

# VOCABULARY SIZE AND LEXICAL FREQUENCY PROFILE USED BY EFL UNIVERSITY STUDENTS IN ESSAY WRITING

*by Aridah Aridah*

---

**Submission date:** 14-Jan-2022 07:15AM (UTC+0700)

**Submission ID:** 1741393222

**File name:** AND\_LEXICAL\_FREQUENCY\_PROFILE-LLTC\_2017\_Proceedings-150-158.pdf (563.05K)

**Word count:** 4156

**Character count:** 20985

**VOCABULARY SIZE AND LEXICAL FREQUENCY PROFILE USED BY EFL UNIVERSITY STUDENTS IN ESSAY WRITING****Khairul Shalih and Aridah**Mulawarman University Language Center  
shalih.khairul@gmail.com and aridah27@yahoo.com**Abstract**

By having knowledge about vocabulary size of the students, teachers may provide materials based on students' vocabulary size and may help them increase their vocabulary size when they are exposed to vocabulary items in slightly higher level than what they have possessed (Laufer and Nation, 1999). This study intends to determine the vocabulary size and lexical frequency profile (LFP) used by EFL university students. Forty-three students participated in this study and the results showed that the average vocabulary size of the students was 7402 word families and their LFP showed that 89.21% of the 1<sup>st</sup> 1000 word list were still used by the students while the ideal percentage was 75%. The finding also showed that there was no significant positive correlation between VST score and the LFP percentage. This infers that students preferred using common words to using advanced words even though they have enough knowledge of advanced words.

Keywords: vocabulary size, lexical frequency profile

**Introduction**

Having a bigger size of vocabulary may be an advantage when it comes to perceiving and producing foreign language. Although vocabulary alone might not be enough to determine a person's ability in mastering a language, it is undeniable that vocabulary is one of the important aspects in mastering a language (Nation, 2014). In order to produce something that makes sense, a person needs to understand how to perceive and how to use the language. A person will not be able to produce something beyond his understanding. Thus, the vocabularies used in the writing of a person display the understanding of that person in language mastery. In addition, vocabulary size is related to language skills in general. It is believed that vocabulary size specifically affects the writing skill of a language user (Laufer & Nation, 1995). This is due to the fact that having larger vocabulary size improves the variety of vocabularies used in the writing product.

Furthermore, as language users, both L2 and foreign language users are often considered having less proficiency than native speakers. However, it is possible for those non-native language learners to get closer to the level of native speakers if the learners keep improving the ability in English mastery (Hirsh & Nation, 1992). Since writing is one of the products in language which can be measured and kept as a record, it can be made as an effective instrument to measure the growth of language learners. When a language learner has an

intention to improve the ability in language mastery, that user may use writing as a method to practice the language. Even if it is done little by little, the learner can review the writing some other time.

The writing can also become a record which shows the development of the learner over time. The learner can see the gap between the oldest and the latest writing that a person has ever kept in record. In this way the learner can see how much s/he develops his/her ability in learning the language. For these reasons, it is clear that using writing as a method in language learning is beneficial for non-native language learners in order to increase their vocabulary size and proficiency in writing.

### **Vocabulary Size and Lexical Frequency Profile**

Vocabulary refers to all the words that a person knows or uses. Vocabulary is defined as the stock of words used by or known to a particular people or group of people (Oxford Advanced Learner's Dictionary, 2000). In other words, vocabulary is an element of language which is used to express or to understand the intention of people.

Vocabulary is not usually being studied for the purpose of simply knowing the vocabulary itself, but rather in order to be used by the language user in the forms of communicating or understanding the content of language. An important aim of learning vocabulary is to bring the vocabulary knowledge of students into communicative use, where students are in a situation where there are demands upon them to make use of what they know (Laufer & Nation, 1995).

The number of words mastered by is different for every person. However, there is a standard of how many words a person needs in order to communicate or at least understand when the language is applied in daily context. The number of words known by each person is the vocabulary size that he/she possesses.

