

## Developing a Web-Based Vocabulary Size Test for Indonesian EFL Learners

**Ignatius Tri Endarto<sup>1</sup>, Adaninggar Septi Subekti<sup>2</sup>**  
ignatius.endarto@staff.ukdw.ac.id<sup>1</sup>, adaninggar@staff.ukdw.ac.id<sup>2</sup>

Universitas Kristen Duta Wacana<sup>1,2</sup>

### Abstract

One important aspect of English language learning is vocabulary. However, the availability of vocabulary proficiency tests, both manual and computer-based, is still limited. Almost all currently existing vocabulary tests are monolingual and have not accommodated the needs of EFL learners. The objective of this study was to develop a web-based vocabulary size test for English learners in Indonesia. It consisted of 6 stages, namely: (1) needs analysis; (2) corpus selection and corpus data analysis; (3) sampling to select words from corpus data; (4) generating test items; (5) developing a web-based version of the test and pilot study; and (6) reflection and evaluation. Pilot study involved having a sample group of English learners in Indonesia try out the designed vocabulary size test. A paired-samples t-test was then used to determine whether there was a significant difference between the learners' scores on the designed test and those on a free online vocabulary size test not specifically designed for Indonesian test takers. The t-test indicated that there was a significant difference between the designed test scores ( $M=6433$ ,  $SD=1169$ ) and free online test scores ( $M=5700$ ,  $SD=1616$ );  $t(35)=-3.597$ ,  $p = 0.001$ . The identifiable factors which seemed to contribute to this difference were different corpus data (COCA vs. BNC), different varieties of English (American vs. British), and different formats (bilingual vs. monolingual). Despite the difference, there was a positive correlation between the two tests (Correlation = 0.657) meaning that higher scores on one of the tests equals higher scores on the other.

**Keywords:** corpus, EFL, vocabulary, vocabulary size test

### Introduction

Acquiring a foreign language is an intricate process (Barin, 2011). It comprises several interrelated elements language learners need to master. One of its most basic and pivotal pillars is vocabulary. The more words one picks up, the more he/she becomes proficient in that language. It is in line with Alavi and Akbarian (2012) who emphasize the importance of vocabulary in assessing language proficiency. Standardized English tests, such as TOEFL, often address this aspect specifically in their reading section. Mustafa (2019) also claims that vocabulary is the most fundamental aspects of language which needs to be taught prior to other skills.

In Indonesia, many colleges and schools have given a vital status to teaching vocabulary by integrating it into their English lessons or, for those with English majors, simply offering courses on vocabulary as preparatory subjects. However, there are some issues that arise in relation to the integration of vocabulary in English subjects at schools or the implementation of vocabulary courses at colleges. As stated by Mustafa (2019), no particular research has been dedicated to address the issues of vocabulary teaching in the country.

The first problem is the very small number of sources available to facilitate teachers to test and measure students' vocabulary proficiency. Besides, most of the available

vocabulary tests are not specifically designed for the needs of English learners in Indonesia as they are either monolingual (using English only) or non-Indonesian bilingual (using English and a language other than Indonesian). In fact, bilingual tests are said to be more appropriate for assessing EFL learners' vocabulary knowledge than the monolingual ones (Daller, Milton, & Treffers-Daller, 2017). Due to this lack of bilingual (English-Indonesian) tests, teachers have to develop their own vocabulary tests or modify the available sources to make them more suitable for Indonesian EFL (English as a foreign language) learning contexts.

The second issue is the need for a standard framework of reference or a leveling system which best reflects students' vocabulary acquisition. In the case of vocabulary course implementation, different lecturers teaching separate or parallel vocabulary courses can indeed help to accommodate more diverse learning needs and bring more techniques, knowledge, as well as other positive qualities into the classrooms. However, a teacher's perspective can, and often, be different from one another. For that reason, continuation of learning will be ensured if they refer to a common framework or system for assessing students' vocabulary levels. Since a good selection of materials is crucial in vocabulary classes (Basuki, Damayanti, & Dewi, 2018), such a framework of reference or a leveling system can help to ensure teachers appropriately decide which target words to teach.

