LEXICAL RICHNESS IN INDONESIAN JUNIOR HIGH SCHOOL STUDENTS' WRITING PRODUCTION: A CORPUS-BASED STUDY

David Geba Abi Anandi[1], Fransiskus Xaverius Mukarto[2]
[1]davidgebaabianandi@gmail.com, [2]mukarto@usd.ac.id

SANATA DHARMA UNIVERSITY

ABSTRACT

In the 2013 curriculum, English is not a compulsory subject in primary schools. Students may still have relatively very limited productive vocabulary and they might use certain communicative strategy when they did not know the vocabulary to express themselves. Therefore, this present study aims to examine the lexical richness of junior high school students' writings on descriptive texts. The study focused on examining the lexical variation, lexical diversity, and lexical sophistication in the students' writing production. The study was a corpus-based one. The corpus consists of 18 descriptive texts written by junior high school students in Yogyakarta and analyzed using lextutor, a web concordance English v.9 available on the internet. The results of the analysis reveal that descriptive texts have low, but very likely proper, average ratio on lexical variation, lexical density, and lexical sophistication considering their English proficiency level. It should be noted that the students were those of grade 7.

Keywords: lexical richness, senior high school students, descriptive texts

A. INTRODUCTION

In second language acquisition, vocabulary is considered as important aspect of language to be learned. According to Ramos (2015), vocabulary is the aspect of language that serves as a building block for learners to begin acquiring a second language. Therefore, second language learning highly depends on vocabulary. Vocabulary takes part as an information carrier. Therefore, it plays an indispensable role of language (Zhai, 2016). Moreover, Zhai (2016) also states that vocabulary knowledge is viewed as an essential aspect for second language acquisition, as limited second language vocabulary would hinder successful communication.

Considering the importance of vocabulary in language learning, the idea of a vocabulary learning program is to bring learners' vocabulary knowledge into communicative use (Laufer and Nation, 1995). Learners are expected to use their language competence and knowledge into practical use. When learners are asked to make use of what they know, it is expected that the relationship between direct measures of vocabulary size of the learners and vocabulary richness in their language production are shown (Laufer and Nation, 1995). Hence, this study aims at exploring students' vocabulary knowledge by examining their vocabulary richness in their writing production.

Vocabulary learning is realized through productive tasks. Before looking at the statistics and how the vocabulary is calculated, it is helpful to consider what assumptions they make about effective vocabulary use (Read, 2000). Zhai (2016) states that vocabulary use assessment can be measured by examining the lexical richness. Lexical richness measures seek to quantify the extent to which writers use a diverse and broad vocabulary in their writings (Real et al., 2020). According to Read (2000), in measuring the lexical richness, there are several lexical features which need to be calculated, namely lexical variation or type-token ratio, lexical sophistication, lexical density, and numbers of errors.

1. Lexical Variation
Lexical variation refers to the type/token ratio, or the ratio between the different words in texts and the total number of consecutive words presented in percentage (Laufer and Nation, 1995). According to Read (2000), lexical variation is defined as the extent to which writers have vocabulary knowledge which allows them to avoid using repeated similar words or synonyms, super-ordinates and other kinds of related words. Furthermore, in writing assessment, this part is often called as range of expressions (Read, 2000).

2. Lexical Sophistication
According to Laufer and Nation (1995), lexical sophistication is the percentage of rare or advanced words in texts. Lexical sophistication or lexical rareness is widely approved as a central component of various lexical richness evaluation schemes (Azadnia, 2021).

3. Lexical Density
Including lexical density as a key component in evaluating lexical richness is based on the assumption that the use of more instances of content words facilitates the conveyance of a message denoting complex information through more sophisticated words (Azadnia, 2021). Density level is considered if it contains many vocabulary relative to the total number of words (Signes and Arroitia, 2015).

4. Numbers of Errors
Arnaud (1984), as cited in Read (2000) provides a list of typical errors in his study. They are minor spelling mistakes, major spelling mistakes, derivation mistakes, faux-amis (deceptive cognates), interference from another language on the curriculum and the confusion between two lexemes.

