New Directions In Vocabulary Testing

Stuart A. Webb
Victoria University of Wellington, New Zealand

Yosuke Sasao
Toyohashi University of Technology, Japan

Abstract
There have been great strides made in research on vocabulary in the last 30 years. However, there has been relatively little progress in the development of new vocabulary tests. This may be due in some degree to the impressive contributions made by tests such as the Vocabulary Levels Test (Nation, 1983; Schmitt et al., 2001) and the Word Associates Test (Read, 1993, 1998). In this report, an argument is made that there is a need for the development of new vocabulary tests. The justification for the development of new tests will be discussed and four new tests that are in different stages of development will be briefly introduced. The first two expand on the contributions of the Vocabulary Levels Test. One is a new version of the Vocabulary Levels Test and the other measures knowledge of the different sublists of Coxhead’s (2000) Academic Word List. The second two tests measure a different aspect of vocabulary knowledge, vocabulary learning proficiency. The Guessing from Context Test was designed to measure the ability to guess words in context and the Word Part Levels Test measures knowledge of affixes.

Keywords
Vocabulary Levels Test, affixes, guessing from context, academic vocabulary, depth of vocabulary knowledge

Introduction
There are not many established tests designed to measure what students know about words. Perhaps the most widely used vocabulary tests are the Vocabulary Levels Test (Nation, 1983; Schmitt et al., 2001) and the Word Associates Test (Read, 1993, 1998). The Vocabulary Levels Test measures knowledge of words at different frequency levels and Coxhead’s (2000) Academic Word List. The Word Associates Test measures how well
words are known. These tests were important additions to work on vocabulary assessment and have been of great value to both teachers and learners. However, they are limited to some degree and there is a need for new and improved measures of word knowledge.

The first aim of this report is to look at how existing vocabulary tests might be improved. In particular, the strengths and weaknesses of two of the most important tests of second language vocabulary knowledge, the Vocabulary Levels Test and the Word Associates Test, will be discussed. The reason to look at these tests is that they were designed to measure vocabulary knowledge in different ways. The Vocabulary Levels Test measures the extent to which the form-meaning connection (the extent to which test takers can link L2 form to meaning) is known for sets of words at different word frequency levels. In contrast, the Word Associates Test measures knowledge of different aspects of vocabulary knowledge to determine how well words are known.

The second aim of this report is to describe four new tests that are being developed to fill some of the gaps in vocabulary assessment. The first of these tests is a new version of the Vocabulary Levels Test (NVLT) that improves on the earlier versions by (a) measuring knowledge of the 1000 word level, and (b) measuring knowledge of word frequency levels derived from more current word lists. The second test measures knowledge of the different sublists of the Academic Word List (AWL) rather than measuring knowledge of the AWL as a whole. The third and fourth tests measure vocabulary learning proficiency. These tests should be of great value to teachers because they will help reveal students’ weaknesses in vocabulary learning.

The Vocabulary Levels Test

Perhaps the best known and most widely used vocabulary test is the Vocabulary Levels Test (VLT). The VLT was created by Paul Nation (Nation, 1983) for the purpose of measuring second language learners’ vocabulary knowledge of five different sets of words. Four of the levels measure knowledge of vocabulary within a frequency band: the most frequent 2000, 2001-3000, 4001-5000, and 9001-10,000 word families. The remaining level measures knowledge of academic vocabulary. In the original version of the test, knowledge of Xue and Nation’s (1984) University Word List was assessed. The VLT was updated to its current form by Schmitt, Schmitt, and Clapham (2001). Two useful improvements were made: increasing the number of items at each level from 18 to 30 to increase reliability, and the word list used to evaluate knowledge of academic vocabulary was changed to Coxhead’s (2000) Academic Word List.

