Does the PVLT Provide an Accurate Measure of Productive Vocabulary Knowledge?

Bokyoung Park
(Hanyang University)


Limited research on the PVLT (Productive Vocabulary Levels Test) stimulated the investigation into the relationship between the PVLT and Korean college students’ productive vocabulary use in writing. For the purpose of the study, twenty five students (n=25) were asked to take the PVLT, perform translation tasks, and write an essay. They were also asked to write a reflection paper to examine their perceptions of the PVLT. The LFP (Lexical Frequency Profile) was employed to analyze the students’ compositions to examine the relationships between their vocabulary profiles and the PVLT. The results of the study indicated that there was no significant correlation between the PVLT and the LFP; the relationships between the PVLT and other variables were largely influenced by word frequency levels. These findings imply that the PVLT at the 2000 word frequency level represents the students’ vocabulary knowledge in their compositions, and their writing proficiency. As for the perceptions of the PVLT, most of the students responded that the PVLT assessed their vocabulary knowledge. Interestingly, they attributed both success and failure in providing correct answers to the test format of the PVLT. This study suggests that interpretation of the results of the PVLT should be associated with word frequency levels.

I. INTRODUCTION

Vocabulary has served as a key component of L2 learning and writing because it makes a big contribution to the improvement of L2 learning and provides an effective indicator of overall writing quality (Engber, 1995). This is grounded on many studies that have shown that the quality of vocabulary is closely related to writing. For example, Santos (1988) found that lexical errors are rated as the most serious factor that affects EFL students’ writing. Also, a lack of vocabulary is perceived to cause L2 learners to have more difficulty writing than any other factors (Leki & Carson, 1994). Due to this important role
of vocabulary in the assessment of the quality of written work and learners’ perceptions of writing, many teachers and researchers draw on various ways of measuring learners’ productive vocabulary, such as the Lexical Frequency Profile (Laufer & Nation, 1995) and the Productive Vocabulary Levels Test (Laufer & Nation, 1999). These measures are mostly used to show the percentage of words at different vocabulary frequency levels in learners’ writing and estimate the size of productive vocabulary (Laufer & Nation, 1995).

The Lexical Frequency Profile (LFP) provides teachers with learners’ vocabulary profile, which helps to check the proportion of vocabulary learners use at each of the four frequency levels (the first 1000, the second 1000, the Academic Word List, and the “not in the list” word levels) and the progress they have made in their vocabulary use in writing. The Productive Vocabulary Levels Test (PVLT) assesses learners’ ability to produce words by presenting 18 items at each level (2000, 3000, 5000, Academic Word List, and 10,000 word levels) in the form of a blank-filling test where a few initial letters of the target word are provided. Laufer and Nation (1999) stressed that the PVLT is easy to administer and interpret; in other words, it is simply marked as correct or not, and the score of the test at each level indicates rough number of words “readily available for productive use” (p. 41).

Although these measures are employed in the field of productive vocabulary assessment and claimed to be valid and reliable (Laufer & Nation, 1995, 1999), there are problems inherent in these measures and thus it has been difficult to conclude their effectiveness. Munice (2002) pointed out the complex ways the LFP works. Although the vocabulary at a particular frequency level is not added or removed, the proportion of the vocabulary at that level changes if the vocabulary at other frequency levels is added or removed. This is because the profile always put on a percentage basis. Kim and Ryoo (2009) also raised a question whether the LFP is a reliable measure, reporting the findings that the LFP did not discriminate learners’ different proficiency levels even in the task where two different (difficult and easy) writing topics were given to the same students. This contradicted previous study by Laufer and Nation (1995) which indicated that the LFP shows a stable tendency between learners’ vocabulary use and their writing proficiency; in other words, as learners’ proficiency grows, their use of vocabulary belonging to high frequency level decreases and the use of low frequency vocabulary increases (Laufer, 1998; Laufer & Paribakht, 1998; Morris & Cobb, 2004; Munice, 2002). There are other studies warning that more careful attention is needed when employing the LFP to assess learners’ productive vocabulary knowledge. Learners’ vocabulary improvement does not necessarily mean the change of their vocabulary profile; rather, being able to produce vocabulary, even though it is still at high frequency level, according to context should be considered as an important characteristic that determines learners’ vocabulary improvement (Horst & Collins, 2006).

In addition to the LFP, the PVLT is also problematic. It seems that the PVLT, which
presents differing number of initial letters of the target word, cannot measure actual productive vocabulary knowledge in light of the fact that it does not allow test takers to use vocabulary they intend in the context of actual writing. Laufer and Nation (1999) simply emphasized the validity of the measure by mentioning that the PVLT scores for the four groups of foreign learners of English become higher as their proficiency levels increase. The problem here, however, is that their proficiency level did not involve any written work, but involved their class grades (10th, 11th, and 12th graders, and first year university students). Moreover, in another study (Laufer & Nation, 1995), the subjects’ proficiency level was classified into three groups based on the results of the two different types of tests, and the lowest proficiency group was foreign learners of English with various backgrounds and the other two groups were the Israeli students in the department of English Language and Literature. Therefore, it is hard to believe that the findings are plausible because the validity of the results is reduced with the two different tests and subjects.

Considering the lack of empirical research conducted to verify the effectiveness of productive vocabulary measure, the need for further investigation is inevitable. In order to do so, the PVLT should be dealt with in conjunction with learners’ actual vocabulary use and their actual writing. Learners’ actual vocabulary use involves allowing learners to produce vocabulary when prompted by its meaning, not prompted by a few initial letters of the predetermined words. Learners’ actual writing is also needed in order to verify whether and to what extent the vocabulary knowledge assessed by the PVLT can be explained in association with it. This is because learners’ choice of vocabulary can be affected by various factors such as the organization and content of the writing, their familiarity with the topic, their level of confidence in the words they produce in relation to the surrounding words, the spellings, and so on. These attempts to verify the PVLT can also provide an opportunity to make a meaningful contribution to more reliable and clearer interpretation of the LFP unlike the previous study that simply asserted the LFP’s validity and reliability by comparing with the scores on the PVLT that has inherent problems. Furthermore, a notable problem is that no study has been conducted on how test takers perceive the PVLT; this study is unique in that it deals with test takers’ responses to the test tool, not the opinions from the professionals in the related field. Therefore, this current study aims to examine whether the PVLT is an accurate measure of productive vocabulary knowledge and how the PVLT can be interpreted in association with learners’ actual use of vocabulary, their actual writing, their writing proficiency, and their perceptions of the PVLT. The following research questions are set up to address these issues:

1. What does the PVLT show in terms of learners’ actual use of vocabulary?
2. What does the PVLT show in terms of learners’ actual writing?
3. What are the learners' perceptions of the PVLT?

