Lexical Richness of Students' Writings

Dwi Indarti

Universitas Bina Sarana Informatika
e-mail: dwi.diw@bsi.ac.id

Abstract - *The present study compares the lexical richness of 30 essays taken from the same class of University of Bina Sarana Informatika (henceforth mentioned as UBSI)'s students. The 30 essays were written by female and male of English Department students during their first, second, and third semester. This paper uses Web-based Second Language (L2) Lexical Frequency Profile (LFP), a batch mode of the web-based interface to the lexical richness proposed by (Laufer & Nation, 1995) which allows the users to calculate and to analyze the lexical richness of a corpus of text files at one time. The focus of this study is merely on the score of tokens, types and type-token ratio (TTR). The results show that the average of tokens produced by male students is higher than female students. In term of types, the average types of female students are lower than male students. Meanwhile, the highest TTR score was gained by female students (0.60) and the lowest TTR score was produced by male students (0.55), both in the second semester. A big TTR Score indicates a high lexical richness, while a small TTR shows a low lexical richness. By comparing the lexical richness of students' writings, this study is expected to show the students' writing skill improvement or impairment during first three semesters of studying English.*

Keywords: Lexical Richness, Students' writings, Lexical Frequency Profile

Abstrak – *Penelitian ini membandingkan kekayaan kosa kata (lexical richness) yang terdapat dalam 30 esai yang diambil dari kelas yang sama pada mahasiswa Universitas Bina Sarana Informatika. 30 esai tersebut ditulis oleh mahasiswa wanita dan pria jurusan Bahasa Inggris selama semester pertama, kedua dan ketiga. Penelitian ini menggunakan program online, bernama Lexical Frequency Profile (LFP), yang digagas oleh (Laufer & Nation, 1995). Program online tersebut memungkinkan penggunanya untuk menghitung dan menganalisa kekayaan kosa kata dari sekumpulan teks dalam waktu yang singkat. Penelitian ini hanya berfokus pada jumlah token (kata), types (jenis kata), dan type-token ratio (rasio kata dan jenis kata - TTR). Hasil penelitian ini menunjukkan bahwa mahasiswa pria menghasilkan token (kata) lebih banyak daripada mahasiswa wanita. Dalam hal jenis kata (types), mahasiswa wanita rata-rata memakai jenis kata yang lebih sedikit daripada mahasiwa pria. Sementara itu, jumlah TTR (rasio kata dan jenis kata) tertinggi dihasilkan oleh mahasiswa wanita (0.60) dan jumlah TTR (rasio kata dan jenis kata) terendah dihasilkan oleh mahasiswa pria (0.55) pada semester kedua. Nilai TTR yang tinggi menunjukkan kekayaan kosa kata yang baik, sementara nilai TTR yang rendah menunjukkan kekayaan kosa kata yang belum bervariasi. Dengan membandingkan kekayaan kosa kata, penelitian ini diharapkan dapat menunjuukan peningkatan atau penurunan kemampuan menulis para mahasiswa selama tiga semester pertama belajar Bahasa Inggris.*

Kata kunci: Kekayaan kosa kata, tulisan mahasiswa, Program Lexical Frequency Profile

INTRODUCTION

Writing is a very important skill to be mastered by the students because through writing, the students can preserve ideas, display the improvement of knowledge, improve vocabulary and communication skills. (Walsh, 2010) writes that writing is important because it is used extensively in higher education and in the workplace. Furthermore, (Walsh, 2010) explains that if students do not know how to express themselves in writing, they will not able to communicate well with professors, employers, peers, or others. There are many professional communication need proper writing skills, for examples: proposals, reports, job applications, business writings and many more (Klimova, 2013) concludes several benefits of writing, such as to express one's personality, to foster communication, to

develop thinking skills, to make logical and persuasive arguments, to give a person a chance to later reflect on his or her ideas and re-evaluate them, to provide and receive feedback and to prepare for school and employment.