Although the VLT can be administered to students as one comprehensive test that is made up of five parts, it is not necessary for students to complete the entire test, because the overall score is not of great importance. Instead it is the scores for each level that are of value. Measuring students’ knowledge of one or two levels is very useful for teachers. Completing all five levels of the VLT can take quite a long time for many learners and measuring knowledge of the lower frequency levels is of value for only advanced language learners. Thus, teachers might often choose to measure knowledge of the levels that are most appropriate for their students such as the 2000, 3000, or academic word levels. Positive features of the VLT are that it is easy to administer and grade, online and hard copy versions are freely available, and bilingual versions of the 2000 word level
have been created in Japanese, Korean, Russian, Chinese, Vietnamese, Thai, Tongan, Samoan, and Tagalog.

The results of the VLT indicate the extent of vocabulary knowledge at each level. This in turn indicates where vocabulary learning is needed. Teachers and students should look to find the highest frequency level that has not been mastered. This will indicate where vocabulary learning should be focused. Mastery of a level may be indicated by a score of 26/30 or higher (Schmitt et al., 2001). For example, if a student has scores of 28/30, 21/30, and 11/30 at the 2000, 3000, and 5000 levels, respectively, then he should focus his attention on learning words at the 3000 level, because the 2000 level has been mastered, but knowledge of the next frequency level (the 3000 level) is still lacking. Scores on the VLT can also indicate the extent of vocabulary learning within institutions (Webb and Chang, 2012), and this in turn may indicate the effectiveness of vocabulary learning programs. This is particularly important in the EFL context where vocabulary learning needs to be optimized.

Despite the improvements made by Schmitt, Schmitt, and Clapham, the VLT still has several limitations. First, it does not include a 1000 word level. The relative value of vocabulary is a function of frequency; words that are encountered and used more often have greater value to learners because they will have a greater impact on comprehension and production. This is particularly true at the highest frequency levels. Nation (2001) reported that the most frequent 1000 word families accounted for 84.3% and 75.6% of the words in conversation and newspapers, respectively, and the second 1000 word level accounted for between 5-6%. The decrease in relative value of vocabulary falls dramatically after the highest frequency levels. Webb and Rodgers (2009) found that 85.11%, 4.42%, 1.93%, 0.71%, and 0.15% of words in television programs were in the 1000, 2000, 3000, 5000, and 10000 word levels, respectively. These figures demonstrate the value of learning the highest frequency words and also highlight the importance of measuring knowledge of the 1000 word level; the large percentage of language that is made up of words from the 1000 word level indicates that learning these items will have a far greater impact on second language learning than mastering any other level.

A second limitation of the VLT is that the lists used for the different frequency levels are rather old. West’s (1953) General Service List was used to create the 2000 word level, and frequency criteria from Thorndike and Lorge (1944) and Kučera and Francis (1967) were used to create the 3000, 5000, and 10000 word lists. Although these lists may still reflect vocabulary at current frequency levels to some degree, there is likely to be some variation between the occurrence of words 50-70 years ago and today. For example, words included in the General Service List that are unlikely to be included as high frequency words today are: scold, coward, conquest, scorn, omit, rival, voyage, hollow, punctual, tremble, solemn, veil, coarse, cape, applause, oar, quarrel, scent, tame, tribe, procession, barber, vain, razor, ceremony, rake, hinder, obedient, cultivate, donkey, limb, wax, shilling, sow, haste, remedy, telegraph, thread, clay, axe, carriage, and cork. Although these items still have value to language learners, the need to know words such as cape, donkey, and shilling has diminished considerably.

Table 1 shows how items in the four frequency levels of the VLT are distributed in Nation’s (2006) British National Corpus (BNC) word lists. The table reveals a fairly large amount of variation between the VLT items and the BNC frequency levels.
Twenty-five percent (15/60 items) of items at the 2000 word level of the VLT occur in less frequent BNC frequency levels with one item (inquire) appearing at the 6000 word level. In contrast, just over 33% of the items at the 3000 word level were found in the most frequent 2000 words of the BNC. VLT items from the 5000 word level were found in the 1000 level of the BNC (bleed, document) and at the 9000 level or lower (desolate, fragrant, wholesome), while items at the 10000 word level ranged between the 4000 level (dent) to outside the 14000 level (salve). Although we should expect some variation between the frequency levels of vocabulary in different corpora, Table 1 indicates that the frequency data from which the VLT items were derived may be outdated.

