ELSEVIER

Exploring Indonesian EFL students' lexical diversity and its correlation with academic vocabulary use in an online academic writing environment

Beauty Sholeha Raufi, Herri Mulyono [*], Hamzah Puadi Ilyas, Siti Zulaiha

Universitas Muhammadiyah Prof. DR. HAMKA, Indonesia

A B S T R A C T

This paper reports on a quasi-experimental study designed to explore the relationship between lexical diversity and academic vocabulary use among Indonesian EFL university students engaging in online academic writing discussions and assignments. The study analysed 11,624 tokens and 5437 types collected from students' contributions in online discussions and academic writing assignments. Findings revealed that the level of lexical diversity among the students was significantly varied depending on the text length and the total amount of data. It was found most students were indicated to possess average ESL level of academic writing with a D value ranging from 50 to 70 while some others indicated a high level of adult ESL with developed academic text in the range between 70 and 80 D level. Moreover, it was also found that the degree of lexical diversity was not affected by the writing topic familiarity or students' academic year. Findings revealed a statistically significant negative correlation between measures of lexical diversity and academic vocabulary use. Contrary to expectations, while a diversified vocabulary generally correlates positively with student writing performance, this relationship reverses in contexts emphasizing the academic quality of writing. These results highlight the complex dynamics of lexical usage in academic writing and suggest that mere lexical diversity does not always equate to higher academic writing quality. This study contributes to a critical understanding of how lexical diversity functions in digital academic environments and offers implications for English language teaching practices in higher education settings.

1. Introduction

The rapid advancement of technology has significantly enhanced virtual transactional communication, particularly in the context of social media where Indonesian citizens increasingly engage in discussions on political issues. According to the literature (Lim, 2017; Ritonga and Syahputra, 2019), this engagement signifies a vigorous social capital that reflect the increase of public awareness and a collective readiness to address national concerns. However, this ease of communication also presents challenges, particularly among the youth whose inadequate literacy can facilitate the dissemination of false information which in turn threaten national unity (Lim, 2017; Ritonga and Syahputra, 2019).

The role of social media in fostering or degrading civic engagement links directly to the levels of media literacy and critical capacity among its users. This concern extends into educational settings, where the lack of study on the use of students' vocabulary in online transactional communication complicates the assessment of their critical thinking and

its impact on academic writing skills. Moreover, technology's ubiquitous presence in educational setting is widely documented, especially in enhancing learner's lexical diversity and measuring writing skills (Akbari, 2017; Alshehri, 2022; Hammou et al., 2021; Hassanzadeh et al., 2021; Maamuujav, 2021; Shibata, 2022; Treffers-Daller et al., 2018; Vidal and Jarvis, 2020). Yet, research focusing on the correlation lexical diversity and academic writing, particularly in Indonesian tertiary institutions, remains scarce.

Lexical diversity, reflecting a form of lexical richness, profoundly influences the language abilities of active users, particularly in choosing appropriate diction for transactional communication (Alshehri, 2022; Malvern et al., 2004). Over recent decades, it has become a pivotal tool in evaluating language proficiency, primarily through writing performance. According to Akbari (2017), lexical variation in online discussions and academic assignments can serve as an important metric for evaluating students' critical thinking and writing abilities. Yu (2010) also asserts that lexical diversity is a key measure in analysing student

writing, indicating the levels of cognitive engagement during the learning process. Thus, increasing lexical diversity in transactional written communication, such as forum discussions, can both signal and enhance students' critical thinking capabilities, reflecting significantly in their academic writings.