The last thing to consider is the scarcity of free web-based vocabulary tests which accord with the objectives of the courses. Regardless of its central role in language acquisition (Alqahtani, 2015; Nation, 2006; Schmitt, 2000), vocabulary is still somewhat uncared for when it comes to EFL classrooms in Indonesia (Hananto, 2013; Sudarman & Chinokul 2018). As a result, very few vocabulary tests, fewer of which are web-based, have been developed for Indonesian contexts. In fact, measuring one's vocabulary size or depth is not an easy thing as hinted by the second edition of the 20-volume Oxford Dictionary which lists as many as 171,476 words in English. Therefore, good paper-based vocabulary size tests for a large class, for example, will waste a whole lot of paper. A well-developed web-based test, on the other hand, will easily address this issue and sustain its reusability for as many classes as possible.

In response to the abovementioned issues, this research aimed at developing a bilingual (English-Indonesian) vocabulary size test to measure Indonesian EFL learners' proficiency. Moreover, it focused on designing a web-based version of test to make it more accessible to test takers. The research questions were formulated as follows:

1. What are the characteristics of an appropriate web-based vocabulary size test for Indonesian EFL learners?
2. Is there any significant difference between the designed vocabulary size test and another free online vocabulary size test not specifically designed for Indonesian test takers?

Practically, this research provides English teachers/lecturers in Indonesia with a web-based model of tool for assessing students' vocabulary size (the estimated number of words they know). Theoretically, it seeks to lay the foundation of vocabulary-size-test development in the contexts of Indonesian EFL teaching.

## **Theory**

### **Vocabulary Knowledge and Language Proficiency**

Vocabulary is a very important part of language and thus one of teachers' jobs is to assess their students' knowledge of the target language words (Schmitt, Schmitt, &

Clapham, 2001). There have been various research done on vocabulary knowledge and acquisition. Alavi and Akbarian (2012) in their study categorize the knowledge of vocabulary into two: depth and size. The term 'depth' here refers to how comprehensive a word is known by learners. In contrast, the term 'size' shows how many words students acquire in their target language capacity.

Several other studies show that there is a strong correlation between vocabulary knowledge and language proficiency. Vocabulary size, according to Alavi and Akbarian (2012) is a significant indicator of academic success. Furthermore, the results of some standardized tests such as IELTS and TOEFL, particularly their reading scores, are found to be highly intercorrelated with vocabulary knowledge (Qian, 2002; Akbarian, 2008).

According to Nation and Beglar (2007), English language learners will need to master different numbers of word families when they read different types of texts. For instance, to get 98% coverage of children's movies, students will need to master at least 6000 word families. Alternatively, to enjoy more advanced texts such as novels, they are required to be proficient with approximately 9000 word families.

**Table 1. Vocabulary Sizes to Understand Different Types of Texts  
(Taken from Nation & Beglar, 2007: 9)**

Texts	98% coverage	Proper nouns
Novels	9,000 word families	1-2%
Newspapers	8,000 word families	5-6%
Children's movies	6,000 word families	1.5%
Spoken English	7,000 word families	1.3%

### **Vocabulary Testing and Vocabulary Size Test**

Although vocabulary is a pivotal element of language proficiency which deserves much attention from both language teachers and researchers, there is still lack of appropriate vocabulary tests available for use (Schmitt, Schmitt, & Clapham, 2001). Out of the two types of vocabulary knowledge mentioned earlier, namely depth and size, the latter has attracted more attention from language researchers (Karami, 2012).

Knowing vocabulary size can bring a couple of benefits to language teachers or researchers. In designing a language program, even language testers will often need to measure their students' size of vocabulary (Karami, 2012). More importantly, vocabulary size tests may contribute well to language teaching by helping teachers to know whether students possess enough lexicon to perform a particular task, monitor their vocabulary learning, select the suitable instructions for students, develop an appropriate curriculum, evaluate a language course/program, and many others (Beglar, 2010).

