Measuring EFL Students` Vocabulary Size: Why and How

Iwan Kurniawan

IAIN Raden Intan Lampung
Email: Iwan_stain@yahoo.com

Abstract. *Students` vocabulary size plays very important role toward their success in learning new language. To succeed in language learning 2,000 word families is often mentioned. Having this number, a reader can understand nine from ten words in many written text. Right now, researchers propose number at least 3000 word families and mastering minimum 3000 word families will support students` success in academic reading at university level. To pass tests such as Cambridge First Certificate, TOEFL or IELTS. Vocabulary of 5,000 words are needed. Undergraduate non-native speaker of non-European countries could study successfully in English speaking countries by having word families around 5,000-6,000. A non-native speaker can study PhD level well with vocabulary about 9,000. This article also will discuss the techniques to estimate someone vocabulary size.*

Key words: *vocabulary size, word family, word frequency level*

A. INTRODUCTION

Vocabulary plays very vital role in language learning. It can be seen from advice from David Wilkins, who says ` without grammar very little can be conveyed, without vocabulary nothing can be conveyed. It means if a student spends most time in learning grammar, his or her English will get a little progress but his or her will improve a lot if they learn vocabulary or words (Thonbury, 2002). Having limited vocabulary is also assumed will be barrier for them from learning a foreign language (Subekti and Lawson, 2007). Vocabulary, actually, is the heart of language. In using language, everything is done by words. We listen to words, we speak words, we read words, and we write words.

Another important role of vocabulary is also expressed by Richard and Renandya. They mentioned vocabulary is central part of language proficiency and students with less vocabulary ownership will not be able to use language maximally. Richard and Renandya (2002) state:

> "Vocabulary is a core component of language proficiency and provides much of the basis for how well learners speak, listen, read, and write. Without an extensive vocabulary and strategies for acquiring new vocabulary, learners often achieve less than their potential and may be discouraged from making use of language learning opportunities around them such as listening to radio, listening native speakers, using the language in different context, reading, or watching TV."

From some arguments above, it cannot be denied that the role of vocabulary is very crucial in language teaching and learning. However, the teaching of vocabulary had been for a long time neglected especially in year of 1950s and 1960s when Audio-Lingual dominated language teaching methodology. It happened because at that time language teaching approaches emphasized on grammar and phonology. The Audio Linguist believed the teaching of language should be emphasized on basic grammatical acquisition and after that the building of large vocabularies could be done later easily. Hocket was one of the most famous linguist at that 1950s era argued that vocabulary was the easiest aspect, so it hardly lured formal attention in classroom teaching. Many language teachers attached the teaching of vocabulary as a small part in reading instruction.

In contrast, by decade of 1970s and 1980s, the teaching and learning of vocabulary has developed well. One reason of its resurgence was the present of computer that can provide large amount of information for words analysis. Besides, in this period the concept of communicative language teaching triggered the teaching and learning of vocabulary and partly the natural approach influence. The supporters of this view believed at the early stage of language acquisition, vocabulary comes first

and then grammar. Rivers in the decade of 1980s gave his argument that a learner needs to acquire an adequate number of vocabulary to be able in communication. In addition, Rivers mentioned that the vocabulary acquisition is never slowed down by age and it is very different compared with other aspect of language especially pronunciation. The more mature someone is, the easier for him or her to add vocabulary. It happens likely because an adult has richer knowledge of the world to use those words. In the recent days, there has been an agreement that development of adequate vocabulary is essential for language learning (Nunan, 1991). From some explanations above, the notion that vocabulary plays very strategic position in language learning is unquestionable, so in the recent years the teaching of language has included vocabulary as main component. Vocabulary is not any more treated as additional material and syllabus designers have put it as basic ability that must be acquired at the beginning level.

