

Effects of an Achievable Productive Vocabulary Levels Test on Low English Proficiency Learners' Performance and Perceptions

So-Yeon Moon*
Seoul National University

Abstract

The Productive Vocabulary Levels Tests (PVLTs, Laufer & Nation, 1999) to assess second language learners' size of controlled productive vocabulary have long been used in vocabulary research and assessment. The existing frequency levels of PVLTs between 2000 and 10000 words, however, seem too difficult for low English proficiency (LEP) learners to obtain cut-off scores, usually 75-80%. In order to help and motivate the LEP learners to feel sense of achievement, Abdullah et al. (2013) designed and validated the 500-word level PVLT (henceforth PVLT 500). This paper empirically examines the effects of the low-leveled PVLT 500 upon Korean high school LEP learners' performance and perceptions, through both validating PVLT 500 in comparison with the 2000-word level PVLT (henceforth PVLT 2000) and analyzing test-takers' questionnaire responses. With the proper assessment the study reveals that low-leveled PVLT can be actively applied in Korean EFL high school classrooms. Actually, it is LEP learners, who have been neglected and decentralized in the classrooms, who need 'achievable' tests, not the tests they often fail.

Key words : Productive Vocabulary Levels Test, low English proficiency learners

* Corresponding author (soyeonmoon@snu.ac.kr)

I. Introduction

After years of neglecting, a growing interest in the importance of lexical knowledge and its systematic development has emerged in not only L1 but also L2 (Laufer, 1986; Read, 1988). Based on the idea that lexical knowledge is essential in language competence (Grabe, 1991; Frederiksen, 1982), vocabulary has been regarded as one of the key components of language ability (Bachman & Palmer, 1996; Carroll, 1968; Meara, 1996). Thanks to the ongoing development of the computerized database such as COBUILD, researchers have been prompted to analyze different kinds of lexical knowledge (Laufer et al., 2004).

As lexical knowledge is multidimensional and the relationship between such dimensions and their effects seem very complex and still unclear (Pignot-Shahov, 2012), 'knowing a word' is sometimes defined in terms of various inter-related sub knowledges such as grammatical or morphological knowledge or knowledge of word meaning, otherwise including progressive stages starting from superficial familiarities of words to accurate uses of words in free production (Faerch et al., 1984; Palmberg, 1987). Basically, however, vocabulary knowledge is divided into two categories, receptive and productive knowledge. Receptive vocabulary knowledge is the ability to understand and recognize a word, while productive knowledge is the one to use a word appropriately (Nation, 1990; Read, 2000; Schmitt, 2010; Webb, 2008).

The following are several findings about the properties of receptive and productive vocabulary knowledge:

- Learners have a more receptive vocabulary than productive one (Laufer, 1998; Laufer & Paribakht, 1998; Laufer et al., 2004; Webb, 2008).
- Learners develop receptive vocabulary first prior to productive vocabulary (Pignot-Shahov, 2012).
- Recalling and producing vocabulary is more difficult than recognizing and understanding it (Nation, 1990). Thus, many words are acquired passively first (Laufer & Goldstein, 2004).

- A large number of receptive vocabulary must be obtained in order to develop productive vocabulary (Milton, 2007).
- The gap between the passive and the controlled active vocabulary increases with progress in learning (Laufer, 1998).¹⁾

The last proposal seems inevitable in that learners at the first frequency level, or in other words, the first 1000 words, are prompted to produce their receptive vocabulary, which is most frequently used in communication. In contrast, as the frequency levels increase, learners are not likely to produce the less frequent receptive vocabulary, unless the output is pushed. Thus, diagnosing learners' exact states of receptive and productive vocabulary through appropriate vocabulary tests is a prerequisite. LEP learners need the most attention, for they are likely to give up producing vocabulary more easily, demotivated, according to Hermann's (1980) resultative hypothesis.²⁾

To diagnose whether learners have achieved a certain mastery level of vocabulary and, if not, what can be pedagogically done for them, vocabulary levels tests play important roles. The problem is that almost all Vocabulary Levels Tests (VLTs) and PVLTs have drawn their focus on learners above intermediate levels, not in low-proficiency levels, which make diagnosing and treating LEP learners' insufficient productive vocabulary knowledge difficult. Fortunately, however, Abdullah et al. (2013) recently designed and validated the PVLTs at below 2000-word level, that is 500-word level, functioning as a self-efficiency enhancer.

The present study is in line with Abdullah et al.'s study. The purpose of this study is to apply PVLIT 500 to Korean high school learners and investigate its cognitive effects especially on LEP learners. In addition, the present study also aims at examining LEP learners' perceptions towards the PVLIT suitable for them.