The most widely known vocabulary size test available is the VLT (Vocabulary Levels Test) which was created by Paul Nation (Ishii & Schmitt, 2009). This test uses 4 frequency levels of words, from 2000 to 10000 together with academic vocabulary. It is a monolingual test containing both words and definitions in English.

Another most frequently used test is called the Vocabulary Size Test (Nation and Beglar, 2017). The word family frequency lists utilized in this test were retrieved from the BNC Corpus (Nation as cited in Karami, 2012). In total there are 14 word family frequency levels in the test (each consists of 1000 words) and it takes ten words as the sample from each level. Ordered based on their frequency levels, the 140 words are then presented as the representatives of the 14000 words (Karami, 2012).

Among a number of well-known vocabulary size tests which have been developed—e.g. the Vocabulary Levels Test and the Vocabulary Size Test—none has been specifically designed for Indonesian EFL learners. Therefore, this research will focus on developing a bilingual (English-Indonesian) vocabulary size test aimed at measuring Indonesian EFL learners' proficiency.

### **Research Method**

This research employed a quantitative corpus-based method, whose corpus data were COCA's collection of various texts with a total of 120 million words. The data were used to generate a list of the most frequent words in the corpus. Based on the list of 20,000 most frequent words, a sample of 100 words was selected systematically based on their frequency ranking to be, at a later stage, developed into a complete set of vocabulary size test questions.

There were 6 main stages of this research and each of them can be explained in detail as follows:

1. Needs analysis

The needs analysis was done by interviewing a head of English department, vocabulary lecturers, and an English education department student at a university in Indonesia. The purpose of this needs analysis was to gather some ideas about the most ideal type of vocabulary size test. The interview results indicated that a web-based vocabulary size test was certainly needed in their institution to provide supporting data or evidence related to students' approximate number of acquired words. The data were said to be useful for teachers or lecturers to design both curricula and materials for their vocabulary courses.

2. Selecting the corpus and analyzing corpus data based on word frequency

At this stage, an online corpus was selected to generate a list of word families ranging from low to high frequencies. Corpus of Contemporary American English (COCA) was chosen as the data source, considering that it is an online corpus often used as a reference and has relatively complete features with a very large collection of academic texts.

3. Systematic sampling to generate a sample from each frequency level

From the most frequent word families on the list, a 100-word sample was generated systematically to represent the other lexicons on the list. All the words in the sample were later used to develop the vocabulary size test.

4. Creating test items for each word in the sample

The sample consisted of 100 words representing 20,000 words on the list. From those 100 words in the sample, 100 multiple-choice questions were made.

5. Developing a web-based version of the test and conducting a pilot study

After the test items were made, a web-based version of the test was developed. A pilot study involving Indonesian EFL learners (36 students of an Indonesian university) was conducted. Paired-samples t-test was then used to see whether there was a significant difference between the results of the developed test and those of other online monolingual test not specifically designed for Indonesian EFL learners.

6. Evaluation and Reflection

Evaluation and reflection were done by examining the correlation between the students' test results from the pilot study (using the designed test) with those from another vocabulary size test not specifically designed for Indonesian EFL learners.

As stated earlier, in addition to examining the correlation between pilot study scores and another vocabulary size test scores, the validity and reliability of the designed test will be ensured by conducting a paired-samples t-test on the results of the pilot study.

## Findings and Discussion

### The Characteristics of Web-based Vocabulary Size Test Designed in this Study

As stated by Nation (2012), there are some features that a good vocabulary size test has. They are:

1. The test is applicable to students from various levels of proficiency
2. It is consistent with “what to measure”.
3. It shows some expected behavior such as giving learners the right proficiency levels, consisting of test items with various difficulties in accordance with the word frequency levels, and helping learners to measure their level progression over time.
4. It has the reliability and consistency in measuring vocabulary proficiency, regardless of different circumstances such as genders of test takers.
5. It's not difficult to score and the scores are easy to interpret.
6. The test items avoid ambiguity.
7. It minimizes inefficiency.