Writing skill, as one of the four skills to be acquired by Foreign Language Learners (FLA), is considered as the most difficult skill and is viewed as the predictor of academic success (Graham & Perin, 2007). (Klimova, 2013) explains that writing requires mastering other skills, such as meta-cognitive skills. Learners need to set an objective for their writing, making a plan in form of drafts, using logical structure, then revise and edit. In the process of writing, students have to use cognitive skills; they have to analyze the sources then synthesize them in a compact piece of writing. (Klimova, 2013) suggests that one of the best ways to attract students to write is to ask them write at the beginning of the learning process as freely as possible and evoke in them the feeling of creativity.

Writing task is important in assessing English language learning because it shows the students' ability to explore the vocabulary building, to express their ideas in a comprehensive text, and to convey their ability of English learning. How to measure the quality an essay can be seen from its lexical richness as (Laufer & Nation, 1995) stated that, "The quality of an essay can be seen from its lexical richness as the result of learning new vocabulary, activating previously known vocabulary or proficient in the writing skill." (p. 308). In other words, lexical richness is the variety of vocabulary size used by an author to write an essay. Furthermore, (Laufer & Nation, 1995) also stated that, "a well-written composition, among other things, makes effective use of vocabulary. This need not be reflected in a rich vocabulary, but a well-used rich vocabulary is likely to have a positive effect on the reader." Meanwhile, (Milton, 2008) states that, "Measuring the vocabulary richness of learners can helo give a much better impression of the scale of learning which is taking place than is possible with other measures of language proficiency."

Hoover (2003) as cited in (Djiwandono, 2016) defined the lexical richness as the ratio of types of words to total words (token) written in a text or TTR. In details, (Djiwandono, 2016) explained that, "Type refers to the types of words, while token encompass the total number of words used in a particular text." (p. 210). A big TTR score indicates a high lexical richness, while a small TTR shows a low lexical richness.

(Siskova, 2012) divided the term of lexical richness into three groups: lexical diversity (saying how many different words are used), lexical sophistication (saying how many advanced words are used) and lexical density (saying what the proportion of content words in the text is). Meanwhile, (Laufer & Nation, 1995) proposed several measures of lexical richness; Lexical Originality (LO), Lexical Density (LD), Lexical Sophistication (LS), Lexical Variation (LV) and Lexical Frequenxy Profile (LFP).

1. Lexical Originality

It is the percentage of words in a given piece of writing that are used by one particular writer and no one else in the group.

$$LO = \frac{Number\ of\ tokens\ unique\ to\ one\ writer\ x\ 100}{Total\ number\ of\ tokens}$$

The Lexical Originality index measures the learners' performance relative to the group in which the composition was written (Laufer & Nation, 1995)

2. Lexical Density

It is defined as the percentage of lexical words in the text, such as nouns, verbs, adjectives, and adverbs.

$$LD = \frac{Number\ of\ tokens\ x\ 100}{Total\ number\ of\ tokens}$$

Lexical Density does not necessarily measure lexis, since it depends on the syntactic and cohesive properties of the composition (Laufer & Nation, 1995)

3. Lexical Sophistication

It is the percentage of 'advanced' words in the text.

$$LS = \frac{Number\ of\ advanced\ tokens\ x\ 100}{Total\ number\ of\ lexical\ tokens}$$

What is labeled as 'advanced' would depend on the researchers' definition. To decide what vocabulary is advanced, it is necessary to take the learner's level into consideration. Lexical Sophistication is determined by the researchers' definition of advanced or sophisticated words, its uses are limited (Laufer & Nation, 1995)

4. Lexical Variation

It is the type/token ration in per cent between the different words in the text and the total number of running words.

$$LV = \frac{Number\ of\ types\ x\ 100}{Total\ number\ of\ tokens}$$

The type/token ratio has been shown to be unstable for short texts and can be affected by differences in text length; even more sophisticated formular have been shown to be unsuitable for short texts like learners' essays. LV distinguishes only between the different words used in a composition (Laufer & Nation, 1995)

5. Lexical Frequency Profile

The LFP (Lexical Frequency Profile) shows the percentage of words at different vocabulary frequency levels. The calculation is done by an online computer program which compares vocabulary lists against a text that has been typed to see what words in the texts are and are not in the list and to see what percentage of the items in the texts are covered by the lists. The program can calculate the lexical frequency on the basis of word tokens, word types or word families (Laufer and Nation, 1995, p. 312). The classification of words is based on the General Service List (GSL) published by Michael West (1953) which represents the most frequent words of English. It consists of K1 words or a thousand most frequent words of English (1-1000) and K2 words or two thousand most frequent words of English (1001-2000). The program also shows the AWL (Academic Word List) that is words with high frequency appearance in English academic texts. Below is the terminology of lexical richness (Indarti, 2017).

