoo, 2016).

The students of the study were told to fill in the cloze items in class during 20 minutes for each one. The first cloze text (C1) was taken from Brown et al. (2001), and the second cloze text (C2) was from Kobayashi (2002a). C1 contained 30 blanks, and C2 had 25 (see Appendix 1 and Appendix 2, respectively). Two cloze scores from the two different types of cloze texts were merged because it was assumed that a combination of the two different cloze texts would improve the prediction of Korean EFL students' written productive vocabulary in writing. Accordingly, the maximum score was 55 under the exact-word method. Under the exact-word method, students' cloze scores were the number of correctly answered items (i.e., original words in the cloze texts). On the other hand, under the acceptable-word method, the numbers of answers increased to 105 (C1) and 66 (C2) respectively because any possible words in terms of both meaning and grammar were counted correct.

The two cloze passages were examined by *VocabProfile* (Cobb, 2002) to determine which type of vocabulary each cloze passage carries. Table 1 shows the percentages of each vocabulary list, total number of words (token), and number of different words (type) that *VocabProfile* provides. As Table 1 displays, C1 seems more difficult than C2 because C1 carries more percentages of advanced vocabulary (i.e., AWL and OLW) and greater number of types (C1: 206, C2: 170). If a text contains more types of words, it means the text has more various vocabulary. The length of the two tests, however, is almost the same (C1: 363 words, C2: 354 words). As for the cloze test items, the number of function words was 35 (C1: 23, C2: 12) and the number of content words 20 (C1: 7, C2: 13) out of 55 blanks in total. Thus, these two different cloze texts were used to provide the students with more various contextual clues for cloze test items in terms of lexical distribution, text difficulty, and cloze test item characteristics.

TABLE 1

Lexical Distribution of Cloze Texts

Cloze Text	K1 (%) (FW, CW)	K2 (%)	AWL (%)	OLW (%)	Token	Type
C1	78.79 (43.25, 35.54)	6.89	9.91	4.41	363	206
C2	90.11 (44.92, 45.20)	3.11	3.11	3.67	354	170

Note. Token = total number of words; Type = number of different words.

One point was given for every correct answer (i.e., original vocabulary in the text) under the exact-word method. Every acceptable word was also given one point under the acceptable-word method, but only semantically and grammatically acceptable answers got credit. Obvious spelling mistakes were counted correct as well.

The glossary for the acceptable-word method was built on the acceptable words already provided from the two studies (Brown et al., 2001; Kobayashi, 2002a) and the Korean students' actual responses to each cloze test item. The students' responses were reviewed and discussed by the researcher and two native speakers to reach an agreement about all the acceptable alternatives. The two native speaker English teachers have been teaching English for about five and six years respectively in Korea. The students were told not to leave any blanks unfilled and asked to give their best guesses for each blank. The student responses to the cloze items were marked by the researcher using the final vocabulary list containing possible answers as well as original vocabulary from the texts.

TOEIC was employed as a general English language test. The participants were required to take a TOEIC test at the start of the semester. The TOEIC test was administered to freshmen to find out their English proficiency levels. TOEIC contains 200 questions: 100 listening comprehension questions and 100 reading comprehension questions. In particular, the reading comprehension questions of TOEIC assess test takers' abilities to comprehend passages and fill in the blanks of given sentences with the most appropriate expressions in terms of sentence structure and vocabulary. It could be assumed that the students tried their best to take the TOEIC test because they were not allowed to take the English course if their TOEIC scores were below 300.

IV. RESULTS AND DISCUSSION

1. Cloze Test Scores and Written Productive Vocabulary

A MANOVA was performed to examine if there is any significant difference between the Korean EFL learners' cloze scores and their written productive vocabulary in writing. To answer the first research question, the effects of the exact-word method were firstly examined. The total score under the exact-word method was 55. Data analysis revealed that the highest score was 46 and the lowest one was 9. The average score was 23, and six students with the score of 23 were excluded to maximize the proficiency differences between groups. Thus, 49 students were assigned to the higher proficiency group and 59 to the lower proficiency group (see Table 3). In this paper, the MANOVA results will be presented before the summaries of mean differences between variables to provide convenience for readers. As shown in Table 2, significant differences were noted in every vocabulary list except for CW. In fact, this means that the exact-word method yielded significant differences in all of the vocabulary lists because CW is a subcategory of K1. Table 3 demonstrates mean differences between groups differentiated by the cloze test scores under the exact-word method.