In developing the web-based vocabulary size test, this research tried to meet those Nation's criteria. However, due to time limitations, not all of them were able to be fulfilled.

*Criteria 1 & 2: The test is applicable to students from various levels of proficiency & consistent with “what to measure”.*

In order to maintain its consistency of measuring what is meant to be measured, the test was developed using COCA, a reliable source of corpus data with a 120-million-word collection of texts from various genres. Based on the list of 20,000 most frequent words in the corpus, a sample of 100 words was selected systematically based on their frequency ranking. Some examples of words in the sample are:

**Table 2. Examples of Words in the Sample**

No.	Word Frequency Ranking	Word	Pos	Field
1	200	run	v	
2	400	music	n	Hum
3	600	third	m	
4	800	growth	n	
5	1000	sexual	j	Soc

As those 100 words were systematically chosen based on their frequencies of occurrence, each was used to represent the other 199 words which occurred more frequently in the word frequency ranking. A hundred multiple choice questions were then formulated from the 100 words in the sample, and the web-based vocabulary size test developed in this study was made up of those 100 questions. The designed test was applicable to students from various proficiency levels because it contained a sample of the most frequent (basic level) to the least frequent (advanced level) words in the corpus.

*Criteria 3 & 4: It shows some expected behavior such as giving learners the right proficiency levels, consisting of test items with various difficulties in accordance with the word frequency levels, and helping learners to measure their level progression over time. It also has the reliability and consistency in measuring vocabulary proficiency, regardless of different circumstances such as genders of test takers.*

Due to time limitations, Nation's criteria 3 and 4 were not fully assessed in this research. Further studies are needed to see whether the designed test meets the two criteria. Nonetheless, both of them were partly measured in this research through a pilot study which involved 36 students of an Indonesian university and a paired-samples t-test to see whether there was a significant difference between the results of the developed vocab test and those of other well-established test not specifically designed for Indonesian EFL learners. The pilot study and t-test results are reported in the next section.

*Criterion 6: The test items avoid ambiguity.*

Nation's criterion 6 was ensured in this research through the test format. Each question in the designed test consisted of both a target word and an example of English sentence containing the word. The sentence examples were given to provide each target word with its intended context. There were 4 options, all of which were in Indonesian language, for each question: 1 correct option, 2 wrong options/distractors, and an "I don't know"/"tidak tahu" option. The "I don't know"/"tidak tahu" option was employed to avoid the practice of guessing the answer by test takers who did not know the meaning of the word being asked. Here are some examples of the test items and complete test instructions in both languages (English and Indonesian):

**This test consists of 100 multiple-choice questions.**  
For each question, choose 1 (one) best option that most closely matches the meaning of the word in bold. If you don't know the word, select "Tidak tahu".

\*\*\*\*\*

Untuk masing-masing pertanyaan, pilih 1 (satu) jawaban yang paling sesuai dengan arti kata yang dicetak tebal. Jika Anda tidak tahu arti dari kata yang dimaksud, pilih opsi "Tidak tahu".

4. \*

**Run:** The athletes **run** as fast as they can to the finish line.

Mark only one oval.

☐ Menuju  
☐ Berlari  
☐ Berjalan  
☐ Tidak tahu

5. \*

**Music:** Listening to **music** is one of his greatest joys.

Mark only one oval.

☐ Susunan suara yang mengandung irama, lagu, nada, dan harmonisasi  
☐ Perangkat elektronik yang memproduksi sinyal atau gelombang suara  
☐ Instrumen penghasil suara yang dimainkan dengan jari-jemari tangan  
☐ Tidak tahu

**Figure 1. Examples of Test Items**

*Criteria 5 & 7: It's not difficult to score and the scores are easy to interpret & minimizes inefficiency.*

By using a web-based version of the test, both scoring and interpretation of the scores were done automatically. The test was also designed in an effort to minimize inefficiency given that it comprised only 100 questions to represent 20,000 most frequent words in the corpus.