Hassanzadeh et al. (2021) conducted a study to investigate the effects of computer-aided concept mapping (CACM) on the lexical diversity (LD) of English as a Foreign Language (EFL) learners' writing, comparing it to traditional outlining methods. A total of 53 university EFL undergraduate students participated in the study, where both pre-tests and post-tests were administered to evaluate writing quality and lexical diversity. The findings indicated a statistically significant difference in LD indices between the CACM group and the outlining group, suggesting that CACM positively influenced lexical diversity. However, the study found no statistically significant correlation between LD and writing quality scores in either group, indicating that while CACM enhanced lexical diversity, it did not necessarily improve overall writing quality. Another study conducted by Nasseri and Thompson (2021) examined lexical density and diversity in dissertation abstracts written by postgraduate students who are native English speakers (L1) and those learning English as a Foreign Language (EFL). The study analysed a collection of 210 dissertation abstracts from various disciplines, employing several indices to quantify lexical density and diversity, including NDWESZ, UBER, MSTTR, and HD-D. The findings revealed significant differences in lexical density and diversity between L1 and EFL writers. Notably, specific measures such as MSTTR and HD-D were particularly effective in capturing these differences.

To understand the significance of lexical function in academic writing within this context of transactional written communication, it is crucial to accurately measure the extent of lexical diversity. Analytical tools like Token-type ratio (TTR) (Malvern et al., 2004) and Parameter D (Durán et al., 2004; Shibata, 2022; Yu, 2010) have been extensively employed to measure lexical diversity with the aid of digital applications such as WordSmith Tools (P. McCarthy, 2005). Studies like that of Shibata (2022) have validated the precision of digital methods in determining lexical diversity levels, which are significantly affected by students' comprehension of the topic and available sociocultural resources. Despite the study findings above, studies specifically targeting the correlation between of lexical diversity and academic writing skills in Indonesian tertiary settings using digital modalities are still lacking. Recognizing this, the current study aims to develop a better understanding of how lexical diversity in online written discussions and assignments influences writing performance among Indonesian undergraduates. Specifically, the current study aimed to answer two research questions.

(1) What is the level of Indonesian EFL university students' lexical diversity when participating in online academic discussion and writing assignments?
(2) Does the students' lexical diversity correlate with their use of academic words in an online academic discussion and writing assignment environment?

To achieve these objectives, a quasi-experimental research design was implemented over a one-year period, focusing on the degree of lexical diversity among students in online discussions and academic writing assignments and analysing the relationship between students' lexicon diversity and their academic vocabulary use in a postsecondary environment. This approach not only addresses the current research gap but also contributes to the broader academic dialogue on enhancing education quality through technology-mediated language learning strategies.

2. Method

2.1. The study design and description

The current study employed a quasi-experimental design to analyse improvements in thirty-two EFL university students' academic writing abilities in reference to their lexical diversity level. This approach allows for the examination of these parameters within a single group of students without the use of control groups. Importantly, the study did not employ a pre-test and post-test structure; instead, it focused on evaluating student performance through the analysis of twenty-six different assignments. These assignments comprised a combination of individual writing tasks and contributions to written discussions, all conducted on an online discussion and assignment platform. Each assignment was developed around specific topics designed to measure students' lexical diversity and enhance their academic writing skills through critical engagement. In this quasi-experimental design, each student's progression was monitored and evaluated over time through continuous submissions, providing a dynamic assessment of their academic capabilities.

2.2. Data collection and the study discourse

The data composition consisted of 11,624 tokens and 5437 types in total were obtained from university students' online discussion and assignment in an academic writing course. Data selection involved two stages. At stage 1, samples were selected according to the two requirements simultaneously. First, participants completed at least half of the total required writing in the discussion platform. Thus, 18 participants with above 70% ($n \geq 40$) of writing completion have remained from 32 candidates. In the first phase of data analysis, the data samples were transcribed with no grammatical corrections (Wang, 2014). However, the misspelling was corrected according to the following principles since recognizing different conjugations from the same word as distinct types could confute in determining lexical with grammatical knowledge (Yu, 2010). The first procedure in this method is, that it is acceptable if there is one missing letter, one misplaced letter or one wrong addition letter to be selected.