1) The passive vocabulary is in line with the receptive vocabulary, and the controlled active vocabulary are the words learners produce when prompted to do so. This is different from free active vocabulary without any specific focus on particular words (Laufer, 1998).

2) According to Hermann (1980), learners' motivation and effort put forward for a task is triggered by positive results and progress.

II. Productive Vocabulary Levels Tests for Low English Proficiency Learners

A. Productive Vocabulary Levels Tests

Lexical knowledge tests can focus on one sub-knowledge such as comprehension of meaning (Meara & Buxton, 1987; Nation, 1983), production of meaning (Laufer & Nation, 1999), vocabulary use (Arnaud, 1992; Laufer & Nation, 1995) or word associations (Read, 1993), or several sub-knowledges (Schmitt, 1999). The focus can also be drawn to test-takers' progress along the continuum of the lexical knowledge (Wesche & Paribakht, 1996). Although Laufer and Goldstein (2004) pointed out oversimplified vocabulary size tests, suggesting strength test models, in which hierarchical difficulty levels of lexical knowledge were drawn, vocabulary size tests still gave a representative picture of the learners' overall state in vocabulary (Read, 2000).¹⁾ Additionally, size tests provide instructors in language teaching programs with efficient placement, diagnosis, and admission information (Laufer et al., 2004).

The first vocabulary size test is known as VLT (Nation, 1983, 1990), for which there have been several validation studies (Beglar & Hunt, 1999; Read, 1988; Schmitt, Schmitt & Clapham, 2001; Xing & Fulcher, 2007). VLTs require learners to link 3 of 6 given words to their meanings. Developed along with five frequency levels between 2000 and 10000 words, VLTs function as placement or diagnostic tests. *The checklist* (Read, 1988) as an alternative format of VLT and *Eurocentres Vocabulary Size Test* (Meara & Buxton, 1987) also measure learners' receptive vocabulary size.

Meanwhile, the necessity of reliable and valid productive vocabulary size tests has been voiced, for receptive vocabulary tests alone were insufficient to measure learners' productive

1) Based on two criteria, recall and recognition as well as form and meaning, vocabulary strength models provide hierarchical difficulty levels of lexical knowledge from passive recognition to active recognition to passive recall to active recall (Laufer et al., 2004).

vocabulary (Hughes, 1989). Through the washback of productive vocabulary measurement, learners can recognize their under- or over-uses of certain words, being motivated to continue progress, while teachers get the information of learners' vocabulary levels as well as the feedback about their teaching materials and instruction styles. As the examples of productive vocabulary size tests, both *The Lexical Frequency Profile* (Laufer & Nation, 1995) and *P_Lex* (Meara & Bell, 2001) elicit as many words as possible, not restricting test-takers' production of words. Other than these, PVLTs (Laufer & Nation, 1999) can be used to examine test-takers' controlled productive vocabulary.

PVLTs investigate learners' ability to draw words when required to do so in constrained but meaningful contexts of fill-in tasks, in which the first part of the target word is shown to prevent similar answers. As modeled on the VLTs, PVLTs also sample their items at each 2000, 3000, 5000, University Word Level, and 10000 word level. Despite several limitations of giving little freedom in learners' responses (Meara & Bell, 2001; Schmitt, 2010), practical benefits of PVLTs including easy administering, marking, and interpreting have made it easier for instructors and researchers to adopt the tests. Actually, Laufer (1998) traced learners' controlled active vocabulary development with the PVLt in his study of examining three types of passive, controlled active and free active vocabulary, and their relationships with learners' progress in learning.

B. PVLts for Low English Proficiency Learners

Nation and Chung (2009) once compiled three kinds of newspaper corpora and classified each text coverage by 1000 frequency levels. As shown in Table 1 the first row of 1000 words covers the very high percentage; 76.59, 75.72 and 75.11% in *Independent*, *New York Times* and *Dominion Post*, respectively. These high frequent words need intensive attention as "it takes at least a year, and usually much longer, to increase vocabulary size by a thousand words" (544). With more general corpus, *Brown* corpus containing various text types and registers, Francis

and Kučera (1982) also revealed the remarkable gap of text coverage between the first 1000 frequency level and the others. Accordingly, teaching attention seems reasonable to be drawn towards the high frequent words at the first and the second 1000 levels (Laufer & Nation, 1999). Then, learners' scopes of vocabulary can be expanded to the third, and fourth 1000 levels and onwards, through utilizing strategies, such as guessing words in context, memorizing or learning parts of words.

