To know is not to use: The gap between students' productive vocabulary knowledge and their actual use in free writing

Bronson Hui
S.K.H. St. Simon's Lui Ming Choi Secondary School
Daniel Fung
University of Oxford

Abstract

One major goal of vocabulary learning is to empower learners to use selected words whenever they need. However, there seems to be a phenomenon that some learners possess the necessary productive word knowledge, such as spelling and grammatical features of the target words, but still find it difficult to use the words correctly. This piece of research reports the process of identifying the potential developmental gap between knowing a word and using it in writing. It involved 12 Form 6 students from a local secondary school using English as the medium of instruction. They completed the Productive Vocabulary Levels Test which measured their productive word knowledge, and the results were adjusted and compared with the lexical profiling of their four pieces of writing for investigating if the words used by the learners belong to: (1) high frequency K2 words (i.e., the second most frequent 1,000 words in English); (2) mid frequency K3 words (i.e., the third most frequent 1,000 words); or (3) low frequency K4-K5 words (i.e., the fourth and fifth most frequent 1,000 words). In order to find out whether these two measures revealed a difference between knowledge and usage, a correlation test was first performed. Results confirmed that knowledge and usage were two different concepts. A follow-up t-test comparing the expected and actual proportion of use showed a significant difference with
To know is not to use: The gap between students’ productive vocabulary knowledge and their actual use in free writing

high frequency K2 words, but not K3 and K4-K5 words. We argued that there was a gap between knowing and using a word as well as students’ strategy to opt for lower frequency words. Pedagogical implications are discussed in relation to how teachers can assist learners in bridging this gap.

Introduction

Vocabulary is an essential part of language learning. In the case of receptive skills (i.e., reading and listening), studies have demonstrated that a foreign language user needs to know a high percentage of the words within a piece of text in order to gain adequate comprehension (e.g., Hu & Nation, 2000; Laufer, 1998; van Zeeland & Schmitt, 2012). This percentage of words known by foreign language users out of the total number of words in the text is termed lexical coverage (Adolphs & Schmitt, 2003). For example, when the lexical coverage for a learner is 80%, it means that s/he knows 80% of the words in the text.

Hu and Nation (2000) used non-words (i.e., words that do not exist) to replace real words in fiction texts in order to mimic language learners’ encounter of unknown words. The researchers created four levels of lexical coverage (i.e., 80%, 90%, 95% and 100%). That is, in the case of the 80% lexical coverage, 20% of the real words were replaced by non-words (e.g., ‘rane’). In other words, it was expected that out of all the words of the text, 20% were unknown to the participants prior to the study. Participants were asked to read the text and complete a 14-item comprehension test in the format of multiple choice questions. Adequate comprehension was defined as being correct in at least 12 items (i.e., 86% of 14). Results showed that none could attain adequate comprehension at 80% lexical coverage; and only 4 out of 16 (25%) and 6 out of 17 (35%) passed the comprehension threshold at 90% and 95% coverage respectively. It was only when the coverage reached 100% that the majority (15 out of 17) showed adequate comprehension in the test. From these figures, the authors suggested that the lexical coverage threshold for adequate comprehension should be between 95-100%, and hence proposed the figure of 98%.

One of the key implications about this figure is that researchers can enumerate the number of words a learner needs to understand authentic texts and set a pedagogical goal for vocabulary learning. For example, van Zeeland and Schmitt (2012) suggested that a learner
needs to know 2,000 to 3,000 words based on a 95% coverage figure in order to understand spoken text. When the calculation was based on a 98% coverage figure, an understanding of 6,000 to 7,000 words was required.

For language teachers in Hong Kong, the question is whether or not this target is achievable as most of their students learn English as a foreign language. One problem is that the goal of language learning is not confined to comprehension. Helping them to use the language (i.e., to speak and write to express meaning) with proper vocabulary is also an important target. Vocabulary knowledge can be receptive (or passive) and productive (or active) in nature (Pignot-Shahov, 2012). This distinction is closely related to the question of what it means to know a word. To address the issue, Nation (2013) identified three aspects in this connection. The first aspect is knowledge of a word’s form such as its written form (i.e., spelling), spoken form (i.e., pronunciation), and its word parts (e.g., prefixes like pre-). The second aspect is the understanding of its meaning, including the concepts and/or referents that the word refers to, as well as its associations with other words, such as ‘good’ being an antonym of ‘bad’. The last one is the knowledge of its usage, namely its grammatical features (e.g., parts of speech), collocations and other constraints on use. The mastery of the first two aspects (i.e., form and meaning) would suffice for comprehension purpose. An understanding of the third aspect (i.e., use) is essential when one needs to produce language in a sensible manner. As a result, there appears to be a gap between receptive and productive knowledge. Indeed, using their Computer Adaptive Test of Size and Strength (CATSS), Laufer and Goldstein (2004) confirmed that receptive vocabulary is generally larger than productive vocabulary of an individual, and that the gap is likely to be a difficulty hierarchy.