One way to assess students' vocabulary knowledge is by examining their lexical richness. Lexical richness is the quality of vocabulary used by someone in a language product (Malvern and Richards, 2013). Corpus can be used as a way to draw the descriptive pattern of language features on various sub-registers (Puspita, 2019).

Many corpora analysis studies have been undertaken by researchers. Some studies focus on analysing lexical richness of essays written by students. For instance, a study written by Ha (2019) focuses on identifying and explain how lexical richness manifest argumentative essays written by thirty-five undergraduates. Similar studies also focus on identifying the lexical richness on students' narratives, such as the study written by Siskova (2012). The study focuses on comparing the different of lexical richness measurements written by Czech EFL learners. Another example of related study is the study written by Azodi et al. (2014). They focus on measuring L2 lexical richness of productive vocabulary in the written production of Iranian EFL university students. As an addition, several studies only focus

only on some lexical features of lexical richness. For instance, a study written by Juanggo (2018) focuses on investigating the Indonesian EFL learners' lexical diversity and lexical sophistication of productive vocabulary in their written discourse.

The research question in this study is "What is the lexical richness level of junior high school students' written descriptive texts in terms of lexical variation, lexical density and lexical sophistication?". This study applies corpus-based study. Particularly, this present study aims at investigating the overall vocabulary profile of senior high school students' writing proficiency by using existing vocab-profiler tools. In order to do so, the researcher explain the lexical diversity, lexical variation, and lexical sophistication quantitatively.

B. METHOD

The present study focuses on examining the lexical richness of students' descriptive text writings. Hence, this study employed a corpus-based analysis. The corpus consists of 18 descriptive texts. The 18 students were from one of the junior high schools in Yogyakarta. The students had various level of English proficiency. The selection of the participants was because the students were willing to be the participants. The writing task was done by students as one of the assignments in English Subject. The students had to produce a descriptive text, describing their daily routine. The results were submitted to the teacher to be scored.

The data were gathered from students' works on descriptive texts. The works were in forms of images and word files. The texts were submitted as a writing assignment in the English subject. After the students submitted their works, the teacher gave the works to the researcher.

The data analysis technique consists of several steps. Firstly, since some texts were in forms of images, the researcher typed the texts. Secondly, the researcher uploaded every file to one of the text-analyser webs via www.lextutor.ca. After accessing the website, the writer clicked Vocabulary Profile button, then VP-Classic. Furthermore, the writer input the texts into the column and clicked 'SUBMIT_Window' button. Below are the descriptions of the lexical richness measured in this study:

Table 1. Lexical Richness Measured

Tool	Type	Measure
Lextutor.ca	Lexical Variation	The proportion the different words in the text and the total number of words found in texts
	Density	The proportion of content words to the total number of words found in texts
	Sophistication	The proportion of words found at different frequency levels, in terms of K1 Words (1-1000), K2 Words (1001-2000), and AWL Words

As additional data, the researcher also analyzed the part of speech produced by the students via https://parts-of-speech.info/. The website helped the researcher to show the percentages of each of the parts of speech written by the students. It is done to see how variative their writings are.

C. FINDINGS AND DISCUSSION

This part discusses the findings of this study. The research question in this study is "What is the lexical richness level of junior high school students' written descriptive texts in terms of lexical variation, lexical density and lexical sophistication?". This part, thus, focuses on categorizing the lexical richness found in the descriptive texts produced by the junior high school students.