TABLE 2

Tests of Between-Subjects Effects: Exact-Word Method

Source	Dependent Variables	Mean Square	*df*	*F*	*p*
Exact Scoring	K1	318.793	1	15.24	.000
	FW	137.664	1	10.70	.001
	CW	37.525	1	2.42	.123
	K2	17.113	1	4.62	.034
	AWL	11.600	1	5.09	.026
	OLW	105.737	1	6.01	.016

TABLE 3

Descriptive Statistics: Exact-Word Method

Dependent Variables	Groups	*M*	*SD*	*n*
K1	High	84.38	4.025	49
	Low	80.93	4.982	59
FW	High	49.88	2.714	49
	Low	47.61	4.172	59
CW	High	34.50	2.912	49
	Low	33.31	4.622	59
K2	High	4.46	1.585	49
	Low	5.26	2.167	59
AWL	High	2.27	1.288	49
	Low	2.92	1.672	59
OLW	High	8.90	3.711	49
	Low	10.89	4.556	59

Interestingly, as seen in Table 3, it was found that the higher proficiency group used more K1 and FW that are considered the most frequent vocabulary (i.e., lower-level vocabulary). This finding is interesting because the most frequent vocabulary has been agreed to be learned earlier than less frequent vocabulary in second language acquisition (e.g., Milton, 2009). It could be explained with the possibility that the higher proficiency learners used relatively more amounts of FWs (one subcategory of K1) in writing than their counterparts did because they were more efficient in using function words in their writing. In contrast, the lower proficiency students might have difficulties using function words in their writing properly and tend to use fewer of them in their writing because they have a lack of knowledge of function words such as articles, prepositions, and auxiliary verbs. This could have led to a relatively smaller amount of K1 that includes function words in their writing even though K1 is considered lower-level vocabulary. However, this

assumption has to be further studied by looking at how they actually use them and what kind of function words they actually choose in their writing.

To illustrate one of the function words, Korean does not have articles equivalent to English, so most of the beginner learners of English have difficulty using English articles in writing. This would have led to the unexpected result even though it has been agreed that L2 learners with lower proficiency tend to produce a relatively greater amount of lower-level vocabulary and vice versa (e.g., Laufer, 1991; Laufer & Nation, 1995; Meara & Bell, 2001). Morris and Cobb (2004) stated that the higher the learners' proficiency levels are, the greater their reliance on FW is, citing Morris and Tremblay (2002) in their study. Morris and Cobb (2004) pointed out that FW is an important indicator of L2 proficiency level. As mentioned earlier, further research is obviously needed to examine what kind of function words were actually used in the higher proficiency students' writing through thorough manual check.

However, the students' use of CWs (a subcategory of K1) did not yield a significant mean difference between the two groups. This means that FW is a more powerful predictor of Korean EFL learners' written productive vocabulary levels differentiated by the cloze test under the exact-word method.

The other vocabulary lists indicate that the less proficient students' writing carried significantly more amounts of K2, AWL, and OLW in writing (see Table 3). This finding is also interesting because one would expect that proficient learners would use more infrequent words in their writing. It could be assumed that those less proficient students might have faced the difficulties paraphrasing the original texts in their summary writing because of their limited language proficiency in English words and structures (Keck, 2006, 2014). Consequently, it seems that the higher proficiency students' more use of FW led to more percentages of K1 and FW (one of the subcategories of K1) and the decrease in the remaining percentages of other vocabulary lists because the total percentage from all the frequency levels should be added up to be 100%.

Next, the effects of the acceptable-word method were also explored to see if this acceptable-word method better predicts students' vocabulary in writing. Under the acceptable-word method, the students' highest score was 53 and the lowest one was 12. Five students who obtained 29 (the average score) were excluded again to show a clearer proficiency difference between groups. As shown in Table 5, the higher proficiency group consisted of 50 students and the lower proficiency group 59 students. The MANOVA was performed again to determine if there is a significant difference between the two groups. The results are presented in Tables 4 and 5.