**New Vocabulary Levels Test**

A new version of the Vocabulary Levels Test (NVLT) is currently being developed to address the limitations of the VLT. The NVLT measures knowledge of vocabulary at five word frequency levels: 1000, 2000, 3000, 4000, and 5000. These levels provide a measure of vocabulary at the levels that are of greatest value to the vast majority of L2 learners. There were two reasons for eliminating the 10000 word level. First, mastery of the 5000 word level may be challenging for all but advanced learners, so assessing knowledge at the five most frequent levels may represent the greatest range in vocabulary learning for the majority of L2 learners. Second, the earlier versions of the VLT have often been misinterpreted as a measure of vocabulary size. The VLT is not a valid measure of vocabulary because there are gaps in vocabulary knowledge that are not assessed.
knowledge of the 4000, 6000, 7000, 8000, 9000, and 10000+ word levels are not measured. Nation and Beglar’s (2007) Vocabulary Size Test is a far superior measure of second language vocabulary size as a whole (but not knowledge of particular levels) because it assesses knowledge of the most frequent 14000 words. Limiting the NVLT to measuring knowledge of the most frequent 5000 word families, may clarify its purpose.

The word lists used to assess knowledge of the different frequency levels were derived from American and British text and were created by Paul Nation (Nation, 2012). The 1st and 2nd 1000 word lists were created using different criteria than the 3000-5000 lists. The first two lists were derived from a greater proportion of spoken text than written text in order to have high frequency word lists that were appropriate for English as a foreign language teaching and course design. The 3000-5000 level lists were based on BNC and the Corpus of Contemporary American English (COCA) word frequency rankings after items from the first two lists were excluded. These new BNC/COCA lists should be representative of current English, and provide a far better indication of the vocabulary being used by native speakers today than the lists used for the creation of the earlier versions of the VLT. The NVLT does not provide a measure of academic vocabulary. The reasons for this were that (a) Schmitt, Schmitt, and Clapham’s (2001) version of the VLT provides a valid measure of the Academic Word List as a whole, and (b) a test measuring the different sublists of the AWL is being developed in concert with the NVLT to provide a more precise measure of knowledge of academic vocabulary (see below).

The NVLT uses the same matching format as the earlier versions of the VLT and includes ten clusters that measure knowledge of 30 items at each level. Fifteen items at each level are nouns, nine items are verbs, and six items are adjectives. Test takers need to match three words from six options to three definitions in each cluster. Appendix A shows one noun cluster for each of the five frequency levels.

**Academic Word Levels Test**

The rationale behind the creation of the Academic Word Levels Test is to provide an assessment tool that reveals test takers’ knowledge of the different sublists of Coxhead’s AWL. There are 10 sublists in the AWL that were created according to the frequency of the items in Coxhead’s academic corpus. The sublists represent academic vocabulary frequency in the same manner that Nation’s (2006, 2012) word lists represent the frequency levels of vocabulary in the BNC and BNC/COCA. Thus, the AWL sublists provide a measure of the relative value of academic words in the AWL. The relative value of the different items is indicated by the lexical coverage of each list and the interval between encounters in text (see Table 2). Lexical coverage of a list is the percentage of text that those items account for. For example in the academic corpus, Sublist 1 provides twice as much coverage as Sublist 2, four times as much coverage as Sublist 4 and 18 times as much coverage as Sublist 9. This clearly indicates that knowing words in Sublist 1 provides much greater value to language learners than knowledge of any of the other AWL sublists. Similarly, the interval between encounters in academic text was 4.3 pages in Sublist 1 and 82.5 pages for items in Sublist 10. Thus, when teaching the AWL, it is important to teach the lists in order because learning Sublist 1 will provide greater benefit to students than Sublist 2, which in turn will provide greater benefit than Sublist 3, and...
When assessing knowledge of the AWL, it is useful to measure knowledge of the items according to the sublists rather than the AWL as a whole. This will allow teachers to determine where they should focus their teaching; beginning with Sublist 1 for students with very little knowledge of the AWL, or beginning at other sublists for those learners who already have mastery of the first sublists.