In order to measure the vocabulary size of a person in number, using word frequency is commonly used by language user. Word frequency is how often a word is used on a page of written English. For example, the word "the" is a word which frequently used in a written article. According to data in Nation (1990), the word "the" is one of the words frequently used in written English, which makes it very common in the view of word frequency. Related to this, in order to measure the vocabulary accurately, this study also used the word family system.

Word family is a group of words which include the root word along with its variations. The word "demolish" for example, includes the word "demolition", "demolishment", and "demolisher". All these words are included in one word family. Although there was no exact standard of how many word family should be known by an English native speaker, there were other studies conducted by Dupuy (1974) and Nation and Goulden (1990) which stated that there were a total of around 54.000 word families existed in the 3<sup>rd</sup> edition of Webster's Dictionary.

Nowadays, the method of using word families as the measurement unit of vocabulary size has become common and widely used. Nation and Beglar (2007) published the standard of word families needed in order to understand certain kinds of text, which was featured by many other studies at the present.

2  
Table 1: Vocabulary size needed to get 98% coverage (including proper nouns) of various kinds of text

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%

Source: Nation & Beglar (2006) in JALT (2007)

By looking at Table 1, it could be concluded that in order to understand necessary language directed for children or teenagers, a language user would need to know around 6,000 word families, while in order to understand newspaper or novel requires the user to recognize around 8,000 to 9,000 word families. By this data, it is also concluded that the standard of vocabulary size an English language user should have in order to use the language in daily conversation is around 7,000 word families or more.

Moreover, the standard for non-native language learners in undergraduate student level would be 5000-6000 word families (Nation & Beglar, 2007). In this standard, it could be interpreted that a vocabulary size level of at least 5,000 word families is required for language learners in order to cope with the lesson in the education of non-native university environment.

Lexical Frequency Profile (LFP) is the amount of word family produced in a set of composition, usually in form of percentages. Laufer and Nation (1995) stated that LFP shows the percentage of vocabularies used by language users at different vocabulary frequency levels in writing. These frequency levels are divided based on the category used in a corpus. The categories listed by Laufer and Nation (1995) were divided into four categories, which were 1<sup>st</sup> 1,000, 2<sup>nd</sup> 1,000, University Word List (UWL), and Not in List. Words included in the 1<sup>st</sup> 1,000 are the most common used words used by language users, while 2<sup>nd</sup> 1,000 includes less frequent words which are not included in the 1<sup>st</sup> category. UWL, now known as Academic Word List (AWL), includes the least frequent words used by language users. Not in list category includes words which are not included in the previous categories, proper nouns, for example.

To interpret LFP for an intermediate language learner, Laufer and Nation (1995) conducted a study and set the standard of LFP taken from essays. From an essay of 200 word families, 150 were in the 1<sup>st</sup> 1000, 20 were in 2<sup>nd</sup> 1000, 20 were in AWL, and 10 were in N category. After being converted to percentage, the composition had 75% - 10% - 10% - 5% accordingly. In this standard composition, 1<sup>st</sup> 1000 had the highest value because it had the most words in its list compared to other categories, while N category had the lowest percentage simply because it only consisted of unstandardized words, such as proper noun or foreign words.

Since LFP describes the variety of vocabularies used by language users in writing essays, it is believed to have a close relation to the vocabulary size of language users. Based on this belief, the main objectives of this study were to determine the vocabulary size and LFP of EFL university students, and to

9  
determine the correlation between the vocabulary size and LFP of EFL university students.

### Method

This study applied correlation design. This correlational study was conducted in English department of Teaching and Education Faculty of Mulawarman University. There were 43 students taken as the sample of study, which were taken from two classes of the fourth semester students. All participants had taken writing class in their respective classes; therefore they already had the basic knowledge of how to write a proper essay.