II. LITERATURE REVIEW

1. Receptive and Productive Vocabulary Knowledge

There have been many attempts to explain what is meant by knowing a word. This reflects the difficulty associated with defining vocabulary knowledge since it involves more than understanding the meaning of a word. However, many researchers agree that learners' vocabulary knowledge can be divided into receptive and productive vocabulary knowledge (Laufer, 1998; Laufer & Paribakht, 1998; Nation, 2001; Read, 2000). Receptive vocabulary knowledge generally involves comprehending the form of a word while listening or reading. Productive vocabulary knowledge involves expressing meaning through the appropriate spoken or written word form. According to Nation's (1990) classification, productive vocabulary knowledge is a higher level of knowledge than receptive one since it requires more comprehensive knowledge beyond understanding a word.

Along with the terms for receptive and productive vocabulary, many researchers use alternative terms in the field. For example, receptive vocabulary is referred to as passive vocabulary, comprehension, or recognition; on the other hand, productive vocabulary is replaced with active vocabulary, production, or actual or possible use (Melka, 1997). Laufer (1998) used the terms "passive" and "active" and divided the active vocabulary into controlled active and free active vocabulary. Controlled active vocabulary knowledge is defined as being able to produce vocabulary elicited by a certain task where the learners are compelled to do so by teachers while free active vocabulary knowledge is referred to as the ability to use a word without being prompted by a task.

It is generally assumed that receptive vocabulary knowledge is viewed as being easier to learn and thus acquired first; then later, productive knowledge is obtained (de la Fuente, 2002; Laufer & Paribakht, 1998). In terms of the size of receptive and productive vocabulary, some studies indicated that the gap between those two is very small; for example, the vocabulary Finnish learners of English produced was almost as large as their receptive vocabulary (Takala, 1984, as cited in Melka, 1997). However, following several estimates of receptive and productive vocabulary knowledge, it is well accepted in the field that receptive vocabulary is larger than productive vocabulary (Clark, 1993; Fan, 2000; Laufer, 1998; Laufer & Paribakht, 1998; Melka, 1997; Nation, 1990; Waring, 1997; Webb, 2008). More specifically, Makarchuk (2010) showed the degree of difference between the receptive and productive vocabulary sizes was associated with word
frequency level and instructional practices in Korean classroom settings. No significant difference was found at the 1,000 word frequency level but as the frequency level decreased, the size of the receptive vocabulary was larger than productive, which was influenced by a decontextualized manner in which vocabulary is taught.

Even though the gap between the size of the receptive and productive vocabulary knowledge is argued by many researchers, the distance between the two can become narrower as productive vocabulary knowledge develops. Morgan and Oberdeck (1930) reported that receptive vocabulary develops relatively faster than productive vocabulary does at first, but the gap between those two vocabularies diminishes as productive vocabulary grows later.

2. Productive Vocabulary Measures

Although it is generally accepted that vocabulary knowledge can be divided into receptive and productive aspects, it seems that the majority of the tests designed to measure the size of vocabulary tend to focus on receptive vocabulary. This can be presumed from the findings reviewed above that learners learn receptive vocabulary prior to productive vocabulary and thus measuring their receptive vocabulary prior to productive vocabulary seems to be logical. However, the more convincing reason why there are more tests designed for receptive than for productive vocabulary is that accurate measure of productive vocabulary knowledge is quite impossible because the vocabulary produced by a learner “tend to be so context-specific that it is difficult to calculate from a small sample the true size or range of the learner’s productive vocabulary” (Meara & Fitzpatrick, 2000, p. 20).

Meanwhile, there are two measures commonly used in the field of productive vocabulary assessment: the PVLT and the LFP. Nation (2001) introduced the PVLT which samples 18 items at each of the five word frequency levels (2000, 3000, 5000, the Academic Word List, and 10,000 word levels). A sentence in meaningful context is provided with a few initial letters of the target word in order to elicit test takers to produce only the intended word. Each level represents 1000 words, which means if a test taker has 9 out of 18 items correct at a certain level, this roughly indicates he or she knows about 500 out of 1000 words at that level. The guide to the PVLT suggests that it is above around 80 percent of correct answers (approximately 16 items) that is considered as satisfactory mastery of the level.

Another measure, the LFP designed by Laufer and Nation (1995), is also used to obtain information about learners’ productive vocabulary. This measure uses a computer program which shows a proportion of productive vocabulary at a different frequency level (1000, 2000, the Academic Word List, and “not in the list” word levels) in the learners’
compositions. The final ratio of the words produced by learners helps teachers check their stage in their vocabulary development (Laufer & Nation, 1999). It is claimed that the LFP shows consistent patterns for different pieces of writing by the same learner regardless of a writing topic and a stable tendency between learners’ vocabulary use and their writing proficiency. However, other studies hold different opinions on the previous studies (Kim & Ryoo, 2009; Lee & Anderson, 2007).