The AWLT consists of five levels with each level measuring knowledge of two sublists. The first level measures knowledge of Sublists 1 and 2, and each subsequent level measures knowledge of the following two sublists with the fifth level measuring knowledge of Sublists 9 and 10. The rationale for measuring two sublists together was that (a) it represents a manageable vocabulary learning goal (120 words) for a course, and (b) it provides a practical tool that can be used to measure progress in learning the AWL.

The AWLT uses the same matching format as the VLT. The participants are presented with 10 clusters, each of which consists of 6 words from the appropriate level and three definitions. The test takers must match the words with their definitions. The distractors consist of one item each from the target sublists and one item from the following sublist. In the case of Level 5, the third distractor is from the previous sublist. Appendix B presents example items from the AWLT.

We could go a step further than the AWLT and measure depth of knowledge of the sublists. For example, creating a test of collocation for AWL items would provide a better indication of whether learners might be able to use AWL items more effectively than the test of form-meaning connection discussed above. The need for new and improved measures of vocabulary depth will be discussed in the following section.

Word Associates Test

One of the more important innovations in vocabulary assessment was the introduction of the Word Associates Test (WAT) (Read, 1993, 1998). The WAT measures vocabulary knowledge in an original way; rather than indicate whether or not test takers know the form-meaning connection of words, it measures knowledge of multiple aspects...
Webb and Sasao

of vocabulary knowledge to reveal how well words are known. Measuring depth of vocabulary knowledge is useful for several reasons. First, it provides a better indication of what students can or cannot do with words. For example, many EFL students have learned vocabulary through word cards or flashcards. However, the focus on learning form-meaning connection in this activity (and many other activities), while useful for comprehension, may not facilitate use of words; many students have knowledge of the form-meaning connection of words, but are unable to use them. Tests of form-meaning connection such as the VLT and Vocabulary Size Test indicate that learners are able to link form to meaning, but they do not indicate whether learners know which other words are frequently used together with a target word (collocation), or the semantic relationships between words (association). Nation (2001) describes nine aspects of vocabulary knowledge, but most tests only measure one: form-meaning connection.

In the WAT, test takers are presented with a cue word and eight options. Their task is to choose four words that are related to the cue word. These words might be synonyms, collocates, or represent an aspect of the meaning of the cue word. The following two examples are from the WAT.

For the first item, surprising is a synonym of sudden, quick represents an aspect of its meaning, and change and noise are collocates. For the second item, ordinary is a synonym of common, shared represents an aspect of its meaning, and boundary and name are collocates. The WAT does not focus on the extent of knowledge of individual aspects (form-meaning connection, collocations, concept and referents) but the extent of knowledge among the three aspects and provides an overall score. This limits the degree to which vocabulary depth is measured because it is not clear to what extent scores are based on knowledge of form-meaning connection or multiple aspects of vocabulary knowledge.

The focus on depth of vocabulary knowledge in the WAT rather than form-meaning connection alone has generated much interest, and the WAT has become a widely used tool for vocabulary assessment (Greidanus et al., 2004; Qian, 1999; Qian and Schedl, 2004). However, although the WAT provided a valuable step forward in vocabulary assessment, a more useful approach to measuring vocabulary depth might be to develop tests that each examine one aspect of knowledge such as collocation and look at how scores compare to a more traditional test of form-meaning connection. While there have been several attempts at measuring knowledge of collocation (Eyckmans, 2009; Gyllstad, 2009), a validated test is yet to be developed. The value of such a test would be that it might better indicate vocabulary use. Learners often know words but are unable to use
them effectively because they do not know their collocates. Unusual word combinations in speech and writing tend to distinguish non-native speakers from native speakers.

There are many challenges to creating tests of collocational knowledge. Questions that need to be considered are: What is an effective test format that isolates knowledge of collocation? What criteria should be used to select collocations? How should frequency and part of speech be controlled? However, a test of collocation, or another aspect of vocabulary knowledge, could be used to help teachers determine how well students know words, and how their knowledge of collocation is developing in relation to form and meaning.