Lexical richness refers to the vocabulary used by someone in a discourse. It reflects the ability and skills in manipulating the basic units of speech (Failasofah and Alkhrishes, 2018). The researcher input each of the texts into a text analyser website called lextutor.ca. The following table is the results of the findings:

Table 2. The descriptions of the students' vocabulary production

No.	Students	Total words	Types
1	Text 1	134	73
2	Text 2	44	35
3	Text 3	100	56
4	Text 4	130	79
5	Text 5	127	61
6	Text 6	143	52
7	Text 7	114	68
8	Text 8	146	75
9	Text 9	75	43
10	Text 10	90	38
11	Text 11	66	37
12	Text 12	100	60
13	Text 13	235	102
14	Text 14	241	87
15	Text 15	73	44
16	Text 16	115	63
17	Text 17	121	67
18	Text 18	135	67
	Total	**2189**	**1107**
	Average	**121.61**	**350**
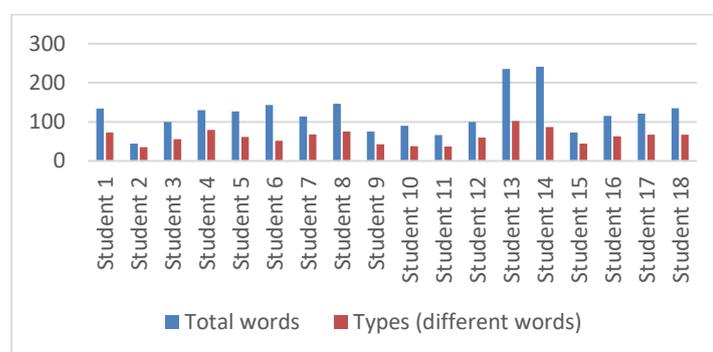
Figure 1. The depiction of the students' vocabulary production

The table and figure above show that the total words produced by the students are 2189 with average of 121.61. In terms of types of words, the number produced are 1107 with the

average of 350. It can be seen that student 14 produced the highest number of words with 241 words. However, student 14 has lower number of different words (types) than student 13. Student 13 has the highest number of different words (types) with 102 types of words.

1. Lexical Variation

Lexical variation is defined as a variety of different words rather than a limited number of repeated words (Read, 2000). The measures applied in this case are the type-token ratio. Below is the result of the calculation of the TTR (type-token ratio) produced by the students:

Table 3. The descriptions of the students' lexical density

No.	Participants	D (lexical variation) type-token ratio	
		Decimal	**Percentage**
1	Text 1	0.54	54%
2	Text 2	0.80	80%
3	Text 3	0.56	56%
4	Text 4	0.61	61%
5	Text 5	0.48	48%
6	Text 6	0.36	36%
7	Text 7	0.60	60%
8	Text 8	0.51	51%
9	Text 9	0.57	57%
10	Text 10	0.42	42%
11	Text 11	0.56	56%
12	Text 12	0.60	60%
13	Text 13	0.43	43%
14	Text 14	0.36	36%
15	Text 15	0.60	60%
16	Text 16	0.55	55%
17	Text 17	0.55	55%
18	Text 18	0.50	50%
	Average	**0.16**	**53%**
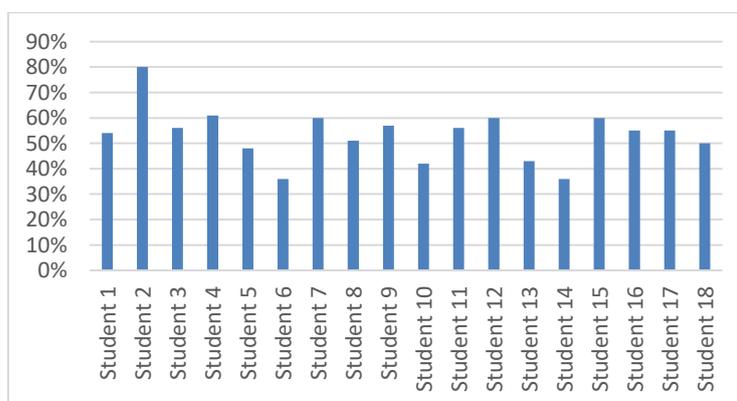
Figure 2. The Depiction of Lexical Variation in Students' Writing Production

The table and figure above show that the average ratio of lexical variation is 53%. From the table and figure above, it can be seen that student 2 has the highest lexical variation with

80%. On the other hand, students 6 and 14 have the lowest ratio of lexical variation with a score of 36%.