There are a number of drawbacks in these two measures. The main disadvantage of the PVLT is that even though it is claimed that the scores represent the approximate number of words that can be produced at each frequency level, it is difficult to extrapolate the size of productive vocabulary from such a small test item (18 items at each frequency level). Particularly, at low frequency levels such as beyond 5000 word level, it would be impossible to provide reliable estimates about the test takers’ productive vocabulary knowledge by testing only 18 words from several thousand words (Meara & Fitzpatrick, 2000). The blank-filling test format is also questionable. It allows test takers to produce only the predetermined words. In addition, since each blank presents differing number of initial letters of the target word, test takers are sometimes required to demonstrate more complex word knowledge such as collocations in order to answer the target words (Shin, Chon, & Kim, 2011). This controlled PVLT prompted by predetermined target words was criticized for not being clear about what the test measures (Read, 2000). However, Laufer and Nation (1999) insisted upon the validity of the test measure by mentioning that learners whose proficiency is higher gain significantly higher scores on the PVLT.

In addition to the PVLT, the LFP is also considered to be problematic in that it is easily affected by the variables such as context and length of the compositions produced by learners. Laufer and Nation (1995) added that those problems can be solved by adopting a broad subject and encouraging learners to generate two 300-word essays to secure enough words to measure their productive vocabulary knowledge. However, it still raises doubts whether the texts produced in that situation genuinely show learners’ vocabulary knowledge since the vocabulary produced by test takers predominantly consists of a small set of high frequency level words (Meara & Fitzpatrick, 2000). Lee and Munice (2006) also pointed out that the LFP can simply deal with individual words without any consideration of lexical phrases.

So far, researchers in the field of vocabulary assessment have used the PVLT along with the Vocabulary Levels Test (VLT) to compare productive and receptive vocabulary knowledge in terms of size. The VLT (Nation, 1990) assesses receptive vocabulary knowledge at five different word frequency levels (2000, 3000, 5000, the Academic Word List, and “not in the list” word levels) in the form of word-meaning matching. The major findings of previous studies revealed that learners tend to possess more receptive vocabulary than productive (Clark, 1993; Fan, 2000; Laufer, 1998; Laufer & Paribakht,
Does the PVLT Provide an Accurate Measure of Productive Vocabulary Knowledge? 165

1998; Melka, 1997; Nation, 1990; Waring, 1997; Webb, 2008). In addition, Laufer and Nation (1995) used the PVLT to prove the reliability of the LFP and concluded that both tools are reliable measures of learners' lexical use.

Whether the PVLT is used to measure productive vocabulary knowledge in the aspect of size or employed as an evidence for the validity of other measure tools such as the LFP, more research is needed to focus on how well the PVLT as an independent measure shows learners’ productive vocabulary knowledge and, if not, to identify what its blank-filling test measures and does not measure. Therefore, this present study aims to investigate the relationships among the PVLT, learners’ actual vocabulary use, and their actual writing in conjunction with their writing proficiency. Also, learners’ perceptions of the PVLT are dealt with to provide more precise insight into the use of the measure.

III. METHOD

1. Participants

The participants in this study were 27 Korean college students who took a three-credit TOEIC course that met four times a week for 75 minutes for four weeks during the 2011 summer semester. However, the number decreased to 25 since two students who failed to complete all the tasks given in class were excluded from the data set. Although it seems that the number of the participants was too small to provide statistically significant outcomes, this study is of great significance in that it focused on challenging general assumptions about the PVLT by comparing the students’ vocabulary knowledge when assessed by the PVLT where learners were required to produce predetermined target words and by an authentic task where learners were allowed to produce vocabulary they intend according to context. This first ever attempt also makes a significant contribution to prompting further investigation in the field of vocabulary assessment.

The class was an elective course. The students were mostly juniors and seniors, and a homogeneous group in terms of age ranging from 21 to 28 with an English education background. They received English instruction for six years before entering the university. The students were highly motivated to learn vocabulary and writing, which they strongly believe helps them improve their L2 proficiency. These students were from various fields of study: English Language and Literature, Korean Language and Literature, Law, History, Engineering, and Business Administration.
2. Instrument

The present study used the PVLT to measure the students’ productive vocabulary knowledge, which is available at http://www.lextutor.ca. The test contains a set of sentences with a blank and each blank is provided with differing number of initial letters of the target word in order to induce test takers to produce only the predetermined words. The test takers were required to fill in the blanks for 36 (18x2) target words, which were drawn from two word frequency levels: the 2000 word (K2) and 3000 word (K3) frequency levels. This study excluded 5000, the Academic Word List (AWL), and 10,000 word levels since the words in these levels are so difficult to create a statistical floor effect (Waring, 1997).

In order to gather the students’ perceptions of the PVLT, they were asked to write a reflection paper answering four open-ended questions right after taking the PVLT. In this way, the students can have more time to reflect on their feelings and provide more organized thoughts about their experience of taking the PVLT rather than impromptu responses through an oral interview. The four questions were included and provided in Korean: the perceived purpose of the test, the variables that make the test difficult and easy, and the opinions about the test. The respondents were allowed to provide as many responses as they want for each question.

The translation task designed by the researcher was used in order to examine whether the students can produce the target words drawn from the same frequency level if presented in a different way than when the PVLT was taken. This task allowed the students to produce the words according to the context while prompted by Korean, not prompted by the artificially controlled test format to elicit the predetermined words. The target words for the translation task were drawn from the other version of the PVLT at the K2 and K3 frequency level respectively. The reason the other version of the PVLT was used is that the validity of the results is reduced if the same words, even though they are two different languages, are tested in the two different tests. About 14 to 15 target words at each frequency level were put together into a text in Korean with some relevant pictures to help the students understand the text better. The students were asked to read the story and translate into English. After the task, some students were selected at random for follow-up interview to examine the difference in the degree of difficulty of the two tasks.

Finally, a writing task was designed to see if there is any relationship between the PVLT and the students’ writing and, if so, whether the PVLT can show the students’ writing proficiency. The students were asked to take a position on the issue “the money is the most important aspect of a job.” The topic was chosen because this kind of question seems likely to be asked on the TOEIC writing test and was relevant and interesting enough for college students to have much to write about. They were encouraged to write more than
Does the PVLT Provide an Accurate Measure of Productive Vocabulary Knowledge?  