Another strong reason to start learning language by words is the argument from some researchers who say that learners at beginning level should be taught a huge number of productive words at least 2000 high frequency words as a threshold level. If a learner has words less than that number, he or she will face difficulties in listening and speaking. In contrast, if they can acquire that number, they can express idea because those 2000 words are often used by native speaker in daily conversation. This number also will enable readers to understand 9 from 10 written texts in English. Another opinion says that 2000 basic words are actually not enough for certain group to succeed. For example, university students need further at least 1000 words high frequency words. So, the whole number will be 3000 high frequency words (Murcia, 2001). From explanations above, it can be concluded that having adequate number of vocabulary is a must and to know how many words someone has, definitely, there must be a test. This article will show the ideal number of vocabulary someone should have, the reasons why to assess someone vocabulary and how to count some vocabulary size.

B. DISCUSSION

a. The Nature of Vocabulary

There are many definitions of vocabulary and in this study; the definition will be quoted from Richard and Campillo. According to Richard (2010), vocabulary is a set of lexemes that can be in the form of single words, compound words, and idioms. Campillo (2015) stated that vocabulary is the total number of words. From two definitions above a conclusion can be made that a vocabulary is total number of words that someone has. These words can be single words, compound words or idioms.

Word is basic unit of language that someone must master from infant age to adult person to communicate. It is important because words are the foundation of a language where bigger structures such as phrases, clauses, sentences, paragraphs, and texts are formed. After having adequate number of vocabulary, a person must know the rules to arrange those words into a meaningful sentence. This rule is called grammar. If someone does not know how to put the words in the correct place in sentence, the vocabulary will be useless. As mentioned before, vocabulary is the heart of language if someone wants to learn a new language he or she must have sufficient number of vocabulary otherwise they will face problem when to hear, to speak, to read, and to write.

Words have two forms: oral or written. **Oral vocabulary** is the word that we know and utilize in listening and speaking. In contrast, **written vocabulary** includes the words that we recognize and use in speaking and writing. Word knowledge also can be divided into two forms: Receptive and productive. **Receptive words** are words that we know when we see or hear and **Productive words** are words when we speak or write. Receptive words are bigger than productive ones.

Words also can be divided into two parts that is function word and content word. **Function word**s play role more in grammar. They have limited meaning and serve

to connect sentences, modify meaning of other words. Content words have many information. For example: article, preposition, pronoun, conjunction, auxiliary, etc. On the other hand, **content words** have meaning. The words that are included in this type are noun, full verb, adjective, and adverb. When we prepare vocabulary size test, the focus will be in this kind of words.

Content words have some varieties of form. For example, we have word *wait* but we also have *waits, waited, waiting*. They are considered different form but still in the same word. There is addition of inflectional endings without changing the meaning and word class of the base. The base and inflectional element is called **lemma** in vocabulary study. However, there are also derivative elements which often change the word classes and meanings. For example: *possible, possibility, possibly, impossible.* All the words are related both in form and meaning. A set of words which has similar form and meaning is called as **word family.** A word family consists of the base plus its inflexions and derivatives.

Words also have classes that often called as **part of speech**. According to Nelson (2001) there are some parts of speech:

- Noun

  This part of speech is words that are used to name person, thing, animal, place, idea, or event. For example: *Camel* is animal that lives in desert.

- Verb

  It is one of the most crucial part of speech and without it, a sentence will never exist. This word that points out an action both physic and mental or to show the status of subject in a sentence by *to be* such as am, is, are, was and were. For example: I *jog* every morning to keep my body fit.

- Adjective

  This part of speech is used to modify a noun. An adjective specifies quality, size and number of nouns. For example: Anita Saraswati is a *graceful* dancer

- Adverb

  Adverb has almost the same function with adjective that is to modify other words. Adverb modifies verb, adjective, and another adverb. Commonly adverbs are formed by adding *ly* after adjective such as *gracefully, calmly, quickly.* For example: The choirs sing *slowly*

- Preposition

  This kind of part of speech has function to show time and place. The words that belong to preposition such as above*, below, outside, on, at, in, near, nearby, since, bottom, up, under, inside, out,* etc. For example: The campaign will start at 10 a.m.