The logical question in practical and pedagogical terms is how a teacher could bridge this receptive-productive gap. Lee and Muncie (2006) asked their participants to read a text on Titanic, the British passenger liner that sank in the Atlantic in 1917. The text contained 42 words or lexical phrases that were explicitly taught as novel vocabulary items. After that, the participants completed a writing task in which they had to describe a hypothetical experience on Titanic. The participants used only as few as 18.4-20.9% of those target vocabulary items in this piece of writing. This result is taken as evidence that reading and explicit vocabulary instruction alone cannot lead to satisfactory vocabulary use. In another attempt, the researchers asked students to
write on the same topic again. This time, the author provided students with scaffolds of a composition structure frame which helped the students focus on the target vocabulary. More than that, the target items were listed for the students, and the students were explicitly encouraged to use the target items in their writing. In this writing, the use of target vocabulary augmented to 67.5-68.7%. In the third attempt after two weeks, the scaffold was still the writing frame. Students were asked to recall the items and write them down before they started to write (instead of being able to refer to the items listed for them). This time, students produced 50.5-63% of the target items.

In a similar vein, Lee (2003) asked her participants to write about cruel sports after reading a passage on bull fighting and receiving explicit instruction on the target vocabulary items. The participants were able to produce 63.2% of the target vocabulary items in the passage that they had learnt. This percentage was compared with 13.2% of use before the vocabulary instruction. From these two studies, it may be concluded that post-reading vocabulary instruction, explicit scaffold, provision of the items as well as encouragement to use the target items could encourage students to use newly learnt vocabulary items in their production.

These studies have brought about two issues. First, two reasons could account for failure to use the target items in writing -- (1) the learners had not actually learnt the items; and (2) the learners have learnt the items but did not use them. The existence of these two possibilities makes the case obscure in the sense that teachers/researchers may find it difficult to pinpoint specifically what caused students’ failure to produce the target items.

Another issue that requires attention is that in both studies, the participants were asked to write on a topic that was very closely related to the vocabulary instruction that they had received. In the case of Lee and Muncie (2006), both the reading text and the writing task were about Titanic. Similarly, Lee (2003) had bull fighting as the topic of the reading, and cruel sports as the writing theme. On one hand, using a similar theme may offer a better context for the learners to try out with the newly learnt vocabulary items. On the other, it is not entirely clear if the learners could produce those target items in another novel context. This uncertainty is particularly important because, after all, one major goal of vocabulary learning is to empower students to use the items freely so that they are able to use the words to express themselves as they wish. In other words, these items should be readily available whenever they need them.
If the ability to produce the newly learnt vocabulary across various contexts is an important goal of teaching vocabulary for production's sake, it implies that there might be multiple levels of productive vocabulary. This notion may address a key problem which affects language teachers in Hong Kong who are used to assessing their students through different formats of vocabulary tests. Students may excel in vocabulary tests in the formats of dictation, fill-in-the-blanks and so on, showing all Nation's (2013) aspects of vocabulary knowledge of a word, such as spelling and grammatical features. However, they may still find it difficult to demonstrate a wide range of vocabulary use when they speak. It is, in fact, likely that many language teachers have been asked by students how newly learnt vocabulary could be used in everyday life.

To answer such a question, it is essential to first empirically identify such a gap between general productive knowledge and the ability to use these vocabulary items. It is only after such a gap is identified that teachers and researchers could look for ways to bridge this gap. Some may comment that this gap is commonly known, implying that teachers should notice students' problems of not being able to produce newly learnt lexical items. While we feel reluctant to agree totally, we acknowledge that some teachers may have a subjective impression from their experience, that knowing a word does not mean using a word. At the same time, we maintain that empirical identification is necessary. As far as we know, no study has specifically documented this final stage of vocabulary development. One of the many reasons why such a gap is merely taken for granted by many is that the measures of general productive knowledge and actual use are often two separate vocabulary measures that researchers employ as dependent variables. For example, looking at the vocabulary size of students in Chinese- and English-medium schools, Lo & Murphy (2010) used both measures without comparing them directly. The present study aims exactly at this comparison, and argues for the importance of bridging the gap between knowing and using vocabulary, once it is identified.

Methodology and Procedures

The participants in the present study included 12 Form 6 students in a local secondary school using English as the medium of instruction in most content subjects, such as Geography and History. These students were recruited from the same Form 6 class whose teacher the researchers of this study had personal contact with. All participants
were native speakers of Cantonese, and had learnt English as a foreign language for at least 11 years (6 years of primary and 5 years of secondary school at the time of participating in this study). These students were expected to sit for the Hong Kong Diploma of Secondary Education (HKDSE) Examination.