Table 1. Percentage Coverage by Five 1000-word Family Lists of Three Newspaper Corpora (adapted from Nation & Chung, 2009 : 541)

Vocabulary level	Independent (UK)	New York Times (USA)	Dominion Post (NZ)
1 st 1000	76.59	75.72	75.11
2 nd 1000	8.68	8.38	8.96
3 rd 1000	2.86	2.66	3.23
4 th 1000	2.21	1.24	2.47
5 th 1000	1.27	0.86	1.38

Yet, the actual EFL classroom situation looks quite different. Though Nation (2010) proposed EFL learners' usual uses of the words in speaking and writing, Japanese university students in Al-Mirtadha's (2014) study showed difficulty in producing the words in speaking. As not only Japanese but also Korean classrooms focus on reading and receptive vocabulary learning, learners tend to feel uncomfortable with producing language in communicative activities. Therefore, especially in EFL situation, scaffolding learners to acquire the reasonable size of the productive vocabulary is critical, for which appropriate PVLts suitable for each learner have to be accompanied.

For LEP learners in Korean high schools, PVLt 2000, the lowest level till 2013, seems too difficult. Overwhelmed by unsolvable word tests, those LEP learners are likely to track 'vicious cycle' of failure, lower motivation, less effort and hence, worse achievement or giving up (Abdullah et al., 2013). In this regard, Abdullah et al. developed PVLt 500 to provide LEP learners with easier and more achievable vocabulary tests, not the ones they used to fail. The present study aims at proving both reliability and validity of PVLt 500 in Korean high school

classrooms. Additionally, LEP learners' perceptions towards 'solvable' PVL 500 are identified. For the purposes of the study, following research questions are investigated:

1. What are the differences of LEP learners' performance when taking the PVL 500 and 2000?
2. How do LEP learners perceive achievable PVL 500 compared with the PVL 2000?

III. Methods

A. Participants

2nd graders of G high school in Gangnam-gu, Seoul participated in the present study. Excluding those who were unable to read an alphabet, which was the minimum requirement for taking the PVLs, a total of 279 students took part in the study as shown in Table 2. They were grouped into 5 levels based on their mid-term English scores. Those whose levels were 'very limited' and 'limited' belonged to LEP group, to whom the preliminary focus was given, whereas those in 'satisfactory' and 'proficient' HEP group.

Table 2. Group Division in the Present Study

group	G1	G2	G3	G4	G5
mid-term score range	0-12.5	12.5-21	21-49	49-83.5	83.5-100
level	very limited	limited	modest	satisfactory	proficient
	LEP learners		HEP learners		
n (N=279)	55	56	56	56	56

B. Controlled Productive Vocabulary Levels Tests

While learners are sometimes reluctant to use certain words,

worrying that they may reflect their imperfect word knowledge, PVLs (Laufer & Nation, 1999) trigger learners to produce words by giving meaningful contexts. One example test item is provided;

The book covers a series of isolated epis_____ from history (37).

Showing the first part of the target word keeps test-takers from filling in other seemingly appropriate words which may come from different frequency levels. Also, the length of each underlined space does not give any cues of how many number of the letters is needed to complete the target word.

Abdullah et al. used 38 items in their 2013 study. 20 items of PVL 500 were newly developed, while 18 of PVL 2000 were adopted from the first version of Laufer and Nation's (1999). The PVL in the present study was composed of 40 items in total (See Appendix). The first 20 items were originated from the PVL 500 in Abdullah et al.'s and the second 20 items came from Versions 1 and 2 of Laufer and Nation's PVL 2000. Test-taking time was limited to 30 minutes as in Abdullah et al.'s study. In scoring, however, half points were given to the answers with grammatical or spelling errors unlike Abdullah et al.'s, in which only correct answers including accurate spelling and grammar were given full scores.

C. Procedures

The overall design of this study is illustrated in Table 3. After filling in the agreement form, a total of 279 participants, including 111 LEP learners, did practice tests to get familiar with test processes. Then, actual tests were conducted within 30 minutes. Two different types, one initiating with PVL 500 and the other with PVL 2000, were randomly given to the participants to remove any possible outcomes from the order of two PVLs. Immediately after the PVLs, questionnaires were given to participants' perceptions towards different versions of PVLs, which took approximately 15 minutes.¹⁾ Additionally, for

the purpose of supporting questionnaire results, some seemingly significant test-takers had individual post-interview sections with the examiner.