Table 1. Terminology of lexical richness

The 1st 1000 words (1-1000)	K1 words
The 2nd 1000 words (1001-2000)	K2 words
Academic Word List	AWL
Number of words	Tokens
Number of different words	Types
Type/token ratio	TTR

Studies of lexical richness using academic writing productions or students' writings have been conducted by many scholars. (Ibrahim et al., 2019) compare the lexical richness of two groups of EFL learners, through a content analysis of 139 essays of entry level university students and 140 essays of third-year university students at an English medium university. They set standard of scored at least a band 6 or 550 in TOEFL. The study uses RANGE program developed by Nation, Heat-ley & Coxhead (2002) and reveals that the two groups of essays show a statistically significant difference in the use of the one thousand, two thousand and Academic Word List (AWL) word level. The results show that there tend to be an over-dependence of the post-sessional students on the one thousand most frequent words while the pre-sessional students show that they are using fewer of one thousand most frequent words.

(Zhang & Daller, 2019) explore the lexical richness of Chinese candidates of different proficiency levels in a graded examination in spoken English (GESE, by comparing 5 lexical indices and the mean length of utterances (MLU) of GESE candidates of three proficiency levels; Grade 2, Grade 5, Grade 7. The results of the study show that measures of lexical richness and MLU are good predictors of success in oral interviews, but that factors other than proficiency play a role when it comes to the placement of students in Grade 7. The study presents further insights in the role that vocabulary knowledge plays in oral interviews.

(Thawarom & Singhasiri, 2020) investigate the quality of vocabulary use by measuring lexical richness (lexical diversity, lexical sophistication and lexical density) of speech production in ESP classroom context. The data was collected from 100 first year university students while performing a speaking task. The study uses the D-tools and AntWordProfiler to measure the diversity and sophistication, while TagAnt was applied for calculating percentages of lexical density and a scale-based approach was employed for technical identification. The results revealed that vocabulary produced by the students was mainly from GSL 1000 words list. Lexical density of the speech was 43,72% and the lexical diversity was moderately rich with a value of D 58.8. The results of the study imply that ESP students need a great awareness that they are parts of a particular group.

(Halim, 2018) finds out the lexical richness of undergraduate theses of English Language and Culture Department. The researcher uses computer software called AntWord Profiler, a kind of software for profiling texts created by Laurence Anthony in order to produce the type, and the token of the text, and uses type-token ratio (TTR) as the method in measuring the lexical richness a part of data analysis. The closer the TTR score to 1, the higher the lexical richness is. However, the result of the study shows that none of the students reached even 0.5 score of TTR.

(Astridya, 2018) investigates the lexical richness, lexical density, lexical sophistication and laxical variation of 90 senior high school students' expository essays. According to the data, grade 12 produces the highest number among others and it concludes that students in grade 12 is the most prepared and have the richest lexical richness.

Meanwhile, (Djiwandono, 2016) investigates the lexical richness in academic papers. He compared between students and lecturers' papers to determine

the difference between type-token ratio (TTR), the use of 2000-word level (or K2), the use of academic words, and the use of off-list words. There are four objectives of the study: (1) to determine the difference between the type-token ratio (TTR) in students' essay and that in their lecturers' essays; (2) to determine the difference between the use of 2000-word level in students' essays as well as in lecturers' essays; (3) to determine the difference between the use of academic words in students' essays and in lecturers' essays; (4) to determine the difference between the students' essays and their lecturers' in terms of the use of words other than the 2000-word level and the academic word

The result showed that the lecturers fare better in term of TTR and academic words, but write slightly fewer 2000-word level and off-list words than the students.