The Productive Vocabulary Levels Test (PVLT) (Laufer & Nation, 1999) was used to measure the general productive vocabulary knowledge of the participants. It requires test-takers to provide a vocabulary item in a given context presented in a sentence. Therefore, it is a test of controlled productive vocabulary knowledge in that test-takers are prompted to offer a word that fits into the context provided. This format of testing also echoes with the assessment practice of some local teachers who give fill-in-the-blanks exercise to students. Therefore, the findings of this test may be of particular relevance to local teachers in Hong Kong. There are two versions of the PVLT, but only Version 1 was used in this study lest doubts might be cast on any potential differences between the two versions. Two examples are given to illustrate the format of the test. They are used to elicit the target words 'hungry' and 'usual'.

1. They sat down to eat even though they were not hu__________.
2. This work is not up to your usu________standard.

Note that the initial letters of the target words are always given so that test-takers could not insert a word other than the target word that might also be semantically appropriate in the context.

The way in which this test could determine one’s productive vocabulary size is based on the frequency of occurrence of the target words in English as represented by the British National Corpus. This corpus contains a huge collection of authentic English texts, from which the frequency of occurrence is determined. The most frequently occurring 1,000 words in English are known as K1 words; K2 words are the next 1,000 words in the line of frequency; and K3 words the next 1,000 words and so on. In the examples above, ‘hungry’ and ‘usual’ are K2 words. For each level, there are 18 questions in the PVLT. Three levels were used in this study, K2, K3 and K4-K5. Laufer and Nation (1999) combined K4 and K5 into one level (i.e., K4-K5) for easier and more practical assessment of knowledge about these relatively low frequency words. With three levels used, the total score for the test in this study was 54. The items were pseudo-randomised using Microsoft Excel’s RAND function which assigns a random
number to each item (one could type ‘=RAND( )’ in any cell). The random numbers were sorted in ascending order to determine the order of the item presented in the actual test taken by the participants. The test was administered by the teacher of the class.

Within a month after the administration of the PVLT, four writing samples were collected from each student. The task requirement and the format resembled those in the questions in the HKDSE. The themes and genres are tabulated below (Table 1). Writings 1 and 2 were assigned as homework for students, and Writings 3 and 4 were parts of the students' formal assessment at school, which were timed. Participants could choose to write on one of the questions in Writings 3 and 4.

Table 1: The themes and genres of the writing tasks

<table>
<thead>
<tr>
<th>Writing</th>
<th>Theme</th>
<th>Genre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing 1</td>
<td>Walkathon</td>
<td>Expository in a poster</td>
</tr>
<tr>
<td>Writing 2</td>
<td>An artist and her exhibition</td>
<td>Feature article in a school newspaper</td>
</tr>
<tr>
<td>Writing 3</td>
<td>An outstanding student award</td>
<td>Nomination</td>
</tr>
<tr>
<td></td>
<td>Hip-pop dance performance</td>
<td>Promotional leaflet</td>
</tr>
<tr>
<td></td>
<td>Teenagers fighting for a girlfriend</td>
<td>Recount in a school newspaper article</td>
</tr>
<tr>
<td>Writing 4</td>
<td>Goal setting</td>
<td>Speech</td>
</tr>
<tr>
<td></td>
<td>Popular culture</td>
<td>Magazine article</td>
</tr>
<tr>
<td></td>
<td>Social harmony</td>
<td>Letter to the Editor</td>
</tr>
</tbody>
</table>

Results and Discussion

In this section, results obtained from the PVLT are first presented, followed by those of the free writing tasks. Recalling that the former is a measure of general productive knowledge and the latter one of actual productive use, this paper seeks to compare the findings from these two measures to shed light on the hypothesized discrepancies of the two levels of productive knowledge.
The PVLT consisted of 18 questions at each level of K2, K3, and K4-K5 words. One mark was given for every correct item. The marking of the test did not take grammatical accuracy into consideration. For example, when the word ‘attached’ was required, members in the same word family such as ‘attach’ and ‘attaches’ were also accepted. In other words, all inflected forms of the words belonging to the same word family were counted as correct. Although grammatical features of a word are part of word knowledge, the PVLT was treated purely as a test of vocabulary instead of a test of grammar in this study. Another rationale is related to the validity of comparison between the PVLT and free writing tasks in that grammaticality was also discounted in the corresponding measure of the students’ free production. As Table 2 indicates, the 12 students who took the PVLT showed a huge variation in their productive vocabulary knowledge. Students showed a decreasing score from K2 through to level K4-K5. This was not surprising because the higher the level, the lower the frequency of the words used in everyday language and hence the more unlikely that students would have got exposed to them.