Tests of Vocabulary Learning Ability

The tests that have been discussed up to this point indicate how many words are known within a particular set of items (VLT, NVLT, AWLT), or how well a particular set of words are known (WAT). These tests can reveal what vocabulary we have learned and what vocabulary we need to learn, but they do not indicate our students’ ability to learn words. Measuring vocabulary learning ability has value because it may indicate to teachers and students how to improve vocabulary learning efficiency, as well as diagnose areas of weakness that may reduce learning potential.

Two tests are being developed to fill this void. One test measures knowledge of word parts (The Word Part Levels Test) and the other test measures the degree to which learners are able to successfully guess words from context (The Guessing from Context Test). Knowledge of word parts and the strategy of guessing from context have long been considered two key components of vocabulary development (Nation, 2001). Recognizing and understanding the different affixes in unknown words when they are encountered can help learners to infer an unknown word’s meaning. Aptitude for guessing unknown words when they are encountered will promote vocabulary learning for those who have greater ability. Moreover, diagnosing our learner’s limitations in these two areas will indicate how we might work to improve our student’s ability to learn vocabulary.

The Word Part Levels Test

The rationale behind the creation of the Word Part Levels Test was that knowledge of word parts is a key component of vocabulary learning; knowledge of the form, meaning, and use of the different affixes will aid us in current and future vocabulary learning. A test that measures the extent of learners’ knowledge of word parts will provide an indication of how they should proceed in their learning. Scores reveal to teachers which word parts remain to be learned, as well as gaps in knowledge between learners. A vocabulary learning program that involves further building of knowledge of word parts can then be implemented to help improve vocabulary learning efficiency.

The Word Part Levels Test is innovative in that it measures three aspects of word part knowledge: form, meaning, and use. Knowledge of form is operationalized as the ability to recognize the written form of an affix (e.g. re- and -less are affixes while su- and -gent are not). Knowledge of meaning is operationalized as the ability to recognize the meaning of an affix (e.g. the meaning of the affixes re- and -less are again and
without, respectively). Knowledge of use is operationalized as the ability to recognize the function of an affix (e.g. adding -ness to the adjective *kind* changes it to a noun, while adding -ize to the adjective *modern* changes it to a verb). Examining different aspects of knowledge separately has two advantages over measuring affix knowledge in a single item. First, it raises awareness of what is involved in knowing an affix for both teachers and learners. Teaching affixes can be challenging for native-speakers because they have rarely explicitly learned word parts. Non-native speakers may often have a better understanding of affixes. However, teaching materials may typically focus on the form-meaning connection of the most frequent affixes making it challenging to comprehensively teach affixes. Second, by isolating and measuring different aspects of affix knowledge, the test can better diagnose learner weaknesses. Teachers can then provide a more effective vocabulary teaching program that caters to an individual’s strengths and weaknesses.

The Word Part Levels Test is being developed based on the results of a more comprehensive measure of word part knowledge that assessed knowledge of 118 different affixes (Sasao, 2013:155-314). The Word Part Levels Test offers a more teacher friendly diagnostic instrument because a comprehensive measure of affix knowledge has limited practical value; it is time consuming, many items are too difficult for beginners, and the results are difficult for teachers to interpret. Rasch analysis was used to identify items at three different levels (easy, medium, and hard). There are 40, 39, and 39 affixes in the easy, medium, and hard levels, respectively. Having tests of affix knowledge at different levels has three advantages. First, it allows teachers to select a test at a level that is more appropriate for their students than a more comprehensive and demanding test. Second, it raises awareness of knowledge of the word parts that should be focused on, as well as the progression of knowledge from one level to another. Third, it provides teachers and learners with a teachable number of word parts for a course. A more comprehensive test may list an unmanageable number of affixes for students to learn making it more difficult to teach word parts effectively.

The three aspects of word part knowledge are measured in separate sections. The first section of the test measures receptive knowledge of written form. Test takers are presented with four options one of which is an affix and the other three are distracters. Prefixes and suffixes are presented in separate items so that there are always three prefix distracters for a prefix and three suffix distracters per suffix. Examples 1 and 2 are for the prefixes dis- and micro-, and examples 3 and 4 are for the suffixes -ful and -ness.