- Conjunction

  These words are used to link phrases and clauses. There are two types of conjunctions:

  a. Coordinating conjunction or coordinator to connect elements those are equal. The main coordinator are *and, but*, and *or*. For example: Simon is handsome and wealthy.

  b. Subordinating Conjunction or subordinator introduces a subordinator clause. The main subordinator such as *although, after, before, if, since, as long as, as soon as, in order that, provided that*, etc. For example: *Because he is very tired*, John returns home earlier.

b. New perspective in Vocabulary Test.

Conventional view used more the **Discrete-point approach** in language testing. It is employed to assess learner knowledge of particular element of language such as word meaning, tenses, sentence pattern, sound contrast, speaking fluency, etc. However, in the last thirty years of 20th century, there is a new effort to adopt **the communicative approach** to language assessment. The tests are based on doing of simulating communication activities which are related to outside of classroom. A learner might be asked to write a letter of job application, to show specific detail information from paragraph, to discuss in scholarship interview. The test-takers will be judged on how far they can meet language demands of the test (Read, 2000). The test should be able to draw a conclusion about students` **language ability** that is **language knowledge** and **strategic competence.** Learners need to know vocabulary, spelling, pronunciation, grammar of the language but they must be proficient in implementing that knowledge for purposes of communication.

Read (2000) later on discusses about some procedures that a test designer must know. These procedures, he calls as three dimension of vocabulary assessment which consists of **Discrete-embedded, selective - comprehensive, context – independent - context- dependent.**

Discrete procedure tries to measure learner vocabulary knowledge as different construction from other language competences. In contrast, embedded vocabulary test estimates bigger structure. For example, asking the students to write university admission procedures improvement. In this point, the construction is academic writing ability and vocabulary scale will be one of five rating which form overall measure. A conventional test is designed on selected words basis and test taker are assessed on how good they show their knowledge of the meaning or use of those word. It is named a selective vocabulary measure. On contrast, comprehensive vocabulary test covers all vocabulary content of spoken or written text. The last dimension is context – independent - context- dependent. For a long time, context

in vocabulary test has been a discussion. Contextualization is a word presented to test takers in a sentence rather than in isolation. In present perspective, the idea of context includes entire text or discourse. The issue is how the test takers are assessed to engage with context of the test or in isolation. The test takers must use information in context to answer the question or just respond the questions in isolated way.

c.   Number of words that someone should have

An educated native speaker will have around 20,000 word families. This number is result of addition 1000 words each year after having 5000 words in the age 5 years. Most adult non-native speaker is considered lucky to have 5000 word families after learning several years. There has been a formal calculation that a non-native speaker needs 18 years to have the same level of a native learner`s word acquisition in one year. It occurs because of exposure or need than aptitude. A person who just travels two weeks to England definitely needs different demand compared to a student who wants to study in England. A research done in New Zealand indicated that native speaker of English get around a 1000 words family a year after three years old so that in 13 years old, a kid has about 10, 500 words family. For a 17 years old teenager possesses around 14000 or 15000 words family. For non-native speaker of English, he or she is really suggested to have around 1000 words family a year to gain 8000 to 9000 words for reading novels and newspapers.

There is a big question about how many words a learner needs to know that can function in all situations. A figure that often mentioned is 2,000 word families. This number is the words which often used by English native speaker in daily conversation. Knowing this number, a reader can understand nine from ten words in many written text. However, there are many strong arguments say that 2,000 is a minimum or threshold level is not enough. Recently, researchers propose number at least 3000 word families. According to Richard and Renandya (2002) mastering minimum 3000 word families will support students` success in academic reading at

university level. Besides, to pass a specific test such as Cambridge First Certificate, TOEFL or IELTS, a working vocabulary of 5,000 word families are needed (Thonbury, 2002). Undergraduate non-native speaker of non-European countries could study successfully in English speaking countries by having word families around 5,000-6,000. A non-native speaker can study PhD level well with vocabulary about 9,000.