Table 2: Results of PVLT

<table>
<thead>
<tr>
<th>N=12</th>
<th>K2 (max= 18)</th>
<th>K3 (max= 18)</th>
<th>K4-K5 (max= 18)</th>
<th>Total (max= 54)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>12.75 (2.56)</td>
<td>8.92 (3.48)</td>
<td>4.58 (3.09)</td>
<td>26.25 (8.31)</td>
</tr>
</tbody>
</table>

Students were also asked to write four essays of 200-500 words each. Their work was then typed onto the computer. Some of the content words which had been given as prompts were removed so that the remaining words were only those given by the students instead of being copied from the questions. Incidentally, quite a number of students started their essays with a large number of words copied from the questions. The processed text was input to the Vocabulary Profilers (available on www.lextutor.ca/vp) for their analysis of their lexical frequency profiles (LFP). Table 3 shows the total number of words in the students’ four writings, in their respective frequency levels of K2, K3, and K4-K5. Again, it was not surprising to find a decreasing number of words used by students as the frequency levels went up. In fact, it is normal for low frequency K4-K5 words to appear less frequently in a text given that it is their low frequency of existence in English that defines them as words at the K4-K5 level. Although this trend does echo with the results of the PVLT, we still need a valid comparison to identify the potential differences between the two measures.
To know is not to use: The gap between students’ productive vocabulary knowledge and their actual use in free writing

Table 3: Lexical frequency profiles of the free writing tasks

<table>
<thead>
<tr>
<th></th>
<th>K2 words</th>
<th>K3 words</th>
<th>K4-K5 words</th>
<th>Total number of words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>107.08 (29.60)</td>
<td>45.67 (22.45)</td>
<td>17.83 (11.27)</td>
<td>1500.17 (237.00)</td>
</tr>
</tbody>
</table>

This paper compares students’ PVLT scores and their free writing in order to identify whether there is a difference between general productive knowledge and the actual productive use of vocabulary. To this end, the PVLT score was computed according to the following formula so that it could be compared directly with the corresponding measures in free production.

\[
\text{PVLT score} \times \text{Adjustment factor} = \text{Expected production}
\]

An expected proportion of production was thereby obtained for each participant and for each lexical frequency level. For example, a student (S1) knew 16 K2 words. The knowledge of these 16 K2 words represented 50% of the total of 32 marks he scored in the PVLT (i.e., all K2-K5 words that he knew). Then, this 50% was multiplied by the adjustment factor for K2 words, which is 1.4 (the calculation of the adjustment factor is described below). As a result, the expected production rate of K2 words in his free writing would be 50% \(\times\) 1.4 = 70%. In other words, for every 100 K2-K5 words that he used, 70 K2 words were expected. This expected production could then be compared with his actual production in free writing as reported in the previous section.

The following explains how the adjustment factors were obtained for each level. First, the production of native speakers was considered to be a useful reference point. Corpus data (i.e., data from a huge collection of authentic texts written by native speakers) are of use here. According to Nation (1990: 17; 2001: 15), K2 words have a natural occurrence of 7.7%; K3 words 4.3%; K4-K5 words 4.6%. To put it in another way, there are on average 7.7 K2 words in every 100 words in any authentic texts. Note that these three percentages do not add up to 100% because there are also K1 words as well as K5+ words in a corpus. Rather, K2-K5 words account for only 16.6% of words in a corpus (see Figure 1 for a visual representation of these numbers). Out of this 16.6% of natural occurrence (i.e., K2-K5 words
only), K2 words have a proportion of 46.4% (7.7%/16.6%); K3 words 25.9% (4.3%/16.6%); and K4-K5 words 27.7% (4.6%/16.6%) (see Figure 2 for a visual representation). Given that a native speaker would have a full PVLT score of 100% (33.3% of the items from K2 words, 33.3% from K3 words, and 33.3% from K4-K5 words), an adjustment factor for K2 words would then be 1.4 (i.e., 46.4/33.3). It means that if a learner knows a proportion of 33.3% K2 words out of all K2-K5 words that s/he knows in the PVLT (e.g., knowing 10 K2 words and 30 K2-K5 words in total, or knowing 8 K2 words and 24 K2-K5 words in total), we expect 33.3% X 1.4 = 46.4% of his free writing production to be K2 words. By the same logic. The adjustment factors for K3 and K4-K5 words are 0.78 (i.e., 25.9/33.3) and 0.83 (i.e., 27.7/33.3) respectively (Table 4). However, this calculation was only valid if we could assume that native speakers know all of the K2-K5 words in the PVLT. Such an assumption should be valid because when Laufer and Nation (1999) designed the PVLT, native speakers were first asked to validate the test and the results indicated that all the items of the PVLT were retrieved accurately by at least six out of seven of them.

