VIBE: COMPUTER-BASED VOCABULARY-SIZE AND STRENGTH TEST FOR TESTING AND LEARNING

Hananto
Universitas Pelita Harapan

Abstract
This paper reports the development of VIBE (Vocabulary Item-Bank of English) from its infancy as a dissertation (Hananto 2007) up to its current usage. VIBE is a computer-based vocabulary size and strength test containing almost 2000 high frequency words from the General Service List (West 1953) and 570 words form the Academic Word List (Coxhead 2000). VIBE has a double task measuring two different strengths of word knowledge: active recall and passive recognition. The active recall task requires the test-takers to supply one missing letter of the English target word and the passive recall task asks them to select the meaning of the target word in Indonesian. Because of the double tasks, three different scoring systems are possible based on: (1) the active recall task, (2) the passive recognition task, and (3) both of the two tasks. VIBE was designed not only for testing but also for learning because in computer-based tests it is possible to do so. Recently, VIBE has also been used for VocComp - a new vocabulary competition using a computer connected to a projector. The competition is believed to be more educational, challenging, and fun than the Spelling Bee and Scrabble competitions commonly held in Indonesia.

Introduction
This paper describes the development of a computer-based vocabulary-size test named Vocabulary Item-Bank of English (VIBE). VIBE is unique in terms of its design, i.e. not only for testing purposes but also for teaching/learning. Lately, it was found out that it can also be used for fun as a computer-based competition named VocComp as an alternative to Spelling Bee and Scrabble, the widely used vocabulary competitions in Indonesia. VIBE also offers a new test format that enables us to measure two strengths of vocabulary knowledge. Before this paper reports the development of VIBE, it starts with the discussion of vocabulary studies in Indonesia and vocabulary size tests in general.
Vocabulary Studies in Indonesia

Since the late 1970s, there has been a revival of interest in vocabulary teaching. Vocabulary has been seen as a priority area in language teaching. Wilkins (1972: 111), for example, writes: ‘without grammar, very little can be conveyed, without vocabulary, nothing can be conveyed.’

In Indonesia, however, vocabulary has not been a main concern in English teaching. This can be seen from the fact that vocabulary teaching in Indonesia is not given sufficient time allocation compared to the teaching of English grammar. The results of vocabulary-size studies conducted in Indonesia show that, in general, Indonesian EFL learners do not have sufficient vocabulary knowledge needed to be able to communicate fluently in English. These studies include Quinn (1968), Nurwani (1995), Kweldju (1997), Nurwani and Read (1999), Kweldju et al. (2000), and Kweldju (2004).

Many Indonesian EFL scholars, however, often have misunderstandings about the simple appearance of the vocabulary tests used in those studies. The tests seem to break many, if not, all communicative language teaching principles. Therefore, it is necessary to discuss the nature of what vocabulary-size tests to avoid the misconceptions.

Vocabulary-Size Tests

Some researchers describe lexical knowledge as a taxonomy of components (e.g. form, grammatical pattern, meaning, function, relation with each other). Most vocabulary tests based on Nation’s components of vocabulary knowledge usually measure just one of the subknowledges.

When just one sub-knowledge is tested, it is possible to test a large number of lexical items. Therefore, the test can claim to represent the learner’s total vocabulary. Such tests are called vocabulary breadth or size tests. This type of vocabulary test has been criticized for the superficial treatment of each item (i.e. testing on one sub-knowledge only by focusing directly or indirectly on word-form and word-meaning relationship).

Most vocabulary size tests are discrete in nature and presents words in isolation and without reference to any linguistic context. Others require test-takers to select the meaning for the words with sentence contexts.
Alternatively, vocabulary tests that measure each lexical item for several areas of knowledge are known as vocabulary depth tests. The limitation of these tests is that it is not feasible to cover a large number of items; consequently, the items do not constitute a representative range of the target items. For that reason, it does not represent the true vocabulary of the test-takers.