Table 3. The Design of the Study by Test Types

A type (n=140)	B type (n=139)	time (N=279)
Practice Test		10 mins
20 items from PVL 500 ↓	20 items from PVL 2000 ↓	30 mins
20 items from PVL 2000	20 items from PVL 500	
Questionnaire		15 mins
Individual Post-interview (for those who were selected)		5 mins per one interviewee

D. Data Analyses

For the statistical data analyses, SPSS 18.0 was employed. To confirm the heterogeneity t-tests were conducted both among five groups and between LEP and HEP groups. All *p* values were less than .05, which indicated each group was different from each other. Reliabilities of the two PVLs were measured by Cronbach alpha, as in Laufer and Nation's (1999) study, in which the reliabilities of PVLs fell between .51 to .80.

Descriptive statistics were used to compare the mean scores and standard derivations of two PVLs by proficiency groups. Meanwhile, not assuming a normal distribution of the residuals, a non-parametric method for one-way analysis of variance (ANOVA), the Kruskal-Wallis test, was used to examine the difference between multiple groups. Also, the Wilcoxon signed-ranks test was applied instead of the paired-samples t-tests to compare the difficulty of PVLs 500 and 2000 within groups. Additionally, the item facility (IF) value was examined with the following standards; Those items with IF 0 ~ .2 belonged to 'very difficult' items, .2 ~ .4 'difficult', .4 ~ .6 'moderate', .6 ~ .8 'easy', and .8 ~ 1 'very easy'.

1) The questionnaire was translated into Korean in the actual study.

Table 4. Items in the Questionnaire

Item	Content
1	Which proficiency levels do I believe I belong to?
2	What is the English vocabulary test for me?
3	Could I feel the difference in difficulty levels of Parts A and B?
4	Was Part A easier than Part B?
5	Writing the number of items for each Part : Items I'm sure of Items which were unclear Items I can solve if I study, but I do not know at present Items I totally gave up on solving
6	Writing each Part when I agree : I tried to solve items as far as I could. I almost gave up solving items, for words and structures were too difficult. Items that are difficult like these keep me from learning the English language.
7	Writing free responses about the overall English language learning including English vocabulary tests

For examining LEP learners' perceptions towards the different PVLs questionnaire results were analyzed. Among seven items in total, the question type of the initial four items was a five-point Likert-type scale response ranging from 'strongly disagree' to 'strongly agree'. Two items were fill-in the gap responses, while the last item was an open-ended question. Table 4 summarizes the details of each item. For the quantitative results of the questionnaire ANOVA and Mann-Whitney U tests were accompanied. Open-ended results and responses from individual post interview were analyzed qualitatively.

IV. Results and Discussion

A. LEP Learners' Performance on PVL 500

Initially, the Cronbach alpha reliability coefficients of PVLs 500 and 2000 were calculated both for the total and LEP learners. As shown in Table 5, the reliability of PVL 500 was very high, $\alpha=.93$ for all test-takers, and high, $\alpha=.88$ for LEP learners. In addition to the findings that both PVLs 500 and 2000 are highly reliable, correlations were computed to examine

how the two PVLTs relate to each other. The analysis yielded the significant correlation between PVLTs 500 and 2000 ($r=.896$, $p<.001$).

Table 5. Reliabilities of PVLTs for All and LEP Learners

	Level	Reliability
All learners (N=279)	PVLT 500 (20 items)	.93
	PVLT 2000 (20 items)	.96
LEP learners (N=111)	PVLT 500 (20 items)	.88
	PVLT 2000 (20 items)	.90

Based on the validation procedure of Abdullah et al.'s (2013) the present study investigated whether PVLT 500 was able to decipher among learners in different proficiency levels as the PVLT 2000 did (Laufer & Nation, 1999). Learners with higher English proficiencies achieved higher gains in PVLT 500 than those with lower proficiencies as shown in Table 6. Also, while the mean scores of PVLT 500 increased steadily, those of PVLT 2000 rather showed sudden incremental stages between G1/2 and G3/4. A Kruskal-Wallis Test, a non-parametric test for an ANOVA, showed that significant differences existed in both LEP and HEP groups for PVLTs 500 and 2000 with the $p<.001$. A post-hoc Tukey analyses revealed significant differences in both groups with $p<.001$.

Table 6. Mean Scores of Five Proficiency Groups for PVLTs

	G1		G2		G3		G4		G5	
	M	SD	M	SD	M	SD	M	SD	M	SD
PVLT 500	4.02	3.04	8.11	5.07	12.09	7.79	16.35	3.28	18.40	1.69
PVLT 2000	0.96	1.47	4.39	4.75	4.87	4.79	14.66	3.95	17.22	2.38

In addition, in all levels ranging from G1 to G5, test-takers' mean scores in PVLT 500 were higher than those in PVLT 2000. The gap scores between PVLTs 500 and 2000 of LEP learners in G1 (3.06) and G2 (3.72) were higher than those of HEP learners in G4 (1.69) and G5 (1.18), which proved that LEP learners had much more difficulty in taking PVLT 2000. The Wilcoxon signed ranks test showed the gap scores were statistically significant

with $p<.001$ at all proficiency levels. Therefore, PVLT 500 including easier items with higher frequencies was proved to be an easier test than PVLT 2000.