This paper used a corpus of thirty student essays that have been collected from the first semester until the third semester, therefore this could be considered as a longitudinal study. The objective of this study is to find out whether there is a development of lexical richness in their writings. By comparing the lexical richness of students' writing in every semester, this study is expected to reveal whether there is writing improvement during three semesters of studying English.

RESEARCH METHODOLOGY

This study used the LFP (Lexical Frequency Profile) proposed by (Laufer & Nation, 1995) to reveal the percentage of words at different vocabulary frequency levels. The calculation is done by a computer program which compares vocabulary lists against a text that has been typed. The free online program can calculate the LFP on the basis of word tokens, word types or word families. The classification of words is based on the General Service List (GSL) and Academic Word List (AWL).

Technique of Data Collection and Analysis

The total of thirty essays was taken from the same class of UBSI students during their first, second, and third semester. Each semester, the writer who was also the lecturer, took ten essays from five female students and five male students. The topic of the essays of each semester was determined by the lecturer. The topic for the first semester was *My dream job*, the second semester was *The story of my life*, and the third semester was *The power of social media*. The student essays were processed one by one using the free-web computerized online digital tools via www.lextutor.ca. The writer chose 'the *Vocabulary Profile'* menu, then '*VP-Classic'* menu and the next is to input each text , finished by press the 'SUBMIT_Window' button

RESULTS AND ANALYSIS

Table 1. Tokens, Types and TTR of Students' writings

FEMALE					MALE				
1		1st smtr	2nd Smtr	3rd Smtr	1		1st smtr	2nd Smtr	3rd Smtr
	Tokens	151	276	126		Tokens	242	143	131
	Types	95	147	80		Types	138	87	73
	TTR	0.63	0.53	0.63		TTR	0.57	0.61	0.56
2		1st smtr	2nd Smtr	3rd Smtr	2		1st smtr	2nd Smtr	3rd Smtr
	Tokens	152	149	217		Tokens	198	196	213
	Types	88	93	105		Types	122	103	110
	TTR	0.58	0.62	0.48		TTR	0.62	0.53	0.52
3		1st smtr	2nd Smtr	3rd Smtr	3		1st smtr	2nd Smtr	3rd Smtr
	Tokens	180	185	275		Tokens	175	317	175
	Types	111	114	135		Types	89	154	103
	TTR	0.62	0.62	0.49		TTR	0.51	0.49	0.59
4		1st smtr	2nd Smtr	3rd Smtr	4		1st smtr	2nd Smtr	3rd Smtr
	Tokens	206	149	124		Tokens	151	147	183
	Types	107	94	71		Types	94	93	104
	TTR	0.52	0.63	0.57		TTR	0.62	0.63	0.57
5		1st smtr	2nd Smtr	3rd Smtr	5		1st smtr	2nd Smtr	3rd Smtr
	Tokens	284	152	75		Tokens	169	186	302
	Types	123	92	50		Types	100	92	175
	TTR	0.43	0.61	0.67		TTR	0.59	0.49	0.58