4  
Before writing the essay required for this study, all participants were given a Vocabulary Size Test (VST) based on the guideline of vocabulary size test by Nation (2012). The test was 14.000 word family VST, which was given in order to determine the vocabulary size of the samples. The results of the vocabulary level test were scored by matching with the answer key. There were 140 items in the test and they were graded according to the method suggested by *The Vocabulary Size Test Guideline* (Nation, 2012). According to the guide, the total score of each test was multiplied by 100, thus the maximum score were 14.000 word families. Based on the result of VST, the vocabulary size of the students would be determined and analyzed.

After the test had been administered, each participant had to write an essay of maximum 500 words. The topic of the essay was similar to the topic they were used to in writing essay in their writing class in university. By using a similar topic, they already have enough background knowledge to write the essay, thus this study was able to focus on the vocabulary aspect alone in analyzing the essay (Laufer and Nation, 1995). To calculate the LFP of each essay, an automatic word counter application from [www.lex tutor.ca](http://www.lex tutor.ca) was used. The application automatically revealed the LFP percentage of every essay inputted to the application.

After the data of both vocabulary size and LFP had been obtained, the data were sorted and analyzed by using *Pearson Product Moment* in order to determine the correlation between these two variables. If the result of the analysis shows that there was a significant positive correlation between the two variables, then the alternative hypothesis ( $H_a$ ) would be accepted. On the contrary, if the result of the analysis shows that there was no significant positive correlation between the two variables, then alternative hypothesis ( $H_a$ ) would be rejected and null hypothesis ( $H_0$ ) would be accepted.

### 8 Findings and Discussion

As mentioned above, the objectives of this study were to determine the vocabulary size level and LFP of EFL students, and to determine their correlation. Table 2 below shows the result of vocabulary size level from the sample.

Table 2: Result summary of vocabulary size test from the samples

Total Sample	Lowest Score	Highest Score	Average Score
43	5200	10000	7402

Source: Test result, 2016

Nation and Beglar (in JALT, 2007) concluded that the goal for English users was 8000 word family. However, as elaborated in chapter 2, initial studies concluded that non-native undergraduate students should have at least around 5000-6000 word families, while competent non-native doctoral students should have at least 9000 word families. The table above displayed that the average vocabulary size of the fourth semester students of English department in Mulawarman University were 7402 word families, which were still short on the ideal target of 8000 word families, but could be considered as adequate according to the standard.

To add more details to the result, the total participants who received scores according to the standard set by Nation and Beglar (2007) was listed in the following table:

Table 3: Amount of samples according to the VST score standard

Score Range	Total Participants	Lowest Score	Highest Score	% (of total participants)
> 8000	14	8100	10000	33
6000 – 8000	22	6100	7900	51
5000 – 6000	7	5200	5900	16

Source: Test result, 2016

As we could see from table 3, among all of the students, there were only 7 who were classified in the minimum range for undergraduate students. There were 22 students who achieved score adequately according to the standard, while there were 14 students who achieved ideal result for English learners. From the result described above, it could be concluded that 16% of the samples achieved the minimum result, while there were 51% of the samples who achieved adequate result. In addition, there were 33% of the students who surpassed the ideal target of score for English learners.

To analyze the LFP of the sample, the essays written by them were inputted to a word counter application. Table 4 below shows the summary of LFP result.

Table 4: Summary of LFP result

Value	LFP (Word Family)				LFP (Percentage)			
	1 <sup>st</sup>	2 <sup>nd</sup>	AWL	N	1 <sup>st</sup>	2 <sup>nd</sup>	AWL	N
<b>Minimum</b>	68	3	1	0+?	81.20	1.16	0.29	1.00
<b>Maximum</b>	142	20	22	0+?	97.38	8.80	6.25	10.20
<b>Average</b>	104	10	9	0+?	89.21	4.36	3.01	3.41

Source: Result of Data Analysis, 2016

The ideal composition for non-native learners should be close to the standard set by Laufer and Nation (1995) mentioned in the previous chapter. The ideal average percentage was to have the 1<sup>st</sup> 1000 category close to 75% and the

AWL category close to 10%. However the average result from the samples revealed that the average of LFP in 1<sup>st</sup> 1000 category was 89.21 while the other categories were less than 5%. From this result, it could be concluded that the students were using the most common words available in order to complete their essay and they occasionally used more complex words.