Figure 1  Proportion of occurrence of all levels based on Nation (1990: 17; 2001: 15)
To know is not to use: The gap between students’ productive vocabulary knowledge and their actual use in free writing

Table 4  Calculation of adjustment factors for all levels based on native speakers’ production

<table>
<thead>
<tr>
<th>Levels</th>
<th>Natural occurrence with K2-K5 (%)</th>
<th>PVLT by native speakers (proportions of words known in each level out of K2-K5 levels)</th>
<th>Adjustment factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>K2</td>
<td>46.4</td>
<td>33.3</td>
<td>= 46.3/33.3 =1.4</td>
</tr>
<tr>
<td>K3</td>
<td>25.9</td>
<td>33.3</td>
<td>=25.9/33.3= 0.78</td>
</tr>
<tr>
<td>K4-K5</td>
<td>27.7</td>
<td>33.3</td>
<td>=25.9/33.3=0.83</td>
</tr>
</tbody>
</table>

The computation of the PVLT score with the use of the adjustment factor has a number of advantages. First, as mentioned earlier, the low frequency words should be, logically, more unlikely to appear in an essay. In other words, because low frequency words are by definition less likely to appear in any texts, expecting students to use them may be problematic. For example, ‘abolish’ is a K5 word. It has a frequency of 545 occurrences in a million-word corpus. Therefore, it would be erroneous to assume that ‘abolish’ should appear (at least once) in a student’s writing containing fewer than 500 words. In a similar vein, one cannot claim that students knowing
a certain proportion of low frequency words would produce the same proportion of such words actual writing. Therefore, it was considered appropriate to use an adjustment factor to ensure that the rareness of low frequency words is taken into account in the present comparison. Furthermore, the calculation of the adjustment factors also considered how native speakers actually use language. This was important because to some, at the very least, using English in a similar way to the native speakers may be one of the goals of learning the language. Finally, this way of calculation has also taken into account the ‘imperfect’ vocabulary knowledge of our learners at different frequency levels of K2-K5.

In sum, an adjustment factor helps to translate the PVLT scores (i.e., general productive knowledge) to an expected proportion of occurrence in free writing. For high frequency K2 words, the adjustment factor is 1.40 (46.4/33.3), with the proportion of 33.3% of the PVLT score translating into 46.4% of expected occurrence in free production. For mid frequency K3 words, the factor is 0.78 (25.9/33.3); for low frequency K4-K5 words, the factor is 0.83 (27.7/33.3).

Turning to our participants in this study, we can now calculate the expected percentage of words in the respective K2-K5 levels in a piece of free writing, which can then compared with their actual production as reported in the previous section (Table 5).

Table 5: Comparison of expected and actual production of word families in different levels

<table>
<thead>
<tr>
<th>N=12</th>
<th>K2 (%)</th>
<th>K3 (%)</th>
<th>K4-K5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expected (PVLT X 1.4)</td>
<td>Actual</td>
<td>Expected (PVLT X 0.78)</td>
</tr>
<tr>
<td>Mean</td>
<td>70.45%</td>
<td>61.59%</td>
<td>26.28%</td>
</tr>
</tbody>
</table>

Meanwhile, a similar trend between expected and actual production was observed in that students were expected to produce mostly K2 words, followed by K3 and finally least K4-K5 words, and they have seemingly done so in the free writing tasks. Therefore, it appeared that students’ knowledge of words could be well put to actual use. This might seem contrary to the notion that there are differences between knowing a word and using a word. However, it would be too soon to arrive at a conclusion before conducting statistical tests, which could help us identify potential statistical relationships between the data.
First, we wanted to look more closely into the relationship between expected and actual use of vocabulary and identify any potential correlation. Pearson’s correlation test can help us find out whether these two variables were so closely related to each other that they might be considered one single concept. If a positive and significant correlation is found, it would mean that the expected and actual use of vocabulary generally increase and decrease together. If this is the case, it will suggest that general productive vocabulary knowledge and actual use are likely to be very similar concepts and follow the same developmental pattern. In contrast, if knowledge and use are two different concepts, and hence follow different developmental patterns, negative or no correlation would be expected, meaning that the data increase and decrease in opposite direction, or that they have no linear statistical relationship at all. Shapiro-Wilk tests confirmed that the data fulfilled the normality requirement for the performance of the Pearson’s correlation test (all p-values n.s.). The results are presented in Table 6.

Table 6: Pearson’s correlation test results

<table>
<thead>
<tr>
<th>N=12</th>
<th>K2 words</th>
<th>K3 words</th>
<th>K4-K5 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation between expected and actual production</td>
<td>$r = .098$</td>
<td>$r = -.332$</td>
<td>$r = .181$</td>
</tr>
<tr>
<td>Significance value</td>
<td>$p = .763$ (n.s.)</td>
<td>$p = .292$ (n.s.)</td>
<td>$p = .574$ (n.s.)</td>
</tr>
</tbody>
</table>

All the correlations were non-significant in statistical terms. This lack of correlation can be explained by the arguments presented above that knowing and using vocabulary are two different concepts, and hence there was no statistical linear relationship between general productive vocabulary knowledge and actual use. In other words, we could reject the notion that knowledge is the same as use as some might suggest based on the descriptive data. To put it in a more specific way, the lack of significant correlation showed knowing more K2 (or K3 or K4-K5) words as reflected by the PVLT does not necessarily mean that learners would produce more of those words in free writing.
While the correlation tests suggested that knowing and using vocabulary were indeed different concepts, the picture of the students' use of vocabulary was still very obscure. For example, one logical question to ask was why there were such differences. To seek a clearer picture, a direct comparison of the differences in proportion between the expected and actual use of vocabulary might be revealing. As such, a statistical test of difference was required. Given that normality was met according to the results of Shapiro-Wilk tests (all p-values n.s.), a dependent-samples t-test for each frequency level was carried out. This dependent-samples t-test can identify the potential mean differences between students' expected and their actual production. If there are significant differences found, it can give indications to the existence of a gap between the two concepts. The results are tabulated in Table 7 and presented graphically in Figure 3.