### **Pilot Study and Paired-Samples T-Test**

The participants of the pilot study were 36 college students. They were all Indonesian EFL learners from 2 different batches (2 classes). During the pilot study, they were asked to take 2 vocabulary size tests: the developed vocabulary size test and another online test available on the internet (VocabularySize.com). The results of each test can be seen in the following table:

**Table 3. Pilot Study Results**

Timestamp	VocabularySize.com	Designed Vocabulary Size Test
8/28/2019 9:45:16	6500	6000
8/28/2019 9:46:02	5300	5600
8/28/2019 9:48:20	6000	6500
8/28/2019 9:49:53	7300	7100
8/28/2019 9:52:12	5200	5500
8/28/2019 9:52:27	6000	6200
8/28/2019 9:52:30	4900	7200
8/28/2019 9:53:05	6300	7700
8/28/2019 9:53:39	4500	6000
8/28/2019 9:56:45	4900	7700
8/28/2019 9:56:56	4000	5400
8/28/2019 9:57:48	6200	7200
8/28/2019 9:57:59	3200	4100
8/28/2019 9:58:12	4400	5300
8/28/2019 9:58:41	3200	6500
8/28/2019 9:59:12	3900	6400
8/28/2019 10:00:26	3800	5500
8/28/2019 10:00:51	4700	6100
8/28/2019 10:02:15	7600	7800
8/28/2019 10:04:36	7800	7200
8/28/2019 10:05:51	4400	3800
8/28/2019 10:09:12	8400	7200
8/28/2019 10:12:57	3300	4200
8/28/2019 10:13:24	6400	4300
8/29/2019 12:46:51	3800	5000
8/29/2019 12:48:17	6400	7400
8/29/2019 12:48:28	8000	8300
8/29/2019 12:49:48	9300	8100

8/29/2019 12:49:53	6300	7300
8/29/2019 12:51:46	6700	7300
8/29/2019 12:53:36	7200	6800
8/29/2019 12:53:38	7100	7200
8/29/2019 12:57:18	3500	6800
8/29/2019 12:57:32	6500	7300
8/29/2019 13:03:51	7200	7000
8/29/2019 13:05:23	5000	6600

To compare the scores of the two vocabulary size tests, paired-samples t-test was carried out to see whether there was a significant difference between the two sets of scores. The results of paired-samples t-test are presented as follows:

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VocabularySize.com	5700.00	36	1616.345	269.391
	Designed_Test	6433.33	36	1169.127	194.854

  

Paired Samples Correlations			
		N	Correlation
Pair 1	VocabularySize.com & Designed_Test	36	.657

  

Paired Samples Test								
		Paired Differences			95% Confidence Interval of the Difference		t	df
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper		
Pair 1	VocabularySize.com - Designed_Test	-733.333	1223.111	203.852	-1147.174	-319.492	-3.597	35
								Sig. (2-tailed)
								.001

Figure 2. T-Test Results

The t-test results showed that there was a significant difference between scores of the designed vocabulary size test ( $M = 6433$ ,  $SD = 1169$ ) and those of other test ( $M = 5700$ ,  $SD = 1616$ );  $t(35) = -3.597$ ,  $p = 0.001$ . Although there was a significant difference, there was also a positive correlation between the two tests (Correlation = 0.657) indicating that participants who got a high score on one test tended to have a high score on the other test too.

Based on the evaluation and reflection, some factors that might contribute to the significant difference between the results of the two tests were:

1. Different corpus data which were used to develop the two tests  
The two tests were designed based on two different corpus data. The vocabulary size test in this research was developed using Corpus of Contemporary American English (COCA) as its data source. VocabularySize.com, on the other hand, collected the data from British National Corpus (BNC).



2. Different varieties of English in the two tests.

As the names suggest, COCA is a corpus that contains the data of American English texts while BNC compiles texts which are mostly in British English.