Secondly, in order to verify high inter-rater reliability, all samples received the same topic given by the instructor. At stage 2, all samples were checked with some criteria. First, the length of the text generally consisted of a short to mid-short length during the first half semester and a mid-long to long length in the second half to the end of the semester which indicated the final writing for the final test. Second, the samples were maintained to consist of 50 words at minimum to ensure achieving the text probability and stabilization value. These mentioned principles for revising miscalculations spelling were grounded according to the rationale that the focus of lexical diversity is to examine how different the word range has appeared. Thus, to fulfil this main objective of lexical diversity, the words were presented in a legible, genuine and recognizable form. However, words with more than one incorrect or misplaced letter were removed since it intervened in the process of distinguishing the lexical diversity with two or further incorrect letters as a misspelling issue or causing a non-existent word. Therefore, from 32 participants, 18 participants remained with 900 samples selected from 26 constructs.

2.3. Data analysis

The data analysis of the current study was conducted in two phases. First, the data sample was gathered and transcribed using a computer program called *Text Inspector* to measure the lexical diversity. In the current study, the D parameter of lexical diversity was examined to address the research question one. D is a concept parameter of lexical diversity in a mathematical model function that revised the weakness from the TTR model. The calculation method using D is claimed to be able to establish the relative lexical diversity for either short or long texts. Moreover, according to Yu (2010), parameter D could measure the

lexical diversity and determine the statistical significance of the correlation between the learners' writing and their general language proficiency. Furthermore, a study by Shibata (2022) also claimed that D was frequently used to investigate the in-depth meaning of lexical diversity to elucidate the lexical diversity in language ability, especially writing.

According to Durán et al. (2004), D is a measuring device of lexical diversity. Moreover, *vocd* is known to calculate the sequential sub-samples of Ds. The final output of *vocd* is an index referred to as D, suggesting that the D formula is used to determine the level of LD in the text (P. M. McCarthy and Jarvis, 2007). The value of *vocd* ranges from 10 to 100 with a higher value of D indicating a higher lexical diversity level. Following the indicator, the lower the lexical diversity produces, the lower the D value. McCarthy and Jarvis (2007) perceive that the *vocd* value is based on word occurrence probabilities, and D serves no use other than to transfer the LD value to a new scale. Moreover, *vocd* carries out its curve-fitting procedure on a curve segment of 35–50 tokens. Therefore, a minimum of 50 valid words in the data of the recent study are needed to supply all 16 data points. In this case, data with words less than 50 words are automatically eliminated during the measurement process. Moreover, the measurement of LD using the *vocd* index is significantly affected by the text lengths. Thus, the probability of D will also increase if the text lengths increase (P. M. McCarthy and Jarvis, 2007). However, a higher TTR will always be found in the smaller part of the text concerning tokens. Therefore, the apparent diversity of texts will be exaggerated if that higher TTR is added or fitted to a mean TTR. The equation of *vocd* is done by summing the probability of occurrence for each word multiplied by 1/r (or one divided by the sample size).

The calculation method using D was claimed to be able to establish the relative lexical diversity for either short or long texts. Moreover, according to Yu (2010), parameter D could measure the lexical diversity and determine the statistical significance of the correlation between the learners' writing and their general language proficiency. Furthermore, a study by Shibata (2022) claims that D is frequently used to investigate the in-depth meaning of lexical diversity to elucidate the lexical diversity in language ability, especially writing. Therefore, the current study employed the parameter D index to answer the first research question in determining the level of lexical diversity among students in online discussion and academic writing assignments. In addition, a study by Durán et al. (2004) suggest that if there was found a high level of D in a text, it could be concluded that the text has a greater level of lexical diversity. Hence, the higher the D level, the more substantial lexical diversity was found. In brief, D was claimed to be a more appropriate indicator in measuring lexical diversity than the TTR traditional method.