2. Lexical Density
Another criterion in measuring lexical richness is lexical density. Lexical density refers to the proportion of lexical or content words. Below is the result of the lexical density calculation:

Table 4. The descriptions of the students' lexical density

No.	Students	Lexical density	
		Decimal	**Percentage**
1	Text 1	0.51	51%
2	Text 2	0.32	32%
3	Text 3	0.45	45%
4	Text 4	0.43	43%
5	Text 5	0.46	46%
6	Text 6	0.53	53%
7	Text 7	0.61	61%
8	Text 8	0.48	48%
9	Text 9	0.56	56%
10	Text 10	0.51	51%
11	Text 11	0.42	42%
12	Text 12	0.45	45%
13	Text 13	0.51	51%
14	Text 14	0.46	46%
15	Text 15	0.45	45%
16	Text 16	0.44	44%
17	Text 17	0.49	49%
18	Text 18	0.39	39%
	Average	**0.48**	**47%**
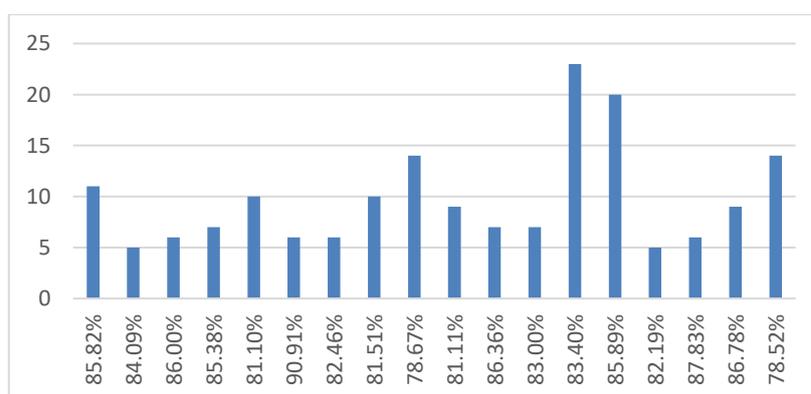
Figure 3. The Depiction of the Lexical Density in Students' Writing Production

Based on the findings, it is shown the average lexical density is 47%. Student 7 has the highest ratio of lexical density with 61%. The table and figure above also reveal that student 2, exceptionally, produced the least density value with 32%. It is also found that 6 students have a density production value of more than 50%, while the rest of the students have lower

than 50% value of lexical density. Theoretically, the highest value of lexical density is 120 (range from 0-120) (Failasofah & Alkhrishes, 2018). It can be inferred that the lexical density production of the students is on average low as the average of lexical density is 47%.

3. Lexical Sophistication

Lexical sophistication is defined as the numbers and percentages of advanced words found in a discourse. The results are presented in the following table:

Table 5. The descriptions of the students' lexical density

No	Participants	K1 Words (1-1000)		K2 Words (1001-2000)		AWL Words	
		Word	Percentage	Word	Percentage	Word	Percentage
1	Text 1	115	85.82%	11	8.21%	2	1,49%
2	Text 2	37	84.09%	5	11.36%	0	0%
3	Text 3	86	86.00%	6	6.00%	4	4.00%
4	Text 4	111	85.38%	7	5.38%	5	3.85%
5	Text 5	103	81.10%	10	7.87%	1	0.79%
6	Text 6	130	90.91%	6	4.20%	0	0%
7	Text 7	94	82.46%	6	5.26%	0	0%
8	Text 8	119	81.51%	10	6.85%	8	5.48%
9	Text 9	59	78.67%	14	18.67%	0	0%
10	Text 10	73	81.11%	9	10.00%	1	1.11%
11	Text 11	57	86.36%	7	10.61%	0	0%
12	Text 12	83	83.00%	7	7.00%	0	0%
13	Text 13	196	83.40%	23	9.79%	4	1.70%
14	Text 14	207	85.89%	20	8.30%	2	0.83%
15	Text 15	60	82.19%	5	6.85%	0	0%
16	Text 16	101	87.83%	6	5.22%	1	0.87%
17	Text 17	105	86.78%	9	7.44%	2	1.65%
18	Text 18	106	78.52%	14	10.37%	5	3.70%
	Total	**1841**	**-**	**175**	**-**	**35**	**-**
	Average	**102.33**	**84.10%**	**9.72**	**7.99%**	**1.94**	**1.60%**
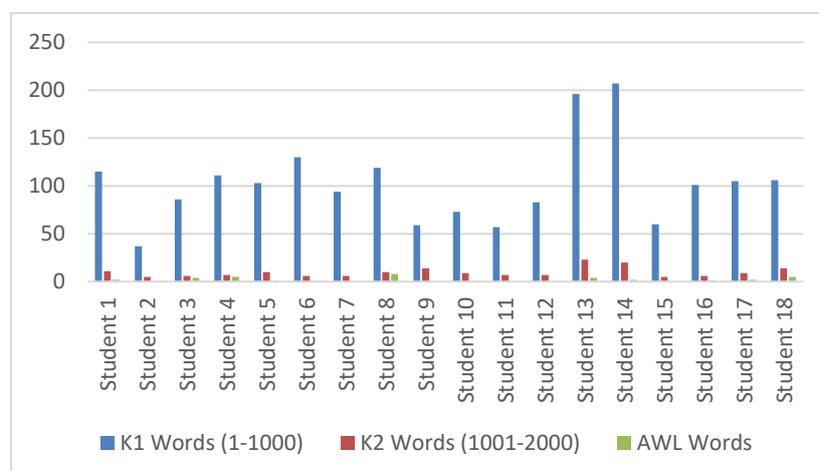
Figure 4. The depiction of the lexical sophistication in students' writing production

The table and figure above show the results of the vocabulary used in the descriptive texts produced by the students. It can be seen that the students rely on the first 1000 words in writing the descriptive texts as the highest percentage of the first 1.000 words list is 84.10%, followed by the second 1000 words list which is 7.99%, and the academic words list which is 1.60%. The academic words produced by the students include assignment, assignments, computer, energy, issues, link, monitor, overall, relax, schedule, schedules, submitted, task, tasks, transportation uniform, and via. The off-list words are the words that are not included in K1, K2, and AWL words list. The words mostly include the names of the students. Students 4 and 18 have the highest number of advanced word production with 5 words.

The researcher also analyzed the parts of speech of the writings. The researcher compiled all of the texts into one single file before checking the parts of speech via https://parts-of-speech.info/. The results show that the part of speech in the writings are dominated by nouns (23%) and verbs (19%), followed by prepositions (16%), and pronouns (16%). Furthermore, the least found part of speech in the writings are adjectives (2%), conjunctions (3%), numbers (5%), adverbs, (8%) and determiners (8%). It surprising that the topic of the writings are about descriptive text. Therefore, it can be concluded that the students had a limited vocabulary size.

Based on the findings, it was found that the students tend to use the first 1000 words in their writing production. There are several reasons why such phenomena occur. Firstly, it is because of their limited productive vocabulary. Therefore, the students tend to repeat the same or similar types of vocabulary in their writing production. This finding is based on the fact that the students are of Grade 7. Hence, they may have very limited vocabulary. It is also supported by the fact that the adjective found in the writings is only 2%, considering the texts were descriptive text. Secondly, the students mainly used the first 1000 words list because they avoid using unknown, unfamiliar, and new words in their communicative production. Thirdly, the task given by the teacher did not give the students the opportunity to use more advanced words.

In relation to other similar studies, the results of this study show that the participants had a low vocabulary size. A study written by Pertama and Ekawati (2022) also has a similar result. The overall results show that the short stories written by the students have quite low lexical richness even though the participants are university students.