First semester

Female student 1 produces tokens 151, types 95, and TTR 0.63. Female student 2 produces tokens 152, types 88, and TTR 0.58. Female student 3 produces tokens 180, types 111 and tokens 0.62. Female student 4 produces tokens 206, types 107, and TTR 0.52. Female student 5 produces tokens 284, types 123, and TTR 0.43. Meanwhile, male student 1 produces tokens 242, types 138, and TTR 0.57. Male student 2 produces tokens 198, types 122, TTR 0.62. Male student 3 produces tokens 175, types 89, and TTR 0.51. Male student 4 produces tokens 151, types 94 and TTR 0.62. Male student 5 produces tokens 169, types 100 and TTR 0.59.
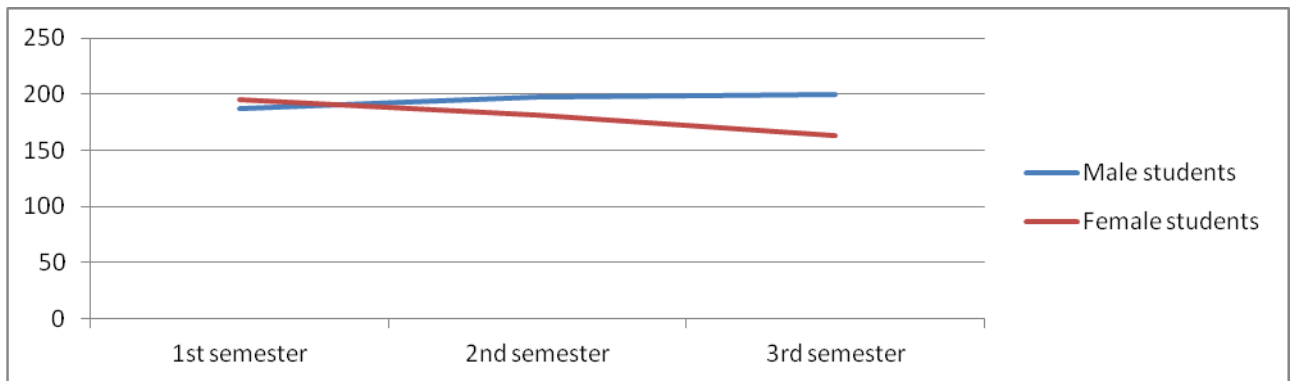
Second semester

Female student 1 produces tokens 276, types 147, and TTR 0.53. Female student 2 produces tokens 149, types 93, and TTR 0.62. Female student 3 produces tokens 185, types 114, and TTR 0.62. Female student 4 produces tokens 149, types 94, and TTR 0.63. Female student 5 produces tokens 152, types 92, and TTR 0.61. Meanwhile male student 1 produces tokens 143, types 87, and TTR 0.61. Male student 2 produces tokens 196, types 103, and TTR 0.53. Male student 3 produces tokens 317, types 154, and TTR 0.49. Male student 4 produces tokens 147, types 93,

and TTR 0.63. Male student 5 produces tokens 186, types 92 and TTR 0.57.

Third semester

Female student 1 produces tokens 126, types 80 and TTR 0.63. Female student 2 produces tokens 217, types 105, and TTR 0.48. Female student 3 produces tokens 275, types 135, and TTR 0.49. Female student 4 produces tokens 124, types 71 and TTR 0.57. Female student 5 produces tokens 75, types 50 and TTR 0.67. Male student 1 produces tokens 131, types 73, and TTR 0.56. Male student 2 produces tokens 213, types 110, and TTR 0.52. Male student 3 produces tokens 175, types 103, and TTR 0.59. Male student 4 produces tokens 183, types 104, and TTR 0,57. Male student 5 produces tokens 302, types 175 and TTR 0.58.
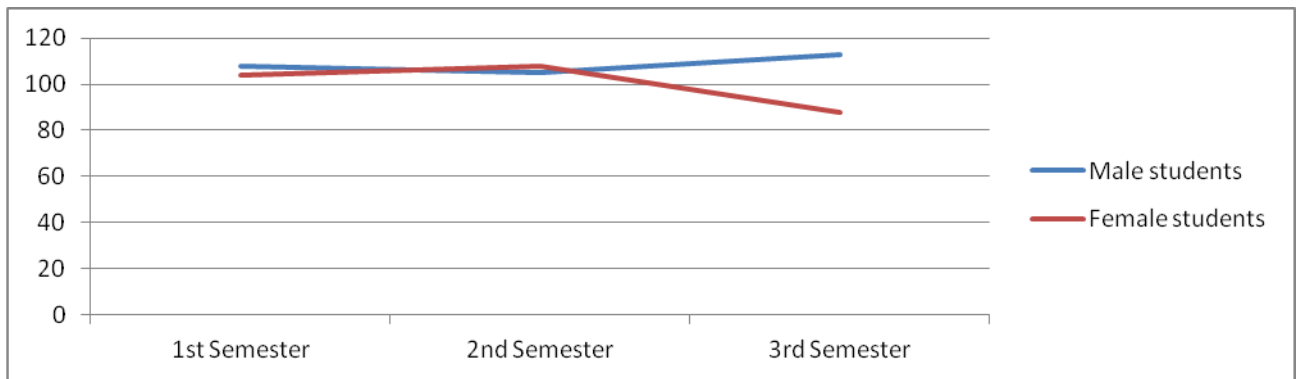
The average of female students' token in the first semester is 195, the second semester is 182 and the third semester is 163. There is a decreasing tendency of female students' token during three semesters. Meanwhile, the average of male students' token in the first semester is 187, the second semester is 197 and the third semester is 200. This shows that male students' tokens are increase during three semesters. The illustration can be seen from the graphic 1 below.