An ideal test of a learner’s entire vocabulary knowledge, therefore, would need to measure the breadth of knowledge, (i.e. how many words s/he knows) and also the depth of knowledge (i.e. how well each word is known). However, the vocabulary sample for such a test would not only have to be large enough to represent the testee’s total vocabulary, but would also enable each word to be tested on all aspects of knowledge. Generally, it is difficult to do both vocabulary size and depth at the same time in one single test.

Compared to other vocabulary tests as they are available at the moment, vocabulary-size tests have some advantages. One of them is that vocabulary size tests have been shown to predict success in reading, writing, and general language proficiency as well as academic achievement (Laufer and Goldstein 2004). Read (2000:115) also admits that “vocabulary size has received more attention because, despite the fact that the tests may seem superficial, they can give a more representative picture of the overall state of the learners’ vocabulary than an in-depth probe of a limited number of words.”

Vocabulary size tests are usually low-stakes tests (i.e. with no important consequences for the test taker). They are most likely to be used by classroom teachers for assessing progress in vocabulary learning and diagnosing areas of weaknesses.

Vocabulary size tests have different degrees of knowledge strengths because knowing a word is an incremental process. Some lexical researchers use different terms such as receptive vs. productive and recognition vs. recall. Laufer and Goldstein (2004) try to overcome the confusion between receptive (passive) and productive (active) vocabulary knowledge by distinguishing four degrees of word knowledge: (1) active recall (supplying the L2 word), (2) active recognition (selecting the L2 word), (3) passive recall (supplying the L1 word), and (4) passive recognition (selecting the L1 word).
The following are some well-know vocabulary size tests:
2. The Eurocentres Vocabulary Size Test (Meara and Buxton 1987, Meara and Jones 1988).
4. The Vocabulary Size Placement Test (Alderson 2005), shown in Figure 1. (downloaded from http://www.lancs.ac.uk/researchenterprise/dialang/about)

![Figure 1. A screenshot of the Vocabulary Size Placement Test](image)

VIBE

The existing vocabulary-size tests have one common feature: they were all designed for vocabulary-size tests only. There was a need to develop a vocabulary-size test for Indonesian EFL students that can be used not only for testing but also for learning purposes. VIBE was started in 2004 as a dissertation project and completed (Hananto 2007).

VIBE used two well-known word-frequency lists: (1) the General Service List of English (West 1953) and (2) the Academic Word List. The latter is a compilation of 570 word
families that occurred with wide range and high frequency in a specially constructed
corpus of academic texts (Coxhead 2000). The target words from the two word lists
were divided into three word-frequency levels: (1) 1000 word level (961 target words),
(2) 2000 word level (985 target words), and (3) academic word level (570 target
words). Each word family was represented by its head word and some words were
excluded from the item-bank. All one-letter words (e.g. l and a) and unimportant two-
letter words (e.g. oh) are not included. In total, there were 2516 target words.
VIBE adopts the data-based approach to creating the questions. Instead of creating the
questions one by one, the program will automatically offer various objective test
formats to be selected. The following are some sample data from the 1000 word level:

ABLE = dapat, bisa, mampu
ABOUT = tentang, kira-kira, hampir
ABOVE = yang di atas; lebih tinggi dari
ACROSS = di seberang
ACT = bertindak, tindakan

The various test formats generated by a computer program are not equally reliable and
valid. Some formats are more reliable and valid than the others. During the validation
(Hananto 2007), it was found that a combination of one-letter gap filling and five-
option multiple-choice format (Figure 2) was promising.
Because it is an item-bank, VIBE can generate almost unlimited number of parallel test forms, both computer-based and paper-based. It can select a certain number of vocabulary items randomly so that anybody who takes the test would likely to get a test form that would be different from the test forms that had been used before.