Table 7. The Number of Test-takers Achieving Mastery Level

	PVLT 500 at 50% (10/20) and above	%	PVLT 2000 at 50% (10/20) and above	%	PVLT 2000 at 70% (14/20) and above	%
G1	2	3.6	0	0	0	0
G2	20	35.7	10	17.9	3	4.6
G3	38	67.9	20	35.7	6	10.7
G4	53	94.6	51	91.1	33	58.9
G5	56	100	56	100	50	89.3

Using cut-off scores of 10 (50%) or 14 (70%) out of 20 items, LEP learners easily failed the PVLT 2000 as shown in Table 7. Instead, LEP learners could have more chance to pass PVLT 500 than PVLT 2000 when the cut-off score was 50% ($39.3>17.9$). Regarding that PVLT 500 shows reasonably high correlation with PVLT 2000, if the aim of the test is not to distinguish HEP learners, test designers or instructors had better encourage LEP learners to take low-leveled PVLT first. Starting from the basic level of PVLTs, LEP learners can be scaffolded to take higher levels of tests in stages after completing vocabulary tests.

In Korean high schools English classes are often differentiated by learners' levels, and the number of learners in lower-leveled classes is usually limited to around 10. Table 8, in which each low-leveled group was divided into 5 sub-groups, reveals the possible usefulness of the present PVLT 500 in LEP learners' vocabulary instruction. The gap scores between two PVLTs were steadily high in the sub-groups. Especially, learners in G1-1,3,4 obtained nearly a zero score in PVLT 2000. No matter how hard LEP learners tried their best to solve items in both PVLTs, PVLT 2000 seemed too difficult for them. Therefore, in terms of lower-leveled classes, it seems much more useful to apply the PVLT 500 test.

Table 8. Mean Scores of Sub-LEP Groups

sub group	G1-1 (n=11)	G1-2 (n=11)	G1-3 (n=11)	G1-4 (n=11)	G1-5 (n=11)	total
PVLT 500	3.9	4.5	3.5	3.2	4.9	4.02
PVLT 2000	0.9	1.3	0.7	0.6	1.3	0.96
sub group	G2-1 (n=12)	G2-2 (n=11)	G2-3 (n=11)	G2-4 (n=11)	G2-5 (n=11)	total
PVLT 500	7.1	6.3	9.4	8.6	9.2	8.11
PVLT 2000	3.4	3.1	4.6	4.8	6.1	4.39

IF indices in Table 9 indicate that one 'very difficult' item with IF of 0 ~ .2 occurred only in PVLT 2000, while PVLT 500 alone included three 'very easy' items with IF .8 ~ 1. 'Difficult' items with IF .2 ~ .4 showed up more in PVLT 2000 (n=5) than PVLT 500 (n=2), whereas 'easy' items with IF .6 ~ .8 appeared more in PVLT 500 (n=6) than PVLT 2000 (n=3). PVLT 500, accordingly, was proved to have easier items than PVLT 2000.

Table 9. Item Facility Indices in PVLTs 500 and 2000

PVLT 500			PVLT 2000		
Item no	Words	IF	Item no	Words	IF
2	important	0.85	22	dozen	0.68
20	morning	0.83	31	introduced	0.66
9	build	0.82	27	cream	0.61
15	power	0.79	21	opportunity	0.57
10	bank	0.7	34	improve	0.55
19	spend	0.7	24	treasure	0.52
3	near	0.68	28	wealth	0.51
11	children	0.62	23	tax	0.49
4	change	0.61	32	admire	0.48
6	expect	0.59	25	charm	0.46
17	enough	0.58	36	roots	0.46
16	effect	0.56	40	brave	0.44
14	start	0.51	30	stretched	0.43
12	good	0.49	35	delivered	0.42
8	evidence	0.47	26	lack	0.37
7	father	0.46	38	wandered	0.35
18	subject	0.45	37	connects	0.31
1	market	0.44	33	burst	0.31
13	away	0.36	29	pupils	0.27
5	carried	0.34	39	limited	0.12

To summarize, the present PVLT 500 establishes similar validity as in Abdullah et al.'s (2013). Then, LEP learners don't necessarily have to take difficult tests like the HEP learners do, which may only result in LEP learners' giving up earlier. Rather, it is likely to be better to assist LEP learners in taking the easier test prior to the more difficult one.