To motivate students increase their vocabularies and as description for them to know the meaning of their vocabularies size, Nation (2012) shows three main frequency levels of high frequency, mid frequency, and low-frequency words as follows:

Tabel 1. Frequency Level

Level	1000 word family lists
High Frequency	1000-2000
Mid Frequency	3000-9000
Low Frequency	10,000 on

d. Why test students` vocabulary

The purpose to test students` vocabulary is to know how effective our teaching is. Without a test, we never know the progress of our students in acquiring English vocabulary, so it can show the students` learning progress and motivation. A teacher can improve his teaching by choosing a better strategy or add more unknown words to be memorized by students. Teacher also can predict students` progress and diagnose weakness in vocabulary learning etc. A test also has positive backwash that is if students know that they will be measured, they will learn more diligently. The test will push them to see back their vocabularies. The test result also can be a prediction to determine whether or not they are ready for study at certain level of education such as undergraduate or graduate level in English speaking countries. Moreover, test vocabulary result is also valuable for syllabus designer when planning a course for students. In reading class, syllabus designer can use the

research result to decide how reading programme is designed and taught. In higher level, vocabulary size research findings can also be used at college or university admission policy because the vocabulary level that students have will play important role to adjust with language demands for study such as reading text books, writing papers, listening lectures and classroom presentation.

e. How to prepare a test

To get the data of the test, a well-designed test definitely must be prepared. Nation (2013) mentions there are two kinds of methodology that can be used to measure vocabulary size. First method is ***Dictionary-based sampling***. This kind of method relies on a dictionary. The steps are as follows: the researcher finds out how many words there are in a dictionary. Then the sample of these words was determined, so the ration between sample and number of words in dictionary is known. The second method is ***Frequency-based sampling***. The person who has developed this way is Thorndike in the 1920s and 1930s economic depression. He developed word frequency list which then used to create vocabulary size test and the sampling words were taken from the high, medium and low frequency. Read (2000) mentions Dictionary-based sampling is commonly used to take estimate native speakers` vocabulary, whereas for EFL learners, frequency-based sampling is normally employed.

The production of Dictionary-based sampling test will be based on the theory of Thonbury (2002). According to him, one way to measure students` vocabularies is by using a dictionary and choosing random words – say every tenth word on every tenth page – and join them to be a test. The test could be a multiple choice and students are simply asked to choose the translation of a word in four options. The number of words correctly chosen will represent the proportion of words in dictionary. For example, a student knows thirty from a hundred words randomly selected (30%), and there are 10,000 headwords in the dictionary, so a rough estimation of a student`s vocabulary size is 30% of 10,000 or 3,000 words.

To set the vocabulary size test based on word frequency, researcher will use list of high frequency words such as the ***teacher`s word book*** or the ***General Service List*** as the basis of sampling that consist of words from 2000 to 3000. Estimating EFL students` vocabulary size is less problematic than native speakers because they have limited words and the vocabulary that they know is also generally from high frequency. The production of test can be done by taking 5% from 2000 words, for example. There are 100 words that later chosen randomly from the list to be items in the test. If a student can choose correct answer sixty from one hundred and then multiplied 2000, so he or she will get score 1,200. It means the student`s vocabulary size is 1, 200 words.

There are some steps that a vocabulary test designer must take care because this kind of test is discrete. ***First,*** the test user must decide the goal of the test. The test can be used to measure first language or second language. The test measures written receptive vocabulary knowledge. ***Second,*** test user must prepare the test specification that is sampling words for the items. The items in the test are taken from each word frequency levels without bias to any level. The word frequency is from word families in the British National Corpus. Only a small number of each level that can be used as sample. ***Third,*** the stem of the test has four multiple choice form and followed by non-defining sentence that includes the word in it. The test is available in the form of monolingual and bilingual. In this case, bilingual test will be used. The test takers must choose translation of a word in four options in their first language that is Indonesian. The test will be in hardcopy and bilingual form. For example:

.