300 words to yield stable results as Laufer and Nation (1995) suggested, but they were not allowed to use a dictionary and discuss with their peers for consultation. They were given 30 minutes to complete the task because it is the amount of time allowed in the TOEIC and TOEFL writing and it is common to allow test takers to spend 30 to 60 minutes in most of the studies related to writing (Silva, 1993).

In order to obtain the Lexical Frequency Profile (LFP), the VocabProfile (Cobb, 2002) was used in this present study. The program, VocabProfile is also available at http://www.lextutor.ca and provides profiles of vocabulary on the basis of the proportion of word families in the following four frequency levels: the first 1000 words (K1), the second 1000 (K2), the Academic Word List (AWL), and words that are “not in the list” (NIL). The VocabProfile indicates the percentage of words at each frequency level learners produce in their writing.

3. Data Collection Procedure

In order to measure the students’ productive vocabulary, they were asked to take the version A of the PVLT (see Appendix A). The test can be taken online, which is available at http://www.lextutor.ca, but it was conducted in a paper-and-pencil format in order for the students not to be exposed to any vocabulary in the other version of the PVLT, which was intended for later use for the translation task. The frequency levels used in this study consisted of two levels: the 2000 and 3000 word frequency levels (K2 and K3). The reason why these two bands were employed is that according to the previous study by Waring (1997), the scores on the PVLT at the 5000 word band were so low that it is likely to create a statistical floor effect. The students were asked to fill in the blanks for all 36 items (18 items for each K2 and K3 band respectively) without resorting to any kind of dictionary and consultation with their peers and teacher. Then, the test was scored by awarding one point for a grammatically and semantically correct word and zero point for an incorrect one. Unlike the study by Laufer and Nation (1999) where spelling and grammatical mistakes were counted as correct, those mistakes were marked as incorrect in this study since it is quite obvious that productive vocabulary knowledge includes correct spellings and grammatically correct words.

Right after taking the PVLT, the students were asked to write a reflection paper answering four open-ended questions and submit it. Since most of the students provided more than one response to each question, the total number of the responses was more than 25. Follow-up interviews were conducted when it was necessary to clarify the answers from their reflection paper. All the responses were analyzed by the researcher. The individual responses were coded first and then clustered under categories into which the related responses can fall. As a result, there were various categories under the four
different questions. Then, the responses were counted and ranked according to the frequency.

For the translation tasks at the two frequency levels (K2 and K3), the students were given the text to translate into English. Each text at the K2 and K3 band was composed of approximately 60 Korean words and contained about 14 to 15 target words drawn from the different version of the PVLT so as to reduce the possibility of giving the students opportunity to practice the words exposed when they took the blank-filling test (the PVLT). The text was presented with some pictures illustrating the main points of the story to help the students understand the text better. The instruction was given to the students for each task by the researcher. They were encouraged to translate what was written as much in detail as possible. Two translation tasks were conducted during class time in one week. Following all the tasks, some students were randomly asked about which tasks (K2 and K3) they felt were more difficult to compare the degree of difficulty of the two tasks. The students’ fifty (25x2) compositions produced for the two translation tasks (K2 and K3) were collected and analyzed. For the assessment, a score of one was given if the target word was grammatically and semantically correct and zero if not.

In order to examine the relationship between the PVLT and the students’ writing, they were asked to perform the writing task in which they were asked to clarify their position regarding “the most important aspect of a job is the money.” They were supposed to write at least 300 words on the topic and were allowed to have 30 minutes to complete the task just as indicated in the guide to the TOEIC. Then, their compositions were rated by the two Korean instructors with more than 4 years of experience in teaching writing. At first, they rated students’ compositions using the ETS assessment rubric independently and compared the scores later. When there was discrepancy between their marks, they discussed the ETS rating schemes together and adjusted their scores. By doing so, they were able to come to an agreement.

For the purpose of the study which was to examine whether there is any relationship between the PVLT and the students’ writing proficiency, the students were classified into a different proficiency group. The students with the score 5 and 6 were assigned to a high proficiency group (n=7) and the ones with the score 1 and 2 to a low proficiency group (n=8). Although the number of the subjects who participated in this study was too small to be divided into different groups and to provide statistically significant outcomes, learners’ writing proficiency should be dealt with in order to verify whether the PVLT is valid as argued in the previous study (Laufer & Nation, 1999), where proficiency was adopted to prove the validity of the PVLT. However, the outcomes should not be considered as an independent indicator to verify the PVLT, but as a supplement to other findings in this study.
4. Data Analysis

The present study used the VocabProfile (Cobb, 2002) to analyze the students’ 25 texts for their LFP. In order to do so, all the texts were typed and stored in the computer. The texts were not corrected at all by the researcher. However, it should be noted that the words with spelling mistakes are not recognized in the program and then categorized as “not in the list” words in students’ LFP since they are literally “not in the list” words. The problem here is that the ratio the LFP shows at each frequency level (1000 word, 2000 word, the Academic Word List, and “not in the list” word level) is affected interactively by the proportion of this four frequency levels. In other words, if unknown words were added in the “not in the list” word level, the ratio of words at K1, K2, and the Academic Word List level would change. Therefore, the words with spelling mistakes were excluded to produce more reliable vocabulary profiles as suggested by Laufer and Nation (1995) and Munice (2002).

For the data analysis, Pearson’s linear correlation coefficient was used to examine the relationships between the PVLT and other variables: the translation tasks, students’ compositions (the LFP), and their writing proficiency. Also, a multivariate analysis of variance (MANOVA) was performed using SPSS version 18 to analyze the differences in the scores on the PVLT (K2 and K3) according to the students’ writing proficiency. In order to obtain the results from the students’ reflection paper, all the answers were categorized and the frequency of each category was counted.