3. Different formats used by the two tests

The vocabulary size test in this study was designed specifically for EFL learners in Indonesia, hence its bilingual format (English and Indonesian). On the contrary, VocabularySize.com is a monolingual test whose instructions, questions, and answer choices are all in English.

### Conclusion

As indicated in Findings and Discussion, the designed test in this research had most, though not all, of the characteristics of a good vocabulary size test proposed by Nation (2012). Those characteristics were: (1) applicable to students from various levels of proficiency; (2) consistent with “what to measure”; (3) avoids ambiguity; (4) not difficult to score and the scores are easy to interpret; and (5) minimizes inefficiency. Due to time limitations, two other criteria were not fully assessed in this study.

The t-test results indicated that there was a significant difference between the scores of the developed test and those of other tests available on the internet. Some identifiable factors that might to the difference were different corpus data (COCA vs. BNC), different varieties of English (American vs. British), and different formats (bilingual vs. monolingual). Despite the significant difference, there was a positive correlation between the two tests which indicated that participants who got a high score on one test tended to have a high score on the other test.

In order to validate the findings of this research, further studies comparing the designed test to another vocabulary size test that has more similar characteristics (similar corpus data, English varieties, and formats) are needed. Besides, more research on the designed test which involves larger number of participants will also be beneficial to the development of such a test in Indonesian EFL contexts.

### Acknowledgments

We thank the Institute of Research and Community Service of Universitas Kristen Duta Wacana (LPPM UKDW) for the support which has made this research possible.

### References

- Akbarian, I. 2008. *The Role of Vocabulary Knowledge in Predicting Performance on Reading Comprehension Item Types*. Doctoral dissertation, University of Tehran, Tehran, Iran.
- Alavi, S. M. & Akbarian, I. 2012. The role of vocabulary size in predicting performance on TOEFL reading item types. *System*, 40, 376-385.
- Alqahtani, M. 2015. The importance of vocabulary in language learning and how to be taught. *International Journal of Teaching and Education*, 3(3), 21-34.
- Barin, M. 2011. An evaluation of university level preparatory class students' feelings in terms of foreign language anxiety. *Journal of Social Sciences*, 11(47), 1-14.
- Basuki, Y., Damayanti, A. & Dewi, S. U. 2018. Vocabulary coursebook for EFL learners of higher education in Indonesia. *International Journal of Education and Literacy Studies*, 6(4), 122-128.

- Beglar, D. 2010. A Rasch-based validation of the vocabulary size test. *Language Testing*, 27(1), 101-18.
- Daller, H., Milton, J. & Treffers-Daller, J. 2007. *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.
- Hananto. 2013. Government-funded computer-based vocabulary project in Indonesia. *Asia-Pacific Collaborative Education Journal*, 9(1), 15-25.
- Ishii, T. & Schmitt, N. 2009. Developing an integrated diagnostic test of vocabulary size and depth. *RELC Journal*, 40(5).
- Karami, H. 2012. The development and validation of a bilingual version of the vocabulary size test. *RELC Journal*, 43(1), 53-67.
- Mustafa, F. 2019. English vocabulary size of Indonesian high school graduates: Curriculum expectation and reality. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 3(2), 357-371.
- Nation, I. S. P. 2006. How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.
- Nation, I. S. P. 2012. *The vocabulary size test*. Retrieved from <https://www.wgtn.ac.nz/lals/about/staff/Publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf>.
- Nation, I.S.P. & Beglar, D. 2007. A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Qian, D.D. 2002. Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning*, 52, 513-536.
- Schmitt, N. 2000. *Vocabulary in language teaching*. Ernst Klett Sprachen.
- Schmitt, N., Schmitt, D. & Clapham, C. 2001. Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing* 2001, 18(1), 55-88.
- Sudarman, S. & Chinokul, S. 2018. The English vocabulary size and level of English department students at Kutai Kartanegara University. *Eternal: English, Teaching, Learning, and Research Journal*, 4(1), 1-15.