Graphic 1. The average of students' token

In term of types of words (types), the average types of female students in the first semester is 104, then in the second semester increases to 108 and in the third semester decreases up to 88. Meanwhile, the average types of male students' in the first semester is 108, then in the second semester decreases to 105 and in the third semester increases to 113. The illustration can be seen from the graphic 2 below.

Graphic 2. The average of students' token

The average of type token ratio (TTR) of students' writings during three semesters can be seen from diagram 1 below. In the first semester, female students' TTR is 0.55, the second semester is 0.60 and the third semester is 0.56. While, the average TTR of male students in the first semester is 0.58, the second semester is 0.55 and the third semester is 0.56.

It can be inferred that there is no significant difference of students' writings during three semesters. The highest TTR was gained by female students in the second semester (0.60) while the lowest TTR was gained by male students in the second semester.
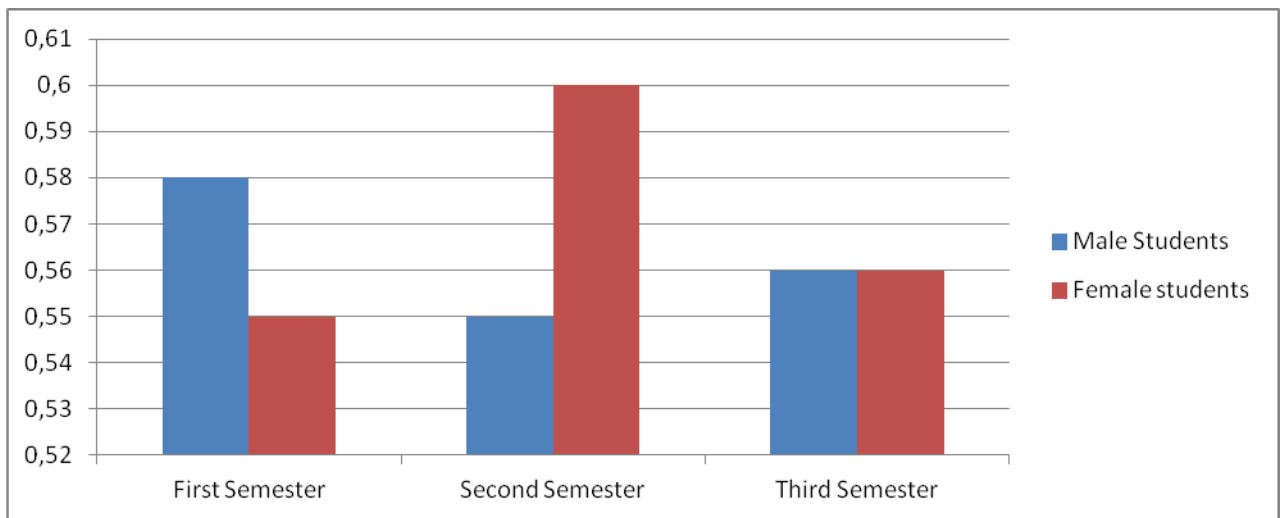

Diagram 1. The average of Type Token Ratio (TTR)

CONCLUSION

A good quality of an essay can be measured by the score of type-token ratio (TTR). A big TTR score indicates a high lexical richness. It means that the essay's writer uses many various types of words in order to make a well-written essay. This study used students' essays taken from the first three semesters of their studying of English. The calculation done by a free-web computerized online digital namely Lexical Frequency Profile proposed by (Laufer & Nation, 1995). The result shows there is non-

significant difference and non-significant improvement of the students' writings during three semesters. In fact, the result tended to decrease in the second semester and then, tended to increase back in the third semester. The reason needs to be investigated deeply in further research.