In addition to the examination of D parameter, the data was analysed statistically using JAMOVI as a statistical analysis software or computational tool. Furthermore, the recent study employed JAMOVI software to estimate the mean, standard deviation, correlation and regression analysis (to observe the outliers and correlation between the parameter D and the quality ratings using the D value), ANOVA statistic (to determine the denotative difference in parameter D appeared between the writing experience and writing topics; writing topics and language proficiency; language proficiency and writing experience).

3. Findings and discussion

The current study analysed the lexical diversity from a sample of Indonesian EFL university students' lexical diversity when participating in a digital academic writing activity. Specifically, two research questions were addressed to investigate the degree of lexical diversity among students in online discussion and academic writing assignments and the correlation between the diversity of their lexicon and their academic vocabulary use in postsecondary settings.

3.1. The degree of university students' lexical diversity

The current study examined lexical diversity using parameter D to address the first research question. Fig. 1 presents the graphic of the mean level of D between the participants while Table 1 illustrates the value of D in the detailed number form. The compositions include eighteen participants with twenty-six constructs of students' writing. Each participant was treated as the sample that produced fifty sub-constructs. The final compositions involved ± 900 data from the selected participants analysed using the *Text Inspector* software. The analysed data represented the participant's level of lexical diversity in online discussion and academic writing assignments.

Table 1 and Fig. 1 show that the lowest level of D was performed by sample 16 ($D = 55,82$) while the highest level was performed by sample 4 ($D = 100.05$). Although the D levels between samples were found to be heterogeneous between the range 50–100, it was found that the mean of each level ($D = 78.80$) indicates the level of adult ESL (Durán et al., 2004). In addition, some samples were found to indicate the adult ESL level individually with a D value ranging from 50 to 70 (i.e., 8 to 9, and 12 to 16). Moreover, several samples (i.e., 1 to 7, 10 to 11, and 17 to 18) demonstrated a high level of adult ESL, with developed academic text in the range between 70 and 80 D level.

Table 2 below demonstrates the descriptive statistic of lexical diversity. The result of lexical diversity analysis indicates the adult ESL mid to high level with a minimum score of 55.82 to a highest of 100.05. The mean is calculated at 78.80 with a standard deviation of 12.79. In addition, the mid score of D was found in 77.52, which means that most of the participants performed the academic text with their lexical diversity level.

Furthermore, Fig. 2 presents the graphic detail of the D levels of the participants in each construct. The highest level of lexical diversity was performed by sample 4 in Discussion 6 with a $D = 153,28$. However, there was also an apparent indication of a low level of D from several samples' values at zero levels due to the differences in students' online discussion and academic writing assignments. Several students produced writing less than fifty writing. Therefore, it was irrefutable to claim that the level of lexical diversity among the students is significantly diverse depending on the text length and the total amount of data.

Despite their familiarity with the subject area, it was found in Fig. 2 above that students' lexical diversity had been constantly fluctuating and uneven. Therefore, the unsteady graphic of students' lexical diversity indicates that there is no guarantee that the degree of lexical diversity was significantly affected by the writing topic familiarity or students' academic year. On the contrary, the degree may change upon both internal and external factors during the writing period.

In brief, answering the first research question, (RQ1. What is the level of Indonesian EFL students' lexical diversity when participating in online academic discussion and writing assignments), the level of lexical diversity among the students varied significantly depending on the text length and the total amount of data. The *Text Inspector* analysis revealed that most of the participants' *D* scores indicated the adult ESL level individually (Durán et al., 2004) with a D value ranging from 50 to 70 while some others demonstrated a high level of adult ESL with developed academic text in the range between 70 and 80 D level. Moreover, it was also found that the degree of lexical diversity was not affected by the writing topic familiarity or students' academic year.