Table 5: The Result of *Pearson Product Moment*

	VST (x)	LFP (y)			
		1 <sup>st</sup> 1000	2 <sup>nd</sup> 1000	AWL	N
r Product Moment	-	0.27	-0.47	0.11	-0.02

Source: Analysis Result, 2016

The result of Pearson product moment showed that the  $r$  value for each category was 0.27, -0.47, 0.11, and -0.02 respectively. According to Santoso (2001) the result of this study could be interpreted that the score of VST and the value of LFP of the 1<sup>st</sup> 1000 category has a slight positive correlation, meaning that there was a low positive correlation between these two variables. Another variable which also got positive result was the AWL category, whose  $r$  value was interpreted as having a low positive correlation. On the other hand, the score of VST and the value of LFP of the 2<sup>nd</sup> 1000 category could be interpreted as a moderate negative correlation, and the N category could be interpreted as a low negative correlation.

From the result shown above, it was initially expected that the higher the score of vocabulary size test, the higher the percentage for more advanced word family category would be. However, the result had shown that the better the score of VST, the more words originated from the 1<sup>st</sup> 1000 category were used in the writing product. Taking account that the vocabulary range from this category is immensely larger than other categories since the 1<sup>st</sup> 1000 category holds the highest number of items and also the most common vocabulary items existed in dictionary, this result could still be considered as normal. On the contrary, the 2<sup>nd</sup> 1000 received negative correlation since more percentage of word family were used in its counterpart category. However, as for the most advanced word family category, the academic word list (AWL) had a positive correlation but the correlation was categorized as a low positive correlation. Even though the correlation was positive, but it was significant.

It can be concluded that achieving high score in vocabulary size test, which describes the vocabulary size level of a language user, might not always guarantee that the writing product contains vocabularies of advanced word family, as described in the lexical frequency profile. This conclusion was taken due to the usage of the 1<sup>st</sup> 1000 word family in the writing product correlated more to the VST score since this category had higher positive correlation, while the usage of AWL category also had positive correlation to VST score but was weaker than that of the 1<sup>st</sup> 1000 category. Moreover, the value of  $r$  table for the 43 students was 0.294. Therefore, the  $r$  value of the 1<sup>st</sup> 1000 and AWL categories, which both had positive value of 0.27 and 0.11, were both less than the value of  $r$  table and

they were not significant. Thus, the alternative hypothesis ( $H_a$ ) was rejected whereas the null hypothesis ( $H_0$ ) was accepted.

To identify the cause of this insignificant correlation, the result of this study was compared to the findings of a similar study conducted by Laufer and Nation (1995) with three groups of language users with different proficiency levels in English. One of the three groups had similar result to finding of this research. One of the three groups, the low intermediate level group was a class of students who studied in New Zealand but originated from countries which used English as foreign language. Furthermore, the group was still in the early period of their education in university, which made the group in similar situation to the samples of this research. The low-intermediate level group achieved 86.5% - 7.1% - 3.2% - 3.3% in their average LFP result.

Comparing the result of group from the previous study and the score of the students in this research, the average LFP (percentage) was 89.21% - 4.36% - 3.01% - 3.41%, which had higher value in the 1<sup>st</sup> 1000 and lower value in 2<sup>nd</sup> 1000, while the value in AWL and N category were relatively similar. Thus, it was speculated that the possible reasons which cause the rise of the 1<sup>st</sup> 1000 category and the decrease of other categories from the ideal result was the learning environment of the samples in the university. The participants studied in a country where English is used as a foreign language, so that complex English vocabularies were likely uncommon to be used in daily life. While in the classroom they might use English as the main language, they might not use English outside of the classroom as much as inside the classroom. The students might know words from categories other than the 1<sup>st</sup> 1000, but chose to use the simplest vocabulary items the most since it was assumed to be enough to convey the message in their essay.