1. (1) sal- (2) cau- (3) lin- (4) dis-
2. (1) micro- (2) cerem- (3) sheph- (4) pecul-
3. (1) -rse (2) -ack (3) -ful (4) -uin
4. (1) -onse (2) -inge (3) -ound (4) -ness

The second section of the test measures knowledge of form-meaning connection using a multiple-choice format. In each item, test takers are presented with an affix and two of the highest frequency words that include that affix and convey the target meaning. Again four options are given, one correct and three distractors. The distractors are all meanings that are conveyed by other affixes. Examples 5-8 are for the affixes mono-, dis-, -ess, and -able.
The final section of the test measures knowledge of the affix function. Some affixes have the function of changing the part of speech of a word. For example, some word parts change a noun to a verb, or an adjective to a noun. For each item, a word part is presented with two example words and the test taker must select the correct part of speech of the example words. In Examples 9 and 10, test takers must recognize that adding the suffix -ment to develop and manage changes their part of speech from verb to noun, while adding -ness to aware and ill changes their part of speech from adjective to noun.

When scoring the test, teachers should look at the scores for each section rather than the test as a whole because this will allow them to give students feedback on each aspect of word part knowledge. Students can then work to improve their knowledge of word parts and this in turn may help them to more effectively learn unknown words that contain these words in the future.

The Guessing from Context Test

Guessing from context is a vocabulary learning strategy that can be used to increase the potential to successfully infer the meaning of unknown words encountered in context.
Because there are thousands of English words and a limited amount of classroom time to devote to vocabulary learning, strategies that can be used independently to aid learning are of great value. Clarke and Nation’s (1980) procedure is perhaps the best known version of the guessing from context strategy. It involves the following five steps:

1: Decide on the part of speech of the unknown word.
2: Look at the immediate context (the sentence in which the word is used).
3: Look at the wider context (the relationship with other sentences).
4: Guess.
5: Check the guess.

The Guessing from Context Test (GCT) provides a means to assess the ability to successfully infer the meanings of unknown words when they are encountered in context. It fills an important gap in vocabulary testing because although several studies have sought to measure the skill of guessing from context (Haastrup, 1991; Schatz and Baldwin, 1986), there are no validated tests that measure this strategy. Justification for the creation of the GCT was that guessing from context is the principal strategy used to independently learn unknown words encountered in context (de Bot, et al., 1997; Fraser, 1999; Hulstijn, 1992; Paribakht and Wesche, 1999). The results of the GCT provide teachers and learners with diagnostic information about the test taker’s skill in three different aspects of the guessing strategy: recognizing the part of speech of unknown words, identifying the information found in context that can be used to derive the meaning of unknown words, and the degree to which learners may successfully recognize the correct meaning of unknown words.

The GCT is made up of 60 questions that are based on 20 passages. For each passage, test takers answer three questions that each measure one subcomponent of the skill of guessing from context identified in earlier studies (Clarke and Nation, 1980; Williams, 1985). The target word in each passage is a pseudoword that replaced a low frequency word. The pseudowords include the same inflectional and derivational suffixes and are about the same number of letters as the original words.

The first section of the GCT involves recognizing the part of speech of an unknown word in a sentence. In the following example, to score correctly test takers must identify the part of speech of the pseudoword *turmilted* as verb.

[Passage] He was brought back to the building on June 9th, but the following day he *turmilted* again and this time was away for 94 days.

[Question] Circle the part of speech of the word ‘*turmilted*’.

A. noun  B. verb  C. adjective  D. adverb

The second section involves identifying the contextual clue that may facilitate successfully guessing an unknown word’s meaning. Each passage contains one type of contextual clue that had been identified in earlier research (e.g. Ames, 1966), and the test taker’s task is to recognize the multiple-choice option that includes the clue. In the following example, to score correctly test takers must identify A as the option that includes information that can be used to help infer the meaning of the pseudoword *turmilted*.
Passage

He was locked up for having injured her. On May 23rd, he ran away for 17 days. He was found and caught when he was in a car taken from another person. He was brought back to the building on June 9th, but the following day he **turmilted** again and this time was away for 94 days.