IV. RESULTS AND DISCUSSION

1. The PVLT and Learners’ Actual Use of Vocabulary

The first research question on the relationship between the PVLT and the actual use of vocabulary was examined using Pearson’s linear correlation coefficient. The results showed that no significant correlation was found between the PVLT and the translation task at the K2 band; however, a significant correlation was noted between the PVLT and the translation task at the K3 band (see Table 1). One of the assumptions here is that the students probably feel more comfortable with the words and context provided for the translation task at the K3 band. However, according to the random interview conducted after the translation tasks, most of the students responded that the words and the context given for the translation task at the K3 band were more difficult to generate than those provided for the translation task at the K2 band.
TABLE 1

Correlations between the PVLT and the Translation Task

<table>
<thead>
<tr>
<th></th>
<th>PVLT (K2)</th>
<th>Translation (K2)</th>
<th>PVLT (K3)</th>
<th>Translation (K3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PVLT (K2)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation (K2)</td>
<td>.348</td>
<td>1</td>
<td>.344</td>
<td>1</td>
</tr>
<tr>
<td>PVLT (K3)</td>
<td>.820</td>
<td>.420*</td>
<td>.644*</td>
<td>1</td>
</tr>
<tr>
<td>Translation (K3)</td>
<td>.624*</td>
<td>.420*</td>
<td>.644*</td>
<td>1</td>
</tr>
</tbody>
</table>

*Correlation is significant at the .05 level.

To further explain why there was no significant correlation between the two variables at the K2 band, all compositions were carefully analyzed. It was found that some students who gained higher scores on the PVLT at the K2 band but lower scores on the translation task at its corresponding level provided many other alternative words, which were not counted as correct answers, that still have the same meaning as the target words; for example, instead of using the target word “wage,” some students wrote the word “pay” or “salary” which are not the target words, but still proper words and belong to the same frequency level. They also produced low frequency words such as “restrict” instead of the target word “limit.” Due to the students’ use of wide-ranging vocabulary, the PVLT was weakly correlated with the translation task at the K2 band. In other words, the PVLT is not effective measure since the items in the PVLT are not large enough to show the full extent of their productive vocabulary.

Unlike the finding at the K2 band, however, a significant correlation ($r = .644, p = .001$) was found between the scores on the PVLT at the K3 band and the translation task at its corresponding band, which implies that learners who obtain higher scores on the PVLT at the K3 band are likely to produce the target words in the translation task and vice versa. Based on the responses from the random interview regarding the perceived degree of difficulty of the words in the two bands and the careful analysis of the students’ translation work, it was discovered that the students could not provide as various alternative words as they did in the translation task at the K2 band; they simply either provided the target words if they could or gave up producing them. It is probably because there were not many alternative words they could produce for the words, for example “lieutenant,” “vein,” “marble,” and “draft.”

This finding is notable because it challenges the assumption that the words properly answered in the PVLT, where a few initial letters are given for the target words to be elicited, are likely to be produced in the learners’ writing. Even though it is claimed that the PVLT scores show the approximate number of words that are ready for productive use at a certain level, it is very implausible because the PVLT excludes the possibilities of displaying other acceptable synonyms of the target words. Furthermore, even though there are many other things involved when students push their vocabulary in mind into
productive words in their writing, such as their past experience of vocabulary use, the
degree of familiarity with the word, educational milieu where the vocabulary is taught,
personal preference for high or low level words, those factors have no bearing on the
PVLT. As pointed out in the study by Meara and Fitzpatrick (2000), since any other words
except the target words are not accepted, the PVLT is more likely to show what the test
takers do not know rather than they know. The finding confirms that it is one thing how
well learners perform for the blank-filling task; it is another whether they produce
appropriate words according to the context.

Particularly notable is that this finding also brings into question the whole purpose of
the classification of the word frequency level in the PVLT. Learners seemed to manage to
choose a variety of proper words they want to use in their writing regardless of the word
frequency level. This implies that just because the students gain higher scores on the PVLT
at a particular frequency level does not necessarily mean that they can produce those
words in their writing. Put differently, even students who gain lower scores on the PVLT
are likely to generate proper alternative words at the same or other frequency levels as
long as more authentic and rich context is provided. According to the guide to the PVLT,
test-takers whose score below 50 percent should work on the higher word frequency level
than that is being tested, and if they score 50 to 80 percent, then they should work on the
level that is being tested. More than 80 percent is considered as successful mastery of a
certain frequency level. However, as seen from the students’ translation tasks, they did not
produce words according to the successive word frequency level. Considering the fact that
the proper vocabulary use does not involve the acquisition of the successive word
frequency levels, the students and teachers should be advised not to take the guide
seriously for effective productive vocabulary use.

The finding implies that the PVLT scores do not necessarily show students’ productive
vocabulary use in the translation task, and that learners’ productive vocabulary should be
assessed in such a way that encourages test takers to generate vocabulary through the
performance involving the use of language in the real-life situation rather than a blank-
filling test. This finding lends support to the suggestion that more careful attention need to
be paid to assessing learners’ productive vocabulary when using the PVLT and designing
the class activities involved in vocabulary.

2. The PVLT and Learners’ Actual Writing

In addition to the investigation on what the PVLT shows in terms of the students’ actual
vocabulary use through the translation tasks, in order to further examine whether there is
any relationship between the scores on the PVLT and the students’ compositions, another
correlation analysis was conducted. Table 2 presents the results.
As seen in Table 2, the scores on the PVLT at the K2 band were highly correlated with the writing scores. However, looking at the finding above (see Table 1), it was revealed that the scores on the PVLT at the K2 band represent neither learners’ actual productive vocabulary nor the size of approximate number of it. Therefore, it should be pointed out that this is more of a correlation between the learners’ receptive vocabulary knowledge and writing since the PVLT at the K2 band did not show vocabulary knowledge ready for production. This is also supported by the argument made by Read (2000), which suggests that the PVLT may be another way of assessing receptive vocabulary knowledge in that it correlates substantially with the VLT rather than the LFP measure. Clearly, this finding indicated that relatively higher level (K3) of vocabulary knowledge is not a good indicator of the overall quality of writing.

To investigate the relationship between the PVLT and the LFP, the students’ written texts produced were analyzed using the VocabProfile. As in Table 3, the results indicated that there were no significant correlations noted between the composite scores on the PVLT and the LFPs of the students’ compositions. The composite score on the PVLT was adopted in this study. Since the PVLT has no levels such as the 1000 and “not in the list” words that the LFP has (Laufer & Nation, 1995), it seems rational to use the composite score in order to properly compare with each level of the LFP.