Davies and Pearse (2000) in Hastuti (2015) explained that writing not only involves low-level skills of handwriting, spelling, constructing grammatical sentences, and punctuations; but also high-level cognitive skills of processing, selecting, organizing the ideas and information into rational sequence in the form of paragraphs, editing the draft, and writing the final product. Since there is a complex process which should be considered when making a writing product, it might be understandable for the writer to use the most comfortable vocabulary items for them. Because writing itself can be considered as a difficult process which involves lots of requirements, the students might not have the intention to use words which were not familiar to them, even if they knew the definition of the words.

In addition to the complex process in writing itself, other factor which could affect the result of this study was the lack of chance for the students to use more advanced words in real-time practice. Waring and Nation (2016) concluded that English learners should work on the strategies for low frequency words. If the learner intends to use English for academic purpose, then it is advised to learn general academic words, which mostly can be found in AWL category. Since there are more vocabulary items which are included in the 1<sup>st</sup> 1000 category, it is understandable for English learners to use words of this category more than the others. However, given the right environment and enough time, it is possible for



the learners to increase the use of vocabularies from other categories. The results from the sample groups of the research by Laufer and Nation (1995) could be a fine example that language practice had an immense effect on language learners.

While the result of LFP percentage of this research was similar to Group 1 in the previous research, Group 1 in the previous research was a freshmen group who came from countries which used English as foreign language. However, the environment of the sample group at that time was a university in a country where English was the main language, thus the use of English inside and outside of classroom environment was obligatory. The development of the vocabulary used in writing product could be seen from Group 3 in the previous research, where the samples from that group were students of the same university but in higher semester of their study.

The situational example above explained that learning environment could bring large contribution to the result of vocabularies used in the writing product. When language learners could take their time to use the language in daily basis, the quality of vocabulary they used would also improve.

Because the sample of this research was in the learning environment which did not require them to use English all the time, it could decrease their chance to use more advanced words in their language usage, whether spoken or written. Should the situation be a little bit different, or there was a special treatment to force them to interact more in English, there was a possibility that the result might be closer to the initial expectation.

### **Conclusion**

This study showed evidence that even though the students have a big vocabulary size; it does not mean that they will use them all when they are writing. Most of the students still prefer to use common vocabulary items listed in the first 1000 words, rather than in the other vocabulary list such as the second 1000 words or AWL category. This implies that the high score of vocabulary size level test does not always determine the high lexical frequency profile of the writing product.

### **References**

- Beglar, David. & Nation, I. S. P. 2007. *JALT Magazine – A Vocabulary Size Test*. Compass Publishing.
- Blanchard, Karen & Root, Christine. 1994. *Ready to Write: A First Composition Text*. Addison-Wesley Publishing Company.
- Clark, Vicki L. & Creswell, John W. 2015. *Understanding Research*. Pearson Education Inc.
- Cole, Jenny & Feng, Jay. 2015. *Effective Strategies for Improving Writing Skills of Elementary English Language Learners*. Chinese American Educational Research and Development Association Annual Conference.
- Hastuti, Saptin Dwi Setyo. 2015. *The Influence of Vocabulary and Grammar Mastery on The Students' Writing Skill at Yogyakarta State University*. Jurnal Media Wisata