B. LEP Learners' Perceptions Towards PVLT 500

Learners who obtained a zero on both PVLTs 500 and 2000 revealed extreme feelings of difficulty in the English language learning in the individual interview section as follows:

- A: *For me, English is too difficult.*
- B: *I hate English. I don't know how to read even a word*
- C: *Too difficult! There was nothing I could solve*
- D: *All words were unknown, so I gave up taking the test.*

For the learners above, the PVLT 500 itself was still too difficult to solve. Among the learners who gained a zero on either the PVLT 500 or 2000, most (88%) belonged to LEP learners and received a zero on the PVLT 2000, not on the PVLT 500. LEP learners were likely to solve items in the PVLT 500 trying their best, but usually gave up solving the more difficult ones in PVLT 2000. If they are mostly exposed to more difficult tests like the PVLT 2000, those LEP learners may be easily demotivated by frequent zero scores.

Among seven questions in the questionnaire examining learners' perceptions towards PVLT 500, initially two were general. The first question asked which proficiency levels learners themselves believed they belonged to. ANOVA revealed group differences ($F=27.926, p<.05$) and a post-hoc Scheffe test indicated mean differences between G1/3, 1/4, 1/5, 2/4, 2/5 and 3/5 were statistically significant. Surprisingly, half of the learners regarded their English proficiency levels as 'under-moderate', 'limited' or 'very limited.' Among them, LEP learners remarkably showed '(very) low' confidence. According to the second question, it was LEP learners who were mostly demotivated to take vocabulary tests, reluctant to be given bad scores again. The Mann-Whitney test with a Bonferroni adjustment showed significant differences existed in G1/3, 1/4, 1/5, 2/4, 2/5, 3/4 and 3/5. Learners in G1 seemed to just give up taking tests which were similar to the ones they usually got bad scores on, following 'vicious cycle', while HEP learners in 'virtuous cycle'

were encouraged to take tests to get better scores.

The third and fourth questions were about whether learners themselves could feel the difference in difficulty levels of the two PVLTs. Rather than two extreme groups, G1 (M=2.9) and G5 (M=2.9), those in the moderate group, G3 (M=3.2), felt small differences between two PVLTs. Among those who indicated the gaps, most learners thought PVLt 500 was (much) easier than PVLt 2000 (G1:78, G2:65, G3:83, G4:56, G5:74%).

Table 10 summarizes the results of the fifth question, summing up the total counts from learners' responses. To begin with, LEP learners in G1 were strongly assured of items in PVLt 500 than those in 2000. Though this phenomenon also happened in other groups, the gap ratio was far higher in G1 than any other group. Also, the number of items learners gave up was bigger in PVLt 2000 than in 500. Indeed, not HEP but LEP learners seemed to be significantly affected by the difficulty levels of PVLts. Thus, if relatively similar reliability and validity are detected in both PVLts, applying easier one first to classrooms, in which LEP learners have been at many times disregarded, would be much more pedagogically beneficial.

Table 10. Total Counts from Learners' Responses of Item Numbers

Item types		G1	G2	G3	G4	G5
PVLt 500	Items I'm sure of	163	269	520	737	869
	Items which are unclear	128	111	136	89	63
	Items I can solve if I study, but I do not know at present	243	256	177	99	45
	Items I totally gave up on solving	242	210	114	31	25
	Items I'm sure of	73	160	308	659	775
PVLt 2000	Items which are unclear	137	95	183	124	123
	Items I can solve if I study, but I do not know at present	253	244	224	121	46
	Items I totally gave up on solving	308	314	209	43	35

The sixth question explored learners' perceptions towards each PVLt. As noted in the first response in Table 11, LEP learners tried to solve items in PVLt 500 more than in 2000, while HEP learners did their best both in PVLts 500 and 2000. One of reasons was drawn in the second response. That is, some

LEP learners had difficulty in understanding complicated words or structures in PVLt 2000. Those LEP learners got demotivated even in the English language learning as shown in the third response.