<table>
<thead>
<tr>
<th>Composite</th>
<th>K1</th>
<th>K2</th>
<th>AWL</th>
<th>NIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1</td>
<td>-.120</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K2</td>
<td>-.015</td>
<td>.183</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AWL</td>
<td>.218</td>
<td>-.837*</td>
<td>-.469*</td>
<td>1</td>
</tr>
<tr>
<td>NIL</td>
<td>-.064</td>
<td>-.710*</td>
<td>-.410*</td>
<td>.402*</td>
</tr>
</tbody>
</table>

*Correlation is significant at the .05 level.

This finding is not in line with the results of Laufer and Nation’s (1995) study which suggested that learners who obtain higher scores on the PVLT produce more words at low frequency bands than their counterpart does. As a result, it can be presumed that higher
scores on the PVLT cannot be used to show learners’ ability to produce “more of the more sophisticated vocabulary” (Laufer & Nation, 1995, p. 317).

Laufer and Nation (1995) used the PVLT in their study as one of the ways to address the concurrent validity of the LFP. However, the finding from this present study indicated that the PVLT, which has inherent problems, cannot be used to represent the students’ actual use of vocabulary (see Table 1) nor can display their use of high and low frequency vocabulary (see Table 3). This raises doubts about whether the LFP itself is a reliable measure to assess students’ productive vocabulary. One of the arguments used to prove the validity of the LFP is that the LFP can discriminate students’ writing proficiency (Laufer & Nation, 1995). For the purpose of this issue, in this study the students’ actual writing was adopted to classify the students into different proficiency groups unlike Laufer and Nation’s study (1995) where students’ written work was not used to identify their proficiency, which would have yielded different outcomes. In order to find out whether the students’ writing proficiency can be explained relating to the LFP, the LFP means of the different proficiency groups at each frequency level were compared. Then, a MANOVA was performed to analyze the differences in the LFP according to different proficiency groups.

As shown from Table 4, the high proficiency group of the students used more words at the K2 and AWL bands than the low proficiency group. In a post hoc analysis of MANOVA using the Scheffe procedure, however, it was found not to be significantly different. This implies that the students showed similar patterns (word frequency level) in their use of vocabulary regardless of their writing proficiency. This finding contradicts with Laufer and Nation’s (1995) argument that the LFP can discriminate students’ proficiency and is “a useful diagnostic tool as well as a sensitive research tool” (p.319). This finding suggests that teachers and researchers should not heavily depend on the previously reported findings of the LFP when assessing the quality of their students’ written work since it cannot show other important factors that affect learners’ writing, such as the ability to produce appropriate words according to context and lexical phrases.

Further investigation was conducted to examine if there was any difference in the scores on the PVLT according to the students’ writing proficiency. Although the number of
students in this study was too small to represent each group of different proficiency and to yield statistically significant results, the findings for this research issue can be used to give a supplementary explanation about how the PVLT can be interpreted in the field of assessment in conjunction with other findings in this study, and would make meaningful contributions to prompting further investigation in this field.

TABLE 5
Descriptive Statistics

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Writing Proficiency</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>PVLT (K2)</td>
<td>H</td>
<td>77.7714</td>
<td>12.82442</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>51.4125</td>
<td>8.27310</td>
<td>8</td>
</tr>
<tr>
<td>PVLT (K3)</td>
<td>H</td>
<td>42.1429</td>
<td>16.45629</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>25.0000</td>
<td>10.52887</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 5 summarizes the mean differences of the PVLT between the high and low proficiency group. The differences seem particularly larger in the score on the PVLT at the K2 band than the K3 band. In order to examine whether there were statistically significant differences in the PVLT between the two proficiency groups, the test of between-subject effects was performed using a MANOVA.

TABLE 6
Test of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent Variables</th>
<th>Mean Square</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing Proficiency</td>
<td>PVLT (K2)</td>
<td>1607.641</td>
<td>2</td>
<td>8.222</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>PVLT (K3)</td>
<td>797.001</td>
<td>2</td>
<td>3.423</td>
<td>0.51</td>
</tr>
</tbody>
</table>

The result suggested that the PVLT scores at the K2 band, not at the K3 band, differ according to the students’ writing proficiency. As seen in Table 2 and Table 6, the PVLT scores at the K3 band did not indicate any correlation \(r = .382, p = .059\) with the students’ compositions and nor did they show any difference in terms of learners’ writing proficiency. This finding contradicts the previous study by Laufer and Nation (1999) arguing that the PVLT “distinguishes among most of the proficiency groups” (p. 40). Contrary to their argument, however, this study indicated that students whose writing proficiency is higher did not gain higher scores on the PVLT at a relatively lower frequency level, which is the K3 band, than those whose writing proficiency is relatively lower. It is logically presumed that it must be difficult to adopt the PVLT as an indicator of learners’ writing proficiency because each frequency level of the PVLT leads to different interpretation in different context.
Judging from the findings above, however, this study instead proposes the PVL T at the K2 band, in comparison with that at the K3 band, as an effective indicator of the students' writing since the PVL T at the K2 band is highly correlated with the students' writing \((r = .598, p = .002)\) and their writing proficiency \((F = 8.222, p < .05)\). This finding confirms that vocabulary knowledge at a lower frequency level does not have an effect on the quality of writing; rather, appropriate vocabulary use according to context and other skills such as content and organization involved in writing can discriminate learners' writing proficiency.

3. Learners' Perceptions of the PVL T

To identify the test takers’ view on the PVL T, the students were asked to write a reflection paper answering four open-ended questions so that they could have enough time to provide their opinion. The students wrote more than one response to each open-ended question provided; thus, the total number of responses was more than 25. Each response in the students’ reflection paper was open coded for emergent themes, and then the answers were categorized and counted for their frequency. When asked the first question, “what do you think this test measures?” 52% (20 responses out of 38) of the students’ responses indicated that the test assessed their vocabulary knowledge, and 26% (10 out of 38) of the responses showed that the test was intended to assess the ability to interpret the sentence. Spelling test was the third most frequent response (13%).