As can be seen in Table 4, the significant mean differences between groups were found in K1, FW, K2, and OLW. Table 5 indicates that the more proficient students used more

amounts of K1, also known as the most frequent vocabulary among the vocabulary frequency lists. The results also demonstrate that the lower proficiency students used more amounts of infrequent words, which is similar to the analysis results of the exact-word method. These findings indicate the same patterns with the exact-word method except for AWL. The same explanations could be applied to these study results as well. Unlike the exact-word method, however, AWL did not make any statistical difference under the acceptable-word scoring method. This study result partially confirms an earlier study by Paribakht and Webb (2016) that investigated the AWL coverage in reading, listening, and cloze texts. The study found that AWL coverage was not correlated with reading and listening comprehension scores, and there was a small, negative correlation between the AWL coverage and cloze test scores. However, Morris and Cobb (2004) reported a correlation between L2 learners' AWL uses and their academic performance in their course.

TABLE 4

Tests of Between-Subjects Effects: Acceptable-Word Method

Source	Dependent Variables	Mean Square	*df*	*F*	*p*
Acceptable Scoring	K1	309.554	1	14.761	.000
	FW	165.292	1	12.957	.000
	CW	22.521	1	1.431	.234
	K2	16.174	1	4.314	.040
	AWL	7.517	1	2.953	.089
	OLW	116.641	1	6.716	.011

TABLE 5

Descriptive Statistics: Acceptable-Word Method

Dependent Variables	Groups	*M*	*SD*	*n*
K1	High	84.29	3.866	50
	Low	80.91	5.105	59
FW	High	50.02	2.577	50
	Low	47.55	4.234	59
CW	High	34.26	2.999	50
	Low	33.35	4.630	59
K2	High	4.48	1.630	50
	Low	5.25	2.161	59
AWL	High	2.36	1.486	50
	Low	2.88	1.683	59
OLW	High	8.89	3.356	50
	Low	10.96	4.746	59

Taking these study results together, the findings can be summarized as follows. First, the cloze test scores can predict the Korean EFL learners' written productive vocabulary in writing. Second, it was found that the two types of cloze test scoring methods did not lead to widely differing differences in student vocabulary use in writing. The only difference was noted in the use of AWL, showing that there was a significant difference in AWL in the exact-word method. Even though the acceptable-word scoring has been perceived as a much fairer way to score L2 learners' cloze tests in previous studies (Brown, 1980; Brown et al., 2001; Oller, 1972; O'Toole & King, 2011), the present study results did not yield remarkable differences between the two scoring methods in terms of Korean EFL learners' written productive vocabulary. Third, the higher proficiency students used significantly more K1 and FW, whereas the lower proficiency students carried significantly more amounts of K2, AWL, and OLW. This finding is not in line with previous studies (Henricksen, 1999; Laufer, 1991; Laufer & Nation, 1995) which suggest that the more proficient L2 learners are supposed to produce more advanced vocabulary in their writing.

2. General English Test Scores and Written Productive Vocabulary

MANOVA was performed again to determine whether Korean EFL learners' written productive vocabulary differences between the two general English proficiency groups are statistically significant. The effects of the general English proficiency levels were examined by using TOEIC test scores. The participants' TOEIC scores ranged from 920 to 350. The TOEIC score used to differentiate the two proficiency groups was 850 because the participants' TOEIC scores showed that below 850, the first highest score started at 630. Therefore, there seemed to be no use calculating and using the average TOEIC score of the participants. The students who obtained 850 or above on TOIEC were categorized as the higher proficiency group (n = 32) and those with below 850 as the lower proficiency group (n = 82) to maximize the proficiency difference between groups.

As shown in Table 6, the general English language proficiency levels did not make more significant differences to the students' written productive vocabulary in writing than the English language proficiency levels obtained from cloze test scores under the two different scoring methods. The two proficiency groups differentiated by TOEIC scores showed significant differences in K1 and FW alone, showing the higher proficiency group used more K1 and FW. On the other hand, the exact-word cloze test led to significant differences in K1, FW, K2, AWL, and OLW, and the acceptable-word cloze test showed significant differences in K1, FW, K2, and OLW. This finding suggests that the cloze test may better predict Korean EFL learners' written productive

vocabulary than a general English language proficiency test. These findings confirm the previous study results (Keshavarz & Salimi, 2007; Saeedi et al., 2011) that indicate significant relationships between cloze test scores and general English language test scores on TOEFL.