Table 11. Learners' Perceptions Towards Each PVLt

	1) I tried to solve items as far as I could.					2) I almost gave up solving items, for words and structures were too difficult.					3) Items that are difficult like these keep me from learning the English language.				
	G1	G2	G3	G4	G5	G1	G2	G3	G4	G5	G1	G2	G3	G4	G5
PVLt 500	5	7	4	0	1	0	0	0	0	0	1	1	0	0	0
PVLt 2000	1	2	4	2	4	7	5	3	0	0	9	7	10	0	0
Both PVLts	21	32	39	44	44	5	7	3	2	3	15	12	3	2	2
Both not	15	8	2	6	6	28	37	40	46	46	16	28	34	46	48
no response	13	7	7	4	1	15	7	10	8	7	14	8	9	8	6

Last but not least, an open-ended final question regarding the overall English language learning was given to learners. As proficiency levels grew, more positive but less negative comments showed up. LEP learners produced the word, '(too) difficult', a lot, but were far from using words like 'fun' or 'satisfied', which appeared in G3 and G5, respectively. With specific questions regarding learning strategies to improve English proficiency levels, HEP learners showed interest in other functions of English language, such as speaking. Indeed, HEP learners who have succeeded in difficult PVLts seem to be triggered to give more effort and hence, greater achievement in learning the English language.

V. Conclusion

PVLts are based on the assumption that one's vocabulary consists of frequency levels (Laufer, 1998). Yet, none of the PVLts at current five levels fully focus on the inside of the vocabulary below 2000 word level. Regarding that the foremost 1000 words work as basic words, with which learners can guess the meaning of other words with higher frequencies, test designers or instructors are encouraged to devise and revise

low-leveled PVLTs continuously. Particularly in Korean high school classrooms, where learners' vocabulary breadth of knowledge varies, more achievable PVLTs seem to be common.

The present study is the first empirical attempt to apply PVL 500 (Abdullah et al., 2013) in Korean high school classrooms, in which most LEP learners, demotivated to take vocabulary tests because of frequent failures, have been decentralized. The study investigates LEP learners' performance on the PVL 500. Firstly, the PVL 500 functioned as a valid discriminator showing that HEP learners' PVL 500 scores were higher than LEP learners. Secondly, PVL 500 was proved to include much easier items than PVL 2000 so that LEP learners could get better scores on the PVL 500 than on the PVL 2000. Also, according to learners' perceptions, LEP learners perceived the solvable PVL 500 more positively than the PVL 2000 whose items were too difficult to understand.

The present study is not without limitations, however. One limitation was incurred by the fact that learners took the PVLs only once. Repeated but slightly different levels of low-leveled PVLs can be conducted systematically in future studies. Also, longitudinal studies following individual learners' productive vocabulary development are required to better understand the roles of low-leveled PVLs. Designing appropriate tasks for developing productive vocabulary at certain frequency levels is also to be considered. For instance, combining extensive reading for incidental vocabulary learning and productive writing tasks can prevent lexical attrition and connect receptive and productive vocabulary learning (Yamamoto, 2011).

Despite the limitations, pedagogical implications are drawn. That is, recent emphasis on productive skills such as speaking and writing in Korean high school English classrooms has revealed the necessity of productive vocabulary. In classrooms instructors are recommended to teach and assess learners' productive vocabulary with appropriate activities and tests. This positive correlation between teaching and assessment can be activated effectively under teaching in levels. When those in lower-leveled classes are exposed to the PVL 500 at first, the

LEP learners are more likely to have positive attitudes and interest towards productive vocabulary. As the first 1000 words is most frequently used in everyday communication, LEP learners can be scaffolded as a positive first step. Further sets of PVLs beyond the 500-word level, *i.e.*, 600, 700, 800, 900, etc, if used, would guide LEP learners' productive vocabulary skills. One caution is that a series of low-leveled PVLs including PVL 500 do not have to be included on a regular mid- or final-term test. That will be too burdensome to learners and may trigger the problem of fairness among different leveled learners. Rather, instructors can diagnose and assist LEP learners' progress in producing productive vocabulary with the PVLs. By doing so, LEP learners can also take part in English language learning more actively with positive perceptions as HEP learners do.