Table 7: Comparison of expected and actual production of word families in different levels

<table>
<thead>
<tr>
<th>N=12</th>
<th>K2 (%)</th>
<th>K3 (%)</th>
<th>K4-K5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expected (PVLT X 1.4)</td>
<td>Actual</td>
<td>Expected (PVLT X 0.78)</td>
</tr>
<tr>
<td>Mean</td>
<td>70.45%</td>
<td>61.59%</td>
<td>26.28%</td>
</tr>
<tr>
<td>Difference (Actual - Expected)</td>
<td>-8.86%</td>
<td>1.38%</td>
<td>-2.52%</td>
</tr>
<tr>
<td>t(11) = 2.586, p=.025*, r=.615</td>
<td>t(11) = -0.588, p=.568, r=.175</td>
<td>t(11) = 1.453, p=.174, r=.401</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 Mean expected and actual proportions of vocabulary at different frequency levels
In the light of the results, with high frequency K2 words, the actual percentage was significantly lower than the expected percentage of production (p<.05) with a large effect size (r=.615). It means that our participants used fewer K2 words in their free writing than what had been predicted by their knowledge in the PVLT. As for K3 and K4-K5 words, there was no significant difference. In other words, while descriptively students used more K3 words and fewer K4-K5 words than expected, these differences were not statistically significant.

Two explanations are proposed here for to account for this pattern. First, students might have given up high frequency K2 words and opt for those with lower frequency (e.g., K3 and K4-K5). Second, there might be a developmental gap between general productive knowledge and actual production. The reasons for each of these proposals are detailed in the following.

First, students could be opting for lower frequency words. There was indeed good motivation for the students to do that because two of the four writings were formal assessments which resembled the public examination format. In the HKDSE writing paper, as in the formal assessments, a good piece of writing (e.g., scoring 5 or above out of a total of 7 in the Language domain) needs to demonstrate a 'wide' vocabulary 'with many examples of more sophisticated lexis' (Hong Kong Examinations and Assessment Authority, 2012: 9). In order to demonstrate their vocabulary knowledge and score higher, they could be trying to avoid the high frequency words (i.e., K2 words) and produce words of lower frequency (i.e., K3 and/or K4-K5 words). However, whether or not this strategy to opt for lower frequency words alone could account for the data fully is reserved for discussion later.

Second, there might be a developmental gap between knowing and actually using vocabulary items. To illustrate, we can focus on the K2 level. The students appeared to have acquired general productive knowledge of the K2 level given an overall attainment of 70% in the PVLT at this level. However, their failure to use the words at this K2 level as much as expected could imply a developmental gap between general productive knowledge and actual production. In fact, the correlation results presented earlier also tended to support the existence of such a gap between knowledge and use. Given that the correlation figures were all non-significant, knowing and using vocabulary items were indeed two different concepts. At this point of discussion, one could logically doubt whether or not either one of the
To know is not to use: The gap between students’ productive vocabulary knowledge and their actual use in free writing

explanations (either a developmental gap or opting for lower frequency words) alone could account for the data fully. It is argued that both factors could have a role to play here.

To explore how both explanations are playing a role in accounting for the t-test results, three hypotheses are proposed here for testing – (1) opting for lower frequency words alone explains our results fully, (2) existence of a developmental gap explains our results fully, and (3) both together explain our results fully.

**Hypothesis 1: Opting for lower frequency words as the sole explanation**

This hypothesis is that students opting for lower frequency words is the sole explanation for the t-test results. In other words, there was no developmental gap between knowledge and use. With the K2 level, the explanation seemed to hold that students might be opting for lower frequency words (i.e., K3 and K4-K5) and hence we found K2 words being produced significantly less than expected. The use of fewer K2 words, then, would logically contribute to the use of more K3 and/or K4-K5 words because a decrease in proportion of actual use at one level should cause an increase at other levels. However, there was no significant difference between the expected and actual use of K3 and K4-K5 words.

With respect to the K3 level, there might still be a chance that hypothesis 1 was correct. One could argue that while students might have replaced some K2 words with K3 ones, they might at the same time have replaced some other K3 words and opted for low frequency K4-K5 ones. In other words, the effect of discarding high frequency K2 words for mid frequency K3 words might have been counter-balanced by the trend of opting for low frequency K4-K5 words. This counter-balancing might explain why we found no significant result for the K3 level and why hypothesis 1 could continue to hold.