As to the second question about “what was difficult when you were taking the test?” the respondents pointed out the manner in which they should provide only the target word starting with a few given letters, which accounted for 34% (12 responses out of 35) of the responses. Some students reported in great detail in their reflection paper that they found it frustrating not to provide any other alternative words even though they were sure of what is appropriate in the blank as far as the meaning of the words is concerned. The second most frequent response, which accounted for 31% (11 responses out of 35) of the responses, indicated their lack of vocabulary as a reason why they felt the test was difficult. It is interesting to note that a few students pointed out that their habits of memorizing the meaning of the L2 in Korean caused them to have difficulty producing the target words according to the context. A small number of responses showed that the students were having a hard time interpreting the sentences in the PVL T.

Taken from the responses to the third question about what makes the test easy when taking the PVL T, a clue from the surrounding context words and from the initial letters of the target word accounted for 35% (7 responses out of 20) and 25% (5 out of 20) of the responses respectively. 35% (7 out of 20) of the responses indicated that some test items were easy because they already knew them.
The students gave more various responses concerning the last question, which asked them to provide their personal feelings and thoughts about the test. Table 7 presents a summary of the results from the students’ reflection paper.

**TABLE 7**

Students' Perceptions of Taking the PVLT

<table>
<thead>
<tr>
<th>Item</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of vocabulary</td>
<td>6 (23%)</td>
</tr>
<tr>
<td>Urgent need for learning vocabulary</td>
<td>5 (19.2%)</td>
</tr>
<tr>
<td>Not multiple choice</td>
<td>4 (15.3%)</td>
</tr>
<tr>
<td>Interesting</td>
<td>3 (11.5%)</td>
</tr>
<tr>
<td>Difficult</td>
<td>2 (7.6%)</td>
</tr>
<tr>
<td>Unfamiliar</td>
<td>2 (7.6%)</td>
</tr>
<tr>
<td>Not effective test tool to identify language proficiency</td>
<td>2 (7.6%)</td>
</tr>
<tr>
<td>Others</td>
<td>2 (7.6%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>26 (100.00%)</strong></td>
</tr>
</tbody>
</table>

Most of the students mentioned that the test reflected their lack of vocabulary and they felt urgent need for studying vocabulary. This type of response was also identified as the second most frequent response to the second question regarding the matters that make the test difficult. Some respondents complained about the PVLT; for example, they said that it could have been better if the test had provided multiple choice answers and that it is not a good test tool for assessing their language proficiency since it requires only the target words.

To summarize all the responses, overall, most of the students perceived that the PVLT assessed their vocabulary knowledge and the test reflected their lack of vocabulary, which led them to feel the urgent need for vocabulary learning. In addition to the response regarding vocabulary, the second most frequent response was in regard to the test format in which each blank is presented with a few initial letters of the target word. It is interesting to note that this type of format was referred to as the contributing factors to making the test challenging and easy at the same time. That is, it seems that the students attributed not only their failure to answer in the PVLT to those a few letters when they do not know the target words but also success in finding a correct answer to those format when they know the words.
V. CONCLUSION

The present study explored whether the PVLT can be an accurate measure of productive vocabulary knowledge. The students were asked to take the PVLT (version A) at the K2 and K3 frequency levels and then write a reflection paper to identify how they perceive the PVLT. In order to see whether the results of the PVLT can represent the students' productive vocabulary use and if not, how and to what extent the PVLT can be interpreted in vocabulary assessment, they were asked to perform the translation tasks where the target words were drawn from the version B of the PVLT at the K2 and K3 bands. The students were also asked to write an essay to investigate the relationship among the PVLT, their writing, and their writing proficiency. All 25 compositions were evaluated by two Korean raters based on the ETS assessment rubric and groups were formed according to proficiency. After the students were assigned to the high and low proficiency group, the students' vocabulary in their compositions was analyzed with the Vocabprofile to obtain the LFPs and to find out if their vocabulary profiles differed in terms of the scores on the PVLT and their writing proficiency.

First, results from this study suggested that there was significant correlation between the scores on the PVLT at the K3 frequency level and the translation task at the corresponding level. This means that learners who obtain higher scores on the PVLT at the K3 band also have a higher tendency to produce the target words in the translation task and vice versa. However, no significant correlation was observed between the PVLT scores at the K2 band and the translation task at the same frequency level. In other words, if students obtain higher scores on the blank-filling test at the K2 band, it does not necessarily mean that they tend to produce those target words in writing and vice versa. The reason why there were inconsistent results depending on the frequency level was that the students provided more various acceptable alternative words regardless of frequency level for the translation task at the K2 band than at the K3 band. This implies that the items in the PVLT at the K2 band were not large enough to cover the full extent of the test takers' productive vocabulary and that the PVLT is not a reliable measure of productive vocabulary use for the higher word frequency level.

Second, for the purpose of the investigation into the relationship between the PVLT and learners' writing, and what the PVLT shows regarding learners' vocabulary profiles, the students' compositions were analyzed. The results reported that there was a significant correlation between the PVLT at the K2 band and learners' compositions \( (r = .598, p = .002) \); however, no significant correlation was observed between the PVLT and learners' LFP. In other words, the scores on the PVLT did not differentiate learners' vocabulary profiles in their writing. This finding was not in line with the previous study (Laufer & Nation, 1995). In addition, with regard to the proficiency, the PVLT scores at the K2 band...
discriminated learners’ writing proficiency; on the contrary, there was no significant
difference in the PVLT scores between the high and low proficiency group at the K3 band.

Finally, in terms of the students’ perceptions of the PVLT, most of the respondents
thought that the test was intended to assess their vocabulary knowledge. Interestingly, the
test format where differing number of initial letters are given for the intended words was
the factor that contributed to what makes the test difficult and easy at the same time. It
seems that if the students are familiar with the target words, that aspect was helpful to fill
in the blanks, but if they do not know the target words and have other alternative words in
mind, this type of test format becomes an obstacle.