1. Soldier. John is a **soldier**.
   a. Siswa   c. Prajurit
   b. Tukang Sol Sepatu   d. Bisnismen

Fourth, the order of the test should be mixed all level of the word frequency and avoid arranging in frequency order. The reason is the frequency level may trigger the test takers surrender at the next level, so mixing high frequency words in entire test is a good idea.

G. Type of Vocabulary Size Test

According to Izzura (2014) and Read (2000) there are some well-known vocabulary size tests that used around the words. The first one is *Vocabulary Level Test* (VLT). This test is proposed by Paul nation and based on word frequency. This test estimates vocabulary size from number of words correctly selected by students taken from five different frequency levels. The frequency levels range from the most frequent to the least frequent and grouped in the bands of 1000 words: The first 2000 words, 3000 words, 5000 words, the university level and 10,000 words. According to Paul nation, the 2000 and 3000 word levels consist of high frequency words that function effectively in English. For example, speaking in daily life or reading easy texts. The 5000 words are upper high frequency words that are important for class activities. For university level words, they will help students to read textbooks and other academic reading materials. Finally, 10,000 words level is the seldom words in use.

The second famous test is *The Eurocentres Vocabulary Size Test* (EVST). The test is computerized and uses ten frequency bands of 1000 words. The test uses the lexical decision paradigm and consists of 150 items. 2/3 items are real words and 1/3 are invented non-words. The third is *the Lexical Test for advanced Learners of English* (LexTALE). This test is based on EVTS and includes 60 items and test takers have express whether they understand or not the words. It is still new recognized test and has been validated for rather high proficiency levels. This test also uses word frequency as the criteria in item production.

The last one is TOEFL. The main function of the TOEFL is to assess foreign students who plan to study in countries where English is tool of instruction. The test is to know whether students have sufficient level of ability to study at academic background at American colleges or universities.

C. CONCLUSION

From explanation above, it is known how important to have information about vocabulary size. This article informs the benefits of having adequate number of vocabulary size both for teachers and for students. It also discusses the steps that someone should pass through to design a good test to measure learners` vocabulary size. Besides, theories about vocabularies and kinds of vocabularies tests are also provided.

D. REFERENCES

Arikunto, Suharsimi. 2010. *Prosedur Penelitian Suatu Pendekatan Prakte*k. Jakarta: PT. Rineka Cipta

Campillo, Rosa M Lopez. 2015. *Teaching and learning Vocabulary: An introduction for English Student.* Dialnet, http://www.dialnet. Unirioja.es/descarga. Accessed on February 15, 2015.

Izura, Cristina, et all. *Lextale-esp: A test to rapidly and efficiently assess the Spanish vacobulary size.* Internet Resources. www.uv.es/revispsi/articulos1.14/3izura.pdf (14th March 2016)

Kasiram, Moh. 1994. *Teknik Analisis Item*. Surabaya: Usaha Nasional.

Murcia, Marianne Celce. 2001. *Teaching English as a Second or Foreign Language.* Heinle and Heinle Thomsong Learning.

Nation, Paul. *The Vocabulary Test*. Internet Resources. www.uvictoria.c.nz/.../vocabulary_size test_Information.

Nelson, Gerald. 2001. *English, an Essential Grammar*. New York: Routledge.

Nunan, David. 1991. *Language Teaching Methodology. A Text Book for Teachers*. New York: Prentice Hall.

Read, John. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Richard, Jack C and Renandya, Willy A. 2002. *Methodology in Language Teaching*. England: Cambridge Univerity Press

Richard, Jack C and Smith, Richard. 2010. *Longman Dictionary Language Teaching and Applied Linguistic*. England: Longman.

Subekti, Nanang Bagus and Lawson, Michael J. 2007. *Vocabulary acquisition strategies of Indonesia postgraduate students through reading.* International Education Journal 8 (2). Shanon Research Press.

Thornbury, Scott. 2002. *How to Teach Vocabulary*. Essex: Pearson Education Limited.