However, this study merely highlighted the token, type and TTR of students' writings, while there are numerous terms to be investigated to reveal the lexical richness of an essay. In the future studies,

scholars can explore about the General Service List (GSL words) that contains K1-words or one thousand most frequent words of English (K-1000), K2 or two thousand most frequent words of English (K-2000) and Academic Word List (AWL). The researchers also can investigate the lexical originality, lexical density, lexical sophistication or lexical variation of students' writings. These topics are significant to measure the quality and the improvement of student's writing skill.

REFERENCES

Astridya, F. W. (2018). Lexical richness of the expository writing in Indonesian Senior High School students. *Lingual: Journal of Language and Culture*, *5*(1), 23–29. https://ojs.unud.ac.id/index.php/languange/article/view/38446

Djiwandono, P. I. (2016). Lexical richness in Academic Papers: A comparison between students' and lecturers' essays. *Indonesian Journal of Applied Linguistics*, *5*(2), 209–216. https://ejournal.upi.edu/index.php/IJAL/article/view/1345

Graham, S., & Perin, D. (2007). *Writing Next: Effective strategies to improve writing of adolescents in middle and high schools: A report to Carnegie Corporation of New York*. Alliance for excellent educaiton. https://www.researchgate.net/publication/230853210_Writing_Next_Effective_Strategies_to_Improve_Writing_of_Adolescents_in_Middle_and_High_Schools_A_Report_to_Carnegie_Corporation_of_New_York

Halim, S. W. (2018). Lexical richness in English Language and Culture Department students' undergraduate theses. *Journal of English Language and Culture*, *8*(2), 140–151. https://journal.ubm.ac.id/index.php/english-language-culture/article/view/1098

Ibrahim, E., Muhamad, A., & Esa, Z. (2019). A comparison of lexical richness in L2 written production. *International Journal of Emerging Technologies in Learning (IJET)*, *14*(20), 174–181. https://online-journals.org/index.php/i-jet/article/view/11467

Indarti, D. (2017). Lexical richness of the Jakarta Post Opinion Articles: Comparison between Native and Non-native writers. *Wanastra*, *IX*(2), 138–142. https://ejournal.bsi.ac.id/ejurnal/index.php/wanastra/article/view/2550

Klimova, B. F. (2013). The importance of writing. *Indian Journal of Research*, *2*(1), 9–11. https://www.researchgate.net/publication/274925223_The_Importance_of_Writing

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied LInguistic*, *16*(2), 307–322. https://www.researchgate.net/publication/249237693_Vocabulary_Size_and_Use_Lexical_Richness_in_L2_Written_Production

Milton, J. (2008). French vocabulary breadth among learners in the British school and university system: comparing knowledge over time. In J. Treffers-Daller, H. . Daller, D. Malvern, B. Richards, P. Meara, & J. Milton (Eds.), *Eurosia monographs series 2 L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 57–78). EuroSLA/John Benjamins.

Siskova, Z. (2012). Lexical richness in EFL students' narratives. *Language Studies Working Papers*, *4*, 26–36. https://www.researchgate.net/publication/305999633_Lexical_Richness_in_EFL_Students%27_Narratives

Thawarom, Z., & Singhasiri, W. (2020). Lexical richness of one-minute speaking task by Science and Technology University students. *The Journal of ASIA TEFL*, *17*(1), 70–86. http://journal.asiatefl.org/main/main.php?inx_journals=63&inx_contents=778&main=1&sub=2&submode=3&PageMode=JournalView&s_title=Lexical_Richness_of_One_Minute_Speaking_Task_by_Science_and_Technology_University_Students

Walsh, K. (2010). *The importance of writing skills: Online tools to encourage success*. https://www.emergingedtech.com/

Zhang, J., & Daller, M. (2019). Lexial richness of Chinese candidates in the graded oral English examination. *Applied LInguistic*, *2*(34), 2–34. https://www.researchgate.net/publication/331098397_Lexical_richness_of_Chinese_candidates_in_the_graded_oral_English_examinations