**The Next Generation of VIBE**

The Next Generation of VIBE is a three-year research project funded by DIKTI/the Directorate General of Higher Education, Ministry of Education. The purpose is to improve the software and the database of VIBE. In the first year, the plan is to enlarge the VIBE database to include the Oxford 3000™ wordlist, available online at: http://www.oxfordadvancedlearnersdictionary.com/oxford3000/

In the second year, the various test formats generated from the database will be investigated to select reliable and valid test formats. In the third year, the effectiveness of the software for vocabulary teaching/learning purposes will be examined.
The database will consist of four fields: (1) the target word, (2) its definition or synonym, (3) Indonesian meaning/translation, and (4) a sentence context, for example (the real database will be in a form of table consisting of four columns):

ABLE  =  good at doing something  
      dapat, bisa, mampu  
      She is an able student.

ABOUT =  relating to a particular subject  
       tentang, kira-kira, hampir  
       He lied about his age.

ABOVE =  more than something else  
      yang di atas; lebih tinggi dari  
      The bird flew above our heads.

ACROSS =  from one side of something to the other  
      di seberang  
      They ran straight across the road.

Based on the database above, more than twenty objective formats can be obtained. Below are fifteen format samples.

1. ABLE  
   a) good at doing something  
   b) relating to a particular subject  
   c) more than something else  
   d) from one side of something to the other

2. good at doing something  
   a) ABLE  
   b) ABOUT  
   c) ABOVE  
   d) ACROSS

3. ABLE  
   a) dapat, bisa, mampu  
   b) tentang, kira-kira, hampir  
   c) yang di atas; lebih tinggi dari  
   d) di seberang
4. dapat, bisa, mampu
   a) ABLE
   b) ABOUT
   c) ABOVE
   d) ACROSS
5. ABLE = good at doing something (True/False)
6. ABLE = tentang, kira-kira, hampir (Benar/Salah)
7. good at doing something = _ _ _ _ 
8. good at doing something = a _ _ _ 
9. good at doing something = a _ _ e
10. good at doing something = a _ le
11. A _ LE
    a) good at doing something
    b) relating to a particular subject
    c) more than something else
    d) from one side of something to the other
12. A _ _ E
    a) good at doing something
    b) relating to a particular subject
    c) more than something else
    d) from one side of something to the other
13. She is an ABLE student. ABLE =
    a) good at doing something
    b) relating to a particular subject
    c) more than something else
    d) from one side of something to the other
14. She is an _ _ _ _ student. (good at doing something)
15. She is an a _ _ e student. (good at doing something)

The test formats could be multiple-choice, True/False, gap-filling, or combination of gap-filling and multiple-choice. They could have different strengths (e.g. format 1 and 2) and sensitivity (e.g., format 9-10). The language could be monolingual (e.g. format 1 and 2) or bilingual (e.g. format 3 and 4). The target words could be without sentence context (format 1-12) or with sentence context (format 13-15).

Recently, VIBE has been used for vocabulary competitions (named VocComp). It is believed to be more educational and challenging than the Spelling Bee and Scrabble competitions. After several internal tryouts, VocComp was officially held by the English
Department of UPH in August 2010 for senior high school students. VocComp is planned to be held annually in August during the UPH Festival. VocComp has been adopted and adapted by some educational institutions.

Conclusions and Recommendations

VIBE has unique properties and advantages. It is a computer-based vocabulary item-bank that can be used for multi purposes: vocabulary size testing, vocabulary list learning, and vocabulary competitions. As a vocabulary size test, it offers various test formats with different degrees and knowledge strengths, and some of them are innovative. They offer different measurements of knowledge strengths and sensitivity. Additionally, it can generate almost unlimited number of parallel test forms.

The Next Generation of VIBE is funded by the Indonesian government. It will offer a new field of research. It will take a lot of time and efforts to investigate the reliability and validity of the various test formats and parallel forms generated by VIBE. Researchers are welcome to explore this new field. It is expected that in the near future, the Next Generation of VIBE will make more contributions to development of vocabulary studies in Indonesia.

Works Cited