- Hirsh, David. & Nation, I. S. P. 1992. *Reading in a Foreign Language*. Victoria University of Wellington.
- Laufer, Batia. & Nation, I. S. P. 1999. *Language Testing – A Vocabulary Size Test of Controlled Productive Ability*. University of Haifa Library.
- Laufer, Batia. & Nation, I. S. P. 1995. *Vocabulary Size and Use in L2 Productive Writing*. University of Haifa Library.
- Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. Cambridge University Press.
- Nation, I. S. P. 2012. *The Vocabulary Size Test*. Victoria University of Wellington.
- Nation, I. S. P. 2014. *What do You Need to Know to Learn a Foreign Language*. Victoria University of Wellington.
- Santoso, Singgih. 2015. *Menguasai SPSS 22*. Jakarta: Elexmedia Komputindo.
- Santoso, Singgih. 2001. *Statistik Non-Parametrik*. Jakarta: Elexmedia Komputindo.
- Sudijono, Anas. 2003. *Pengantar Statistik Pendidikan*. Jakarta: Raja Grafindo Persada.

# VOCABULARY SIZE AND LEXICAL FREQUENCY PROFILE USED BY EFL UNIVERSITY STUDENTS IN ESSAY WRITING

## ORIGINALITY REPORT

12%

SIMILARITY INDEX

11%

INTERNET SOURCES

6%

PUBLICATIONS

%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://mafiadoc.com">mafiadoc.com</a> Internet Source	3%
2	<a href="http://www.acarindex.com">www.acarindex.com</a> Internet Source	1%
3	<a href="http://www.neliti.com">www.neliti.com</a> Internet Source	1%
4	<a href="http://repository.usd.ac.id">repository.usd.ac.id</a> Internet Source	1%
5	I.S.P. Nation, Averil Coxhead. "Measuring Native-Speaker Vocabulary Size", John Benjamins Publishing Company, 2021 Publication	1%
6	<a href="http://dspace.tul.cz">dspace.tul.cz</a> Internet Source	<1%
7	<a href="http://studentsrepo.um.edu.my">studentsrepo.um.edu.my</a> Internet Source	<1%
8	<a href="http://www.scribd.com">www.scribd.com</a> Internet Source	<1%

9	<a href="http://digilib.iain-palangkaraya.ac.id">digilib.iain-palangkaraya.ac.id</a> Internet Source	<1 %
10	<a href="http://academic.oup.com">academic.oup.com</a> Internet Source	<1 %
11	<a href="http://mobt3ath.com">mobt3ath.com</a> Internet Source	<1 %
12	Titilope Omolara Olarewaju, Felix Oaikhena Idumah, Agatha Itohan Oseghale, Lucy Adeteju Orumwense et al. "chapter 9 Training Needs Assessment of Palm Oil Processors in Ijebu North Local Government Area, Ogun State", IGI Global, 2020 Publication	<1 %
13	Karami, H.. "The Development and Validation of a Bilingual Version of the Vocabulary Size Test", RELC Journal, 2012. Publication	<1 %
14	<a href="http://eprints.unisnu.ac.id">eprints.unisnu.ac.id</a> Internet Source	<1 %
15	<a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a> Internet Source	<1 %
16	S. Webb. "The Lexical Coverage of Movies", Applied Linguistics, 03/25/2009 Publication	<1 %
17	<a href="http://kansaigaidai.repo.nii.ac.jp">kansaigaidai.repo.nii.ac.jp</a> Internet Source	<1 %

18	<a href="http://repo.iain-tulungagung.ac.id">repo.iain-tulungagung.ac.id</a> Internet Source	<1 %
19	<a href="http://uwe-repository.worktribe.com">uwe-repository.worktribe.com</a> Internet Source	<1 %
20	<a href="http://mjltm.org">mjltm.org</a> Internet Source	<1 %
21	<a href="http://rdoc.univ-sba.dz">rdoc.univ-sba.dz</a> Internet Source	<1 %
22	<a href="http://tigerprints.clemson.edu">tigerprints.clemson.edu</a> Internet Source	<1 %
23	<a href="http://www.victoria.ac.nz">www.victoria.ac.nz</a> Internet Source	<1 %
24	<a href="http://eprints.soas.ac.uk">eprints.soas.ac.uk</a> Internet Source	<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On