However, when the K4-K5 level is considered, it becomes clearer that hypothesis 1 in fact could not be valid. This is because if students opted for lower frequency words to replace K2 and K3 words in a similar fashion, we would see more actual production of K4-K5 words than expected. However, the t-test revealed no such difference. In fact, the actual use of K4-K5 words was even less than expected descriptively. Taken together, while the decreased proportion of use at K2 level should logically lead to an increase in proportion at one or
more levels (K3 and/or K4-K5), this was not the case and therefore, the hypothesis of students opting for lower frequency words as the sole explanation could not hold in light of the results. Although one might suggest that the students could opt for words of even lower frequency (e.g., K5+), readers are reminded that the calculation of proportion in this study was only based on the range of K2 to K5 words. Figure 4 summarises the arguments towards such a conclusion.

![Figure 4: Summary of arguments related to hypothesis 1: Opting for lower frequency words as the sole explanation](image)

**Hypothesis 2: Existence of a developmental gap as the sole explanation**

The second hypothesis contends that the existence of a developmental gap was the sole explanation for the observed pattern. First, there were significantly fewer K2 words being produced than expected in free writing, which appeared to conform to hypothesis 2. Specifically, this could have resulted from the fact that students were not able to put the words that they knew productively into actual use in free writing.
If hypothesis 2 was also true for K3 and K4-K5 production, we might as well find a similar trend of these words being used less than expected. However, this was not the case since no significant difference at these levels was found as mentioned above. In fact, students even used more K3 words than expected descriptively, counter-arguing that a developmental gap could account for the observed pattern fully. Taken together, while the pattern for K2 production might be explained by hypothesis 2, the findings with K3 and K4-K5 words did not conform to this hypothesis, which, therefore, needed to be also rejected.

**Hypothesis 3: Both explanations contributed to the results together**

Hypothesis 3 holds that the two explanations (i.e., opting for lower frequency words and the existence of a developmental gap) together contributed to the t-test results. First, for high frequency K2 words, the two explanations could both contribute to the observed less proportion of K2 production, that is the students would use fewer K2 words than expected.

If there was an effect of students opting for lower frequency words, more actual production of mid frequency K3 words would be observed together with the less use of fewer higher frequency K2 words (as mentioned above when discussing hypothesis 1). At the same time, this expected increase might be counter-balanced by two opposite effects – that some other K3 words were replaced by low frequency K4-K5 words, and that there was a developmental gap making students not able to produce the K3 words that they had knowledge of. Indeed, there was no significant difference between expected and actual use of K3 words and this result could be accounted for by this counter-balancing of different effects. As a result, hypothesis 3 appeared to hold.

Finally, for K4-K5 words, there was also no statistical difference between the expected and actual production in free writing tasks. Such finding can again be possibly due to the balancing of the increase of the use of K4-K5 words related to students being motivated to replace the K2 and K3 words and the effect of the existence of a developmental gap which brings about an observed less use of low frequency K4-K5 words in free writing. Hypothesis 3 still holds for this frequency level of vocabulary. Figure 5 summarises all these arguments related to hypothesis 3.
To know is not to use: The gap between students' productive vocabulary knowledge and their actual use in free writing

Figure 5: Summary of arguments related to hypothesis 3: both explanations contributed to the results together

So far, the validity of three hypotheses for explaining the t-test results has been discussed. The first one is that students have opted for lower frequency items because they wanted to show off their vocabulary. However, it has to be rejected because K3 and particularly K4-K5 words were not being used more often than expected. Second, the existence of a developmental gap between general productive knowledge and actual use could not be the only explanation either. This hypothesis was rejected because of the non-significant differences between the expected and actual use of K3 and K4-K5 words. Hypothesis 3, that both explanations were playing a role together, appears to be the only logical explanation for the findings.

Although findings have been deliberated and considered through their limits, these potential explanations are merely speculative. Further research is needed if the discrepancies between knowledge and use are to be understood. The use of such methods as interviews and think-aloud protocols might shed some light, for instance, on the
issue about whether writing with more lower frequency words than expected was a conscious choice or not. Meanwhile, one thing is clear, and that is the existence of a discrepancy between expected production as deduced from the PVLT scores and the actual use of vocabulary in free writing. Just as the title of this paper has suggested -- knowing a vocabulary item is different from using it in free writing.

As far as the developmental gap is concerned, there appeared to be multiple levels of productive vocabulary. A student knowing all of Nation’s (2001, 2013) aspects of vocabulary knowledge of a word, even up to the aspect of use, may not be able to produce the word freely in their writing. Indeed, as Nation (2013) has indicated, there is usually a big gap of how ‘vocabulary shown to be known on tests like the Vocabulary Levels Test is actually being used in meaning-focused performance’ (p 562). This identified gap also points to the need to look at the comprehensiveness of Nation’s (2001, 2013) aspects of lexical knowledge. The present study is one of the first to identify the discrepancies between knowing and using vocabulary, and it is believed that understanding more about this issue can produce important pedagogical implications for second language learning.