References

- Abdullah, K. I., Puteh, F., Azizan, A. R., Hamdan, N. N., & Saude, S. (2013). Validation of a controlled productive vocabulary levels test below the 2000-word level. *System*, *41*(2), 352-364.
- Arnaud, P. (1992). Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate-component tests. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 133-145). London: Macmillan.
- Al-Mirtadha, M. (2014). The relationship between productive vocabulary tasks and accuracy, fluency, and motivation of Japanese EFL learners. *Vocabulary Education & Research Bulletin*, *3*(1), 2-4.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing*, *16*(2), 131-162.
- Carroll, J. B. (1968). The psychology of language testing. In A. Davies (Ed.), *Language testing symposium: A psycholinguistic approach* (pp. 46-69). London: Oxford University Press.
- Faerch, C., Haastrup, K., & Philipson, R. (1984). *Learner language and language learning*. Clevedon: Multilingual Matters.
- Francis, W. M., & Kučera, H. (1982). *Frequency analysis of English usage*. Boston, MA: Houghton Mifflin.
- Frederiksen, K. R. (1982). A componential theory of reading skills and their interactions. In Mislevy, R. J. (Ed.), *Advances in the psychology of human intelligence: Vol. 1*. (pp. 125-180). Hillsdale, NJ: Lawrence Erlbaum.
- Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, *25*(3), 375-406.
- Hermann, G. (1980). Attitudes and success in children's learning of English as a second language: The motivational vs. the resultative hypothesis. *English Language Teaching Journal*, *34*(4), 247-254.
- Hughes, A. (1989). *Testing for language teachers*. New York: Cambridge University Press.
- Laufer, B. (1986). Possible changes in attitude towards vocabulary acquisition research. *International Review of Applied Linguistics*, *24*(1), 69-75.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, *19*(2), 255-271.
- Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, *21*(2), 202-226.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399-436.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307-322.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33-51.
- Laufer, B., & Paribakht, T. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, *48*(3), 365-391.
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjæe & J. Williams (Eds.), *Performance & competence in second language acquisition*. Cambridge: Cambridge University Press.
- Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, *16*(3), 5-19.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*(2), 142-154.
- Milton, J. (2007). Lexical profiles, learning styles and the construct validity of lexical size tests. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 47-58). Cambridge: Cambridge

- University Press.
- Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1), 12-25.
- Nation, P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, P. (2010). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P., & Chung, T. (2009). Teaching and testing vocabulary. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 543-559). Malden, MA: Wiley-Blackwell.
- Palmberg, R. (1987). Patterns of vocabulary development in foreign language learners. *Studies in Second Language Acquisition*, 9(2), 201-219.
- Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers*, 4, 37-45.
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 19(2), 12-25.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-71.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge. *Language Testing*, 16(2), 189-216.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, 18(1), 55-88.
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79-95.
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40.

- Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of vocabulary levels tests. *System*, 35(2), 182-191.
- Yamamoto, Y. (2011). Bridging the gap between receptive and productive vocabulary size through extensive reading. *The Reading Matrix*, 11(3), 226-242.

Appendix

Controlled Productive Vocabulary Levels Test

- **Instructions** : Answer ALL the questions. You have 30 min to complete this test. Complete each sentence by filling in the blank with an appropriate word. The first few letters of the word are given.

<Example> He was riding a bicycle.

►Part A (1-20)

1. He asked again if she wanted to go to the mar_____ with him.
2. It is impor_____ that a father be a good example to the children.
3. The person who lives n_____ door is an international student.
4. Liam decided to cha_____ the way he lives his life after the death of a good friend.
5. They c_____ the dead body from the building to the car.
6. Oh, you're back! I did not ex_____ you to return so early.
7. Some young boys have problems speaking to their fat_____.
8. The police are looking for ev_____ to show that he took money from the company.
9. The government will bui_____ a lot of new houses.
10. I need to go to the b_____ to take out money.
11. Mothers should try to understand their ch_____.
12. Students should work together to get go_____ results.
13. The percentage of young girls running a_____ from home is high these days.
14. The machine stopped, so they had to s_____ all over again.
15. He has the pow_____ to make all decisions in the team.
16. Eating too much can have a bad e_____ on our health.
17. The company does not have en_____ money to develop the

business.

18. When I was at school, the s_____ that I was best at was English.
19. I sp_____ most of my money on books.
20. I eat at 7 every mor_____.

►Part B (21-40)

21. I'm glad we had this opp_____ to talk.
22. There are a doz_____ eggs in the basket.
23. Every working person must pay income t_____.
24. The pirates buried the trea_____ on a desert island.
25. Her beauty and ch_____ had a powerful effect on men.
26. La_____ of rain led to a shortage of water in the city.
27. He takes cr_____ and sugar in his coffee.
28. The rich man died and left all his we_____ to his son.
29. Pup_____ must hand in their papers by the end of the week.
30. This sweater is too tight. It needs to be stret_____.
31. Ann intro_____ her boyfriend to her mother.
32. Teenagers often adm_____ and worship pop singers.
33. If you blow up that balloon any more it will bu_____ .
34. In order to be accepted into the university, he had to impr_____ his grades.
35. The telegram was deli_____ 2 hours after it had been sent.
36. Plants receive water from the soil through their ro_____.
37. The railway con_____ London with its suburbs.
38. She wan_____ aimlessly in the street.
39. The organisers li_____ the number of participants to fifty.
40. You must have been very br_____ to participate in such a dangerous operation.