TABLE 6

Tests of Between-Subjects Effects: General English Proficiency Levels

Source	Dependent Variables	Mean Square	*df*	*F*	*p*
General English Proficiency Levels	K1	113.292	1	4.874	.029
	FW	64.200	1	4.701	.032
	CW	6.927	1	.447	.505
	K2	4.547	1	1.204	.275
	AWL	3.956	1	1.590	.210
	OLW	41.881	1	2.246	.137

One possible explanation for this finding could be that the cloze test is more closely involved with actual language production. In other words, both cloze test responses by students and their vocabulary use in writing bear the characteristic of actual language production. Cloze tests require L2 test takers to use their language skills including grammar and vocabulary while reading the text and trying to fill in the gaps with the most appropriate words. Cloze test items are surrounded by ample evidence that can be used for L2 learners to fill in the cloze gaps. Similarly, L2 learners need to use all of their language skills and vocabulary knowledge to produce L2 written productive vocabulary in writing. Thus, completion of cloze test items and production of written vocabulary in writing can be considered a similar process in terms of language production.

Another notable study result was found in K1 and FW, which is the same result with the cloze test scores. As can be seen in Table 7, the higher proficiency students used significantly more amounts of K1 and FW in comparison with the lower proficiency group students. Similarly, Lee and Chen (2009) also revealed that Chinese university students overused function words and high-frequency words in their dissertations. This finding suggests that function words, or high-frequency words can be meaningful indicators of Korean EFL learners' lexical proficiency in writing and obviously need to be further investigated in future studies.

TABLE 7
Descriptive Statistics: General English Proficiency Levels

Dependent Variables	Groups	*M*	*SD*	*n*
K1	High	84.13	3.499	32
	Low	81.91	5.240	82
FW	High	49.96	2.584	32
	Low	48.29	4.041	82
CW	High	34.17	2.531	32
	Low	33.63	4.359	82
K2	High	4.54	1.369	32
	Low	4.98	2.122	82
AWL	High	2.34	1.120	32
	Low	2.75	1.721	82
OLW	High	9.00	3.214	32
	Low	10.35	4.672	82

In a different perspective, this finding about greater amounts of K1 and FW uses of the more proficient learners could be explained in relation to cloze test item characteristics as Kobayashi (2002b) reported that characteristics of cloze items had an influence on cloze test scores. As previously mentioned in the present study, the proportion of function words in cloze test items was higher than content words (number of FW: 35, number of CW: 20). Also, the content words taken as cloze test items mostly belonged to K1. As Kobayashi (2002b) has claimed the effects of cloze test items on cloze test scores, the cloze test items mainly composed of K1 and FW might have led to the significant differences in K1 and FW. It seems likely that the students who obtained the higher scores on the cloze test produced the higher proportion of K1 and FW in their writing. For Korean EFL learners, FW might not be easier to acquire at the earlier stage of learning English or as beginner learners of English even though FW is counted as lower-level vocabulary. Further research is needed to see how different cloze items affect students' cloze test responses.

V. CONCLUSION

The present study examined Korean EFL learners' written productive vocabulary in writing according to cloze test scores and general English language test scores. To look at more closely the relationship between cloze test scores and students' written productive vocabulary, cloze test scores were obtained through the following two scoring methods: the exact-word and the acceptable-word.

The study results demonstrated the following major points. First, it was revealed that the cloze test scores as well as general English proficiency test scores can predict the Korean EFL learners' written productive vocabulary. Interestingly, the cloze test was more efficient than the general English language test in assessing and predicting Korean EFL learners' written productive vocabulary in writing. The general English language test scores made a difference in the use of K1 and FW alone, while cloze test scores led to significant differences in almost all the vocabulary lists provided by *VocabProfile* (Cobb, 2002).

Second, the two cloze test scoring methods did not lead to a very notable difference in students' written productive vocabulary profiles, showing the only difference between the two scoring methods in academic word lists. These findings do not seem to fully support the view that the acceptable-word method is fairer for EFL learners. The two scoring methods did not show a remarkably different vocabulary use in writing. This means that English writing teachers do not have to spend more time marking their students' responses to the cloze test items by using the acceptable-word method. Thus, these study results suggest that the acceptable-word method might not be fairer than the exact-word method, and the exact-word method could serve as a predictor of written productive vocabulary of Korean EFL students.