**Pedagogical implications**

One of the most important implications here is the recognition of the insufficiency for students to possess only general productive knowledge of vocabulary. In other words, that students are able to excel in fill-in-the-blanks exercise and assessment tasks does not automatically guarantee that they will be able to use the relevant vocabulary items in an unguided context. Therefore, one key aim for these learners is to be able to use the words appropriately in writing, rather than simply knowing the spelling and grammatical features of the lexical items. Teachers should not stop pushing students to their limits until the latter demonstrate correct use of the lexical items in writing and speaking.

With this goal in mind, there could be a couple of strategies that teachers may consider using. It must be acknowledged, however, that the effectiveness of the proposed teaching strategies may still be subject to further empirical research.

First, teachers should consider assigning more writing tasks with relatively less guidance, such as free writing tasks. For weak learners, the scaffolding may be crucial assistance for them to complete a
writing task. However, especially with the more proficient learners, guidance, or even the provision of a particular context, could promote the use of certain vocabulary items and at the same time possibly posing constraint to the use of others. This constraint could deprive students of the opportunities to formulate and test their hypotheses regarding the use of vocabulary in a trial-and-error manner. In a similar vein, teachers should encourage students to use newly learnt vocabulary items in their writing, again, allowing them to formulate and test hypotheses regarding actual use. Teachers, for example, could ask students to highlight the items that they would attempt to use so that the teacher could pay special attention to these when marking. These attempts may deserve more feedback and encouragement because they are evidence of students' putting efforts into actually using the words that they have recently learnt.

Second, teachers could teach vocabulary in semantically-related sets. For example, ‘vital’, ‘crucial’, and ‘significant’ may be taught together with ‘important’. Presenting items in this fashion could help students assimilate new knowledge to their existing store and enhance their lexical richness through opting for another word, which is exactly what has been observed in the present study and should be further encouraged. For example, students could replace ‘important’ with ‘vital’ in writing and thereby enhancing their lexical richness. However, it is also worth noting that, in a strict sense, no two words are absolute synonyms, meaning that no two words are totally interchangeable across all contexts. Therefore, a simple substitution may not always work. At the same time, to a certain extent, trying to use the new lexical items albeit possibly making mistakes is an important, if not inevitable, step in language learning. Teachers should not discourage the use of items during classroom exercise simply because of the mistakes that might be made. On the contrary, if teachers are able to identify potential confusion, they could provide even more information about the words to make the case clearer for students. Indeed, the confusion arising from the use of semantically-related sets has been reported in and supported by some studies (e.g., Papathanasiou, 2009). However, we note that Papathanasiou (2009) used vocabulary recall (i.e., asking students to recall the target items when given their meanings) as her learning outcome measure. Recall involved students’ general productive knowledge of the vocabulary, while the focus of our study is at the level of free productive use. Importantly, more recent studies (e.g., Ishii, 2015) questioned whether the source of confusion was in fact the semantically-relatedness, or the visual features of the experimental materials. Until more evidence emerges, teaching semantically-related sets is still worth trying.
Another key, but perhaps indirect, aspect of pedagogical implication of this study is the demonstration of potential use of research tools by in-service teachers. While language teachers may still rely heavily on such assessment forms as dictation and fill-in-the-blanks exercise, teachers can make the best of some of the tools used in this study. These tools are useful and reliable as they are already commonly used in research (e.g., Lo & Murphy, 2010). The PVLT is useful for testing the size of one’s productive vocabulary. The Vocabulary Profiler could not only be used by teachers to analyse students’ vocabulary range in actual production. It could also be introduced to students who may use it as a self-assessment tool. It is unfortunate that these useful tools are yet to be fully utilised in the daily process of teaching and learning by teachers. It is hoped that the present study could inspire readers to give it a go.

Conclusion

This study has identified the discrepancy between knowing and using vocabulary in free writing. This discrepancy was attributed to a hypothetical developmental gap and students opting for lower frequency words in order to show off their lexical range. If we agree that one major goal of vocabulary learning is to produce a range of vocabulary learnt in free writing, then we as education practitioners should put in more effort in teaching vocabulary beyond the level of general productive knowledge. We suggested two examples of what we could do to bridge the gap of vocabulary knowledge and use, although the effectiveness of these suggestions still has to be tested by further empirical research. At this point, we would also like to call for more studies about the use of vocabulary in free writing. After all, to know is not to use.
References


Hong Kong Examinations and Assessment Authority. (2012). *Hong Kong Diploma of Secondary Education Examination: English Language (Practice Paper) Marking Schemes and Tapescript*. Hong Kong: HKEAA.