All the findings in this study are significant in that more accurate interpretation of the
PVLT and the LFP is provided along with the students’ perceptions of the PVLT. Previous
study reported that the PVLT shows approximate number of words that can be readily
produced, indicates the stable pattern of vocabulary obtained from the LFP, and
discriminates learners’ proficiency (Laufer & Nation, 1995, 1998). However, this study
found that the relationship between the PVLT and other variables (students’ use of
productive vocabulary, their compositions, and their writing proficiency) is largely
influenced by word frequency level, and that there was no correlation noted between the
PVLT and the LFP. Therefore, the results of this study provided some implications
applicable to instructional practices. First of all, teachers should be careful when adopting
the PVLT since results of the PVLT should be interpreted differently according to word
frequency levels. Students’ ability to write the missing letters in the blanks can neither be
interpreted as being able to produce those words in other contexts nor can their failure to
answer correct words be considered as not being able to produce them in their writing.
Rather, more attention and weight should be given to how the learners choose proper
words, how they push those in mind into their writing, or why they cannot generate those
in novel sentences despite an awareness of the words. This can be explained in conjunction
with EFL contexts where vocabulary learning is mainly focused on rote memorization
without having much chance to produce vocabulary in meaningful contexts. This habit of
learning was also revealed by the results from the students’ reflection paper, which showed
that they wanted multiple choice answers when taking the PVLT. Therefore, in order to
promote learners’ productive vocabulary knowledge, learners need to be guided by
teachers to experience integrative tasks where learners are encouraged to generate proper
vocabulary according to context for various communicative purposes. Teachers should
also keep in mind that if this type of test format such as PVLT continues to be employed as
a vocabulary measure or vocabulary activities in classroom, this particular format may be
brought to learners’ attention and promote the habit of learning vocabulary by focusing on
the discrete vocabulary knowledge. This washback effect will impede their productive
skills in writing.
Despite some limitations of the PVLT, however, judging from all the findings described above, it is suggested that the scores on the PVLT at the K2 band can be applied to access information about learners’ vocabulary knowledge since significant correlations were noted between the scores on the PVLT at the K2 band and other variables: the scores on the PVLT at the K3 band ($r = .820, p = .000$), the students’ writing ($r = .598, p = .002$), and their writing proficiency ($r = .822, p = .002$).

Although this present study contributed to the empirical experiment investigating whether and to what extent the target words in the PVLT were actually produced in other contexts and whether there were any correlations between the PVLT and the compositions, the findings should not be taken as conclusive evidence. Therefore, a few limitations of the study should be acknowledged for future research. First, this study was conducted with a small number of subjects and one writing composition. Four weeks in the summer semester when this research was conducted was not enough time to ask the students to produce more than one composition. Clearly, more than one composition with various topics is strongly suggested for future study to examine how writing topics affect learners’ vocabulary profiles and writing proficiency in conjunction with the PVLT. In other words, the results as to the relations between the PVLT and the LFPs of the students’ compositions might be affected by the different writing topics. Also the writing topics will affect the writing quality according to which the students’ writing proficiency is evaluated. Second, although the above-3000 word level was excluded in this study in light of a statistical floor effect detected by the previous study (Waring, 1997), results might have been different if the students had been asked to take the PVLT at different word frequency level. Therefore, further research needs to examine whether the PVLT above the 3000 word level would report different interpretation from this study.

REFERENCES


Does the PVLT Provide an Accurate Measure of Productive Vocabulary Knowledge?


**APPENDIX A**

Productive vocabulary levels test (Version A)

The 2000-word level

1. I’m glad we had this opp_________ to talk.
2. There are a doz_________ eggs in the basket.
3. Every working person must pay income t_________.
4. The pirates buried the trea_________ on a desert island.
5. Her beauty and cha_________ had a powerful effect on men.
6. La_________ of rain led to a shortage of water in the city.
7. He takes cr_________ and sugar in his coffee.
8. The rich man died and left all his we_________ to his son.
9. Pup_________ must hand in their papers by the end of the week.
10. This sweater is too tight. It needs to be stret_________.
11. Ann intro_________ her boyfriend to her mother.
12. Teenagers often adm_________ and worship pop singers.
13. If you blow up that balloon any more it will bur_________.
14. In order to be accepted into the university, he had to impr_________ his grades.
15. The telegram was deli_________ two hours after it had been sent.
16. The differences were so sl_________ that they went unnoticed.
17. The dress you’re wearing is lov_________.
18. He wasn’t very popu_________ when he was a teenager, but he has many friends now.
The 3000-world level

1. He has a successful car________ as a lawyer.
2. The thieves threw ac________ in his face and made him blind.
3. To improve the country’s economy, the government decided on economic ref________.
4. She wore a beautiful green go________ to the ball.
5. The government tried to protect the country’s industry by reducing the imp________ of cheap goods.
6. The children’s games were funny at first, but finally got on the parents’ ner________.
7. The lawyer gave some wise coun________ to his client.
8. Many people in England mow the la________ of their houses on Sunday morning.
9. The farmer sells the eggs that his he________ lays.
10. Sudden noises at night sea________ me a lot.
11. France was proc________ a republic in the 18th century.
12. Many people are inj________ in road accidents every year.
13. Suddenly he was thru________ into the dark room.
14. He perc________ a light at the end of the tunnel.
15. Children are not independent. They are att________ to their parents.
16. She showed off her sle________ figure in a long narrow dress.
17. She has been changing partners often because she cannot have a sta________ relationship with one person.
18. You must wear a bathing suit on a public beach. You’re not allowed to be na________.

Applicable levels: secondary education
Key words: Productive Vocabulary Levels Test, Lexical Frequency Profile, writing proficiency, productive vocabulary

Bokyoung Park
Toopyong-dong Guri-si Kyounggi-do, Korea
Email: njpbk@hotmail.com

Received in March, 2012
Reviewed in April, 2012
Revised version received in May, 2012