Third, both the cloze test (regardless of scoring methods) and general English test showed that the higher proficiency students used more amounts of the lower-level vocabulary (K1 and FW). In contrast, the lower proficiency students produced higher amounts of the higher level vocabulary (K2 and OLW) regardless of the cloze scoring methods and general English test scores. This finding suggests that K1 including FW could be a significant predictor of Korean EFL students' productive vocabulary use in writing because both the cloze test and general English test showed that the more proficient students used significantly greater amounts of K1 and FW than their counterparts did. This is noteworthy because conventionally, the students with the higher proficiency level are expected to produce more advanced vocabulary (e.g., AWL and OLW) in their writing. From a slightly different perspective, it also points to the fact that the lower proficiency students might have problems with using grammatical words (FW), such as articles, prepositions, conjunctions, and determiners.

These study results have pedagogical implications in that writing teachers can use the cloze test to predict Korean EFL learners' vocabulary use in writing. Students with the higher scores on cloze test might produce more amounts of K1 words including function words and content words. Also, student responses to cloze items can be used to provide direction for cloze test development and use. Writing teachers can make cloze items function well for students who experience difficulties using function words or more advanced words. Writing teachers should decide which cloze items to choose for their students.

Even though the present study has meaningful implications, it has limitations as well. It is

needed to explore the effects of cloze test item characteristics and different types of general English proficiency tests in future studies. The present study did not use those two variables in the data analysis. Also, future studies need to investigate how Korean EFL learners actually use function words in writing according to their English proficiency levels. There may be still many other variables that have not yet been isolated and studied in relation to cloze testing. The findings from future studies may be helpful in shedding some more light on the research into the cloze test as a valid measure of productive vocabulary used in speaking as well as writing and thus guiding future English classes for Korean EFL learners.

REFERENCES

Aiken, K. G. (1977). Using cloze procedure as an overall language proficiency test. *TESOL Quarterly, 11*(1), 59-78.

Bachman, L. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly, 19*(3), 535-556.

Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *Modern Language Journal, 64*(3), 311-317.

Brown, J. D. (2001). *Teaching by principles: an interactive approach to language pedagogy* (2nd Ed.). New York: Longman.

Brown, J. D., & Hudson. T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*(4)*,* 653-675.

Brown, J. D., Yamashiro, A. D., & Ogane, E. (2001). The emperor's new cloze: Strategies for revising cloze tests. In T. Hudson & J. D. Brown (Eds.), *A Focus on language test development* (pp.143-161). Honolulu, HI: University of Hawaii.

Chae, Eun-Young, & Shin, Jeong-Ah. (2015). A study of a timed cloze test for evaluating L2 proficiency. *English Teaching, 70*(3), 117-134.

Clausing, G., & Senko, D. (1978). On the scoring of cloze tests. *Teaching German*, *11*(2), 74-81.

Cobb, T. (2002). *VocabProfile* program [computer software]. Available from the World Wide Web: http://www.lextutor.ca/vp/

Coxhead, A. (2002). The Academic Word List: A corpus-based word list for academic purposes. In B. Ketteman and G. Marks (Eds.), *Teaching and Language Corpora (TALC) 2000 Conference Proceedings* (pp. 73-89). Atlanta, GA: Rodopi.

Gellert, A. S., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment*, *31*(1), 16-28.

Hanania, E., & Shikhani, M. (1986). Interrelationships among three tests of language

proficiency: Standardized ESL, cloze, and writing. *TESOL Quarterly, 20*(1), 97-109.

Harsch, C., & Hartig, J. (2016). Comparing C-tests and yes/no vocabulary size tests as predictors of receptive language skills. *Language Testing, 33*(4), 555-575.

Henricksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*(2), 303-317.

Ito, A. (2004). Two types of translation tests: their reliability and validity. *System, 32*(3), 395-405.

Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writers, 15*(4), 261-278.

Keck, C. (2014). Copying, paraphrasing, and academic writing development: A reexamination of L1 and L2 summarization practices. *Journal of Second Language Writing, 25*, 4-22.