Question

What word or phrase helped you to work out the meaning of the word? Circle one item that applies best.

A. ran away
B. car
C. for 94 days

The third section involves reading the same passages as those encountered in Section 2 and identifying the unknown word’s meaning from three options. In the following example, to score correctly test takers must identify A as the option that is closest in meaning to the pseudoword **turmilted**.

Question

What is the closest meaning of the word? Circle one item that applies best.

A. escape
B. travel
C. pay

The method of evaluating scores is the same as for the Word Part Levels Test. Teachers should look at the scores for each section rather than the test as a whole because this will allow them to give students feedback on the different components of the guessing from context strategy. Students can then focus their efforts on improving in the areas that the test signals need work. This may help them to more effectively learn unknown words that are encountered in the future.

Conclusion

As we have seen, there have been several very useful tests that were developed to measure vocabulary knowledge such as the Vocabulary Levels Test (Nation, 1983, 1990; Schmitt, et al., 2001) and the Word Associates Test (Read, 1993, 1998). However, as with most areas of research in Applied Linguistics, over time we can find ways to improve on existing research, materials, and tests. This paper has described several new developments in vocabulary testing. These include an updated version of the VLT, a test designed to provide a more precise measure of academic vocabulary, a guessing from context test, and a word part levels test. However, further research on vocabulary assessment is needed. A test measuring knowledge of collocations at different frequency levels would be of great value, and each of the recently developed tests discussed in this article will eventually need improvement.
Funding
This research was supported by Victoria University of Wellington Faculty Research Grants.

References


**Appendix A.** Examples of Clusters at Each Word Frequency Level of the NVLT.

<table>
<thead>
<tr>
<th>1000 level</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>idea</td>
<td>moving air</td>
</tr>
<tr>
<td>2</td>
<td>milk</td>
<td>things</td>
</tr>
<tr>
<td>3</td>
<td>oil</td>
<td>white drink</td>
</tr>
<tr>
<td>4</td>
<td>party</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>stuff</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>wind</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2000 level</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>capital</td>
<td>choice</td>
</tr>
<tr>
<td>2</td>
<td>career</td>
<td>job</td>
</tr>
<tr>
<td>3</td>
<td>committee</td>
<td>test</td>
</tr>
<tr>
<td>4</td>
<td>exam</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>fence</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>option</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3000 level</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>accommodation</td>
<td>end</td>
</tr>
<tr>
<td>2</td>
<td>author</td>
<td>leader of a city government</td>
</tr>
<tr>
<td>3</td>
<td>candidate</td>
<td>person who wrote a book</td>
</tr>
<tr>
<td>4</td>
<td>conclusion</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>division</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>mayor</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4000 level</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>candle</td>
<td>something used for cleaning</td>
</tr>
<tr>
<td>2</td>
<td>diamond</td>
<td>teacher</td>
</tr>
<tr>
<td>3</td>
<td>gulf</td>
<td>valuable stone</td>
</tr>
<tr>
<td>4</td>
<td>salmon</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>soap</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>tutor</td>
<td></td>
</tr>
</tbody>
</table>
Appendix A. (Continued)

5000 level
1 analogy ______ comparison between two things
2 captive ______ person kept somewhere unwillingly
3 remainder ______ what is left
4 renovation
5 ribbon
6 vest

Appendix B. Examples of Clusters at Each Level of the AWLT.

Level 1
1 benefit ______ way something happens
2 focus ______ answer or reply
3 policy ______ official plan or method
4 process
5 reliance
6 response

Level 2
1 attitude ______ total amount
2 code ______ feeling toward something
3 ministry ______ set of rules or laws
4 project
5 sequence
6 sum

Level 3
1 aggregate ______ study of the mind
2 decade ______ act of getting larger
3 enforcement ______ picture
4 expansion
5 image
6 psychology

Level 4
1 advocate ______ person who speaks or acts for someone else
2 device ______ level
3 grade ______ subject
4 manipulation
5 phenomenon
6 topic

Level 5
1 assembly ______ tendency
2 inclination ______ put or connect parts together
3 infrastructure ______ group of experts
4 integrity
5 panel
6 vision