D. CONCLUSION

This study examines the lexical richness of junior high school students' writings on descriptive text. The data were calculated by using a website accessed via www.lextutor.ca. This study used students' writings on descriptive text, which were collected from 18 junior high school students in one of the junior high schools in Yogyakarta. The results show that the lexical variation of the students on average 53%, meaning that the students use repeated words in their writings. The lexical density of the students' writing production is 47%. Lastly, the findings show that the students rely on the first 1000 words list with 84.10% of the total words produced. In conclusion, based on the findings, the students exceptionally produced low lexical richness. It can be caused by the students' limited vocabulary, the avoidance of using unknown and unfamiliar words, and the task which did not give the opportunity for the students to use various words.

The findings of this present study can be useful in language learning and teaching field. Teachers can take into consideration the utilization of lexical richness examination and investigation to examine students' productive vocabulary. Therefore, this present study might be helpful for teachers in reflecting their teaching as well as the teaching materials for the students. Another study which is written by Astridya (2018) also focused on analyzing lexical richness of students' writings. The results show that grade 12 has the highest result compared to grade 10 and 11 because they master a little vocabulary and tend to use words repeatedly.

REFERENCES

Astridya, F. W. (2018). Lexical richness of the expository writing in Indonesian by senior high school students. *Lingual*, 10 (1), 23-29. DOI: https://doi.org/10.24843/LJLC.2018.v05.i01.p04.

Azadnia, M. (2021). A Corpus-based Analysis of Lexical Richness in EAP Texts Written by Iranian TEFL Students. *Teaching English as a Second Language Quarterlya (Formerly Journal of Teaching Language Skills)*, *40*(4), 61-90.

Azodi N., Karimi F., Vaezi R. (2014). Measuring the lexical richness of productive vocabulary in Iranian EFL university students' writing performance. *Theory and Practice in Language Studies*, 4 (9), pp. 1837-1849. DOI: 10.4304/tpls.4.9.1837-1849.

Failasofah & Alkhrishes, H. T. D. (2018). Measuring Indonesian students' lexical diversity and lexical sophistication. *Indonesian Research Journal in Education*, 2 (2), pp. 97-107.

Gregori-Signes, C., & Clavel-Arroitia, B. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia-Social and Behavioral Sciences*, *198*, 546-556.

Ha, H., S. (2019). Lexical richness in EFL undergraduate students' academic writing. *English Teaching*, 74 (3), pp. 3-28. DOI: 10.15858/engtea.74.3.201909.3.

Juanggo, W. (2018). Investigating lexical diversity and lexical sophistication of productive vocabulary in the written discourse of Indonesian EFL learners. *Indonesian Journal of Applied Linguistics*, 8 (1), 38-48. DOI: 10.17509/ijal.v8i1.11462.

Laufer, B. and Nation P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* (16) 3, 307-322.

Malvern, D., Richards, B. (2012). The encyclopedia of applied linguistics. *Measures of Lexical Richness*. 1-5. DOI: 10.1002/9781405198431.wbeal0755.

Puspita, D. (2019). Corpus based study: Students' lexical coverage through business plan report writing. *Proceeding: The 2nd International Conference on English Language Teaching and Learning* (2nd ICON-ELTL), 25-38.

Ramos, F. D. R. Incidental vocabulary learning in second language acquisition: A literature review. *Profile* (17) 1, pp. 157-166. DOI: http://dx.doi.org/10.15446/profile.v17n1.43957.

Read, J. (2010). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Real, D.V.C., Cuizon, P.C., Necesito, R.D., Tuppal, C.P., Oducado, R.M.F., Lee, M.D., Javenes, K.J.C. (2020). Lexical richness of L2 production using Nation and Laufer's lexical frequency profile. *SDCA Asia-Pacific Multidisciplinary Research Journal*, 2, pp. 32-37.

Permata, N. (2022). Lexical richness of short stories written by EFL students. *EFL Education Journal*, 9(1), 102-116.

Siskova, Z. (2019). Lexical richness in EFL students' narratives. *Language Studies Working Papers*, 4, pp. 26-36.

Zhai, L. (2016). A study on Chinese EFL learners' vocabulary usage in writing. *Journal of Language Teaching and Research*, (7) 4, 752-759. DOI: http://dx.doi.org/10.17507/jltr.0704.16.