Keshavarz, H. M., & Salimi, H. (2007). Collocational competence and cloze test performance: A study of Iranian EFL learners. *International Journal of Applied Linguistics, 17*(1), 81-91.

Kobayashi, M. (2002a). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing, 19*(2), 193-220.

Kobayashi, M. (2002b). Cloze test revisited: Exploring item characteristics with special attention to scoring method. *The Modern Language Journal, 86*(4), 571-586.

Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal, 75*(4), 440-448.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307-322.

Lee, Soyoung. (1997). Cloze test as a measure of EFL writing proficiency. *English Teaching*, *52*(3), 151-172.

Lee, D. Y. W., & Chen, S. X. (2009). Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners. *Journal of the Second Language Writing, 18*(4), 281-296.

Meara, P. M., & Bell, H. (2001). P Lex: A simple and effective way of describing the lexical characteristics of short texts. *Prospect, 16*(3), 5-24.

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

Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of teaching English as a second language trainees. *System, 32*(1), 75-87.

Morris, L., & Tremblay, M. (2002). *The interaction of morphology and lexis in the development in ESL learners*. Paper presented at the Canadian Association of Applied Linguistics Conference, Toronto.

Oller, W. (1972). Scoring methods and difficulty levels for cloze tests of ESL proficiency. *Modern Language Journal, 56*(3), 151-158.

O'Toole, J. M., & King RAR. (2011). The deceptive mean: Conceptual scoring of cloze entries differentially advantages more able readers. *Language Testing, 28*(1), 127-144.

Paribakht, T. S., & Webb, S. (2016). The relationship between academic vocabulary coverage and scores on a standardized English proficiency test. *Journal of English for Academic Purposes, 21,* 121-132.

Ryoo, Young-sook (2016). Lexical profiles of Korean EFL summary writing from online spoken texts. *Modern English Education, 17*(3), 79-97.

Sadeghi, K. (2014). Phrase cloze: A better measure of reading? *The Reading Matrix, 14*(1), 76-94.

Saeedi, S., Tavakoli, M., Kazerooni, S. R., & Parvaresh, V. (2011). Do C-test and cloze test procedure measure what they purport to be measuring? A case of criterion-related validity. *International Journal of Human and Social Science, 6*(2), 99-108.

Stubbs, B., & Tucker, R. (1974). The cloze test as a measure of English proficiency. *Modern Language Journal, 58*(5-6), 239-241.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly, 30*(4), 415-433.

Watanabe, Y., & Koyama, D. (2008). A meta-analysis of second language cloze testing research. *Second Language Studies, 26*(2), 103-133.

APPENDIX 1

Text Portion of Cloze Test (C1)

The science of automatic control depends on certain common principles by which an organism, machine, or system regulates itself. Many historical developments up to the present day have helped to identify these principles. For hundreds of years there (1) ____ many examples of automatic control systems, but no connections were recognized among (2) _____. A very early example was a device on windmills designed (3) ____ keep their sails facing into the wind. It consisted simply (4) _____ a miniature windmill which rotated the whole mill (5) _____ face in any direction...

Answers: 1. WERE/existed/have been
2. THEM/systems/these
3. TO
4. OF
5. TO

(Continues for a total of 30 cloze items)

Note. Exact words are capitalized, and acceptable words are lower cases.

APPENDIX 2

Text Portion of Cloze Test (C2)

The industrialized countries between them possess 78% of all existing wealth. This means that the (1) _____ countries, which are usually called the 'Third World', have about 22% of (2)_____ wealth, even though their population is about 76% of the world's total. (3) _____ rich industrialized countries give aid to poorer Third World countries. However, this (4) _____ sometimes does more harm than good. This is because many Western aid (5) _____ are importing Western technology into the poorer countries…

Answers: 1. OTHER
2. THE/the existing
3. MANY/many of the/ the/some/some of the
4. AID/help
5. ORGANIZATIONS

(Continues for a total of 25 cloze items)

Note. Exact words are capitalized, and acceptable words are lower cases.

Examples in: English
Applicable Languages: English
Applicable Levels: Tertiary

Young-sook Ryoo
Department of General English
University of Seoul
163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504
Tel: 02-6490-5217
Email: ysryoo@uos.ac.kr

Received 21 September 2017
Reviewed 24 October 2017
Accepted 11 November 2017