3.2. The correlation between students' lexical diversity and their use of academic words

Before the statistical analysis, students' academic writing competencies were examined through the *Text Inspector* software by performing the AWL test. Academic word list (AWL) is claimed as the ideal that represents the use of words in practice (Biber et al., 1994). Although AWL is not the salient evidence, they considered it to be an essential element for demonstrating that a given text is academic in nature
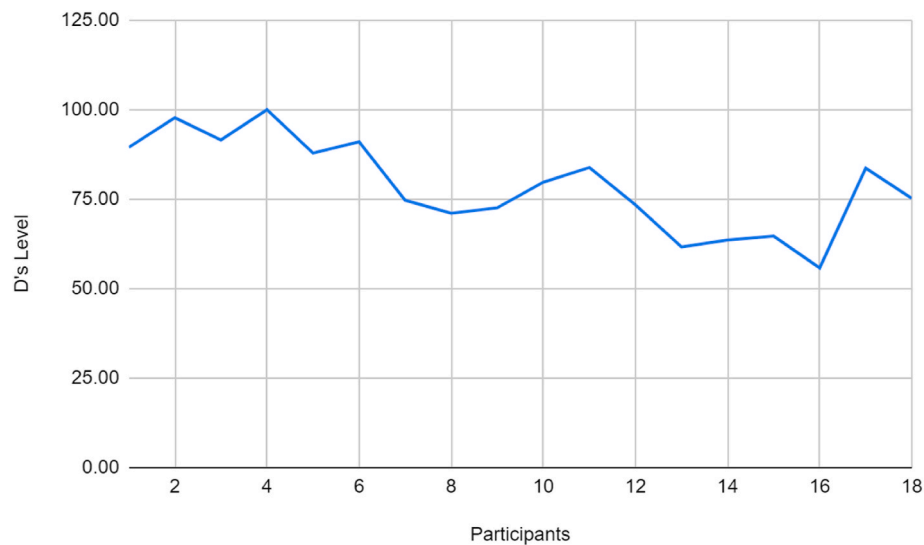
Fig. 1. VOCD and MLTD Mean of 18 participants.

Table 1
Descriptive statistics for D of 18 participants.

Participants	Average D
A	89.53
B	97.80
C	91.59
D	100.05
E	87.92
F	91.08
G	74.75
H	71.11
I	72.62
J	79.79
K	83.91
L	73.44
M	61.67
N	63.60
O	64.72
P	55.82
Q	83.72
R	75.25

Table 2
Summary of descriptive statistic of D.

Mean	78.80
Standard Deviation	12.79
Median	77.52
Min	55.82
Max	100.05

(Coxhead, 2000). Thus, AWL could be seen as a model for determining the correlation between a writer's basic academic vocabulary knowledge and their writing idea. In order to appraise students' academic writing competency and its relation with their lexical diversity, the current study examines student's writing through the frequent appearances of AWL from the samples.

There are 10 types of academic word sub-lists with 60 items for each list except list 10 with only 30 items, namely AWL 1 to AWL 10 (Coxhead, 2000). Sub-list one (AWL 1) contains the most frequent words used in academic writing, while sub-list ten (AWL 10), on the other hand, contains the least frequent words. Additionally, there are also academic phrases, namely K1 to K5 that distinguish the academic phrases that appear in academic writing. Moreover, according to Coxhead (2000), the more frequently the AWL appears in the text, the more

reliable the text as an academic writing text.

Table 3 below presents the analysis of AWL to determine students' academic writing competencies. The total value of tokens ranges from 408 tokens to 1670 tokens. Despite differences in percentages and AWL running words (tokens) between participants, it was found that most participants were demonstrating academic writing skills. For instance, participants number 1 to 6 were found to score a high level of AWL (n > 1000; >50%) from the standard score of the AWL are 570 words. It implies that those 6 participants performed academic writing competencies by using AWL appropriately. Moreover, there are only four participants who scored less than the overall AWL total, such as participants 17, 14, 11, and 18 with total AWL tokens 555 (3.40%), 470 (2.88%), 464 (2.84%), and 408 (2.50%) respectively. Consequently, the academic writing competencies of fourteen out of eighteen participants were found to be excellent.

A summary of a descriptive statistic of AWL is presented in Table 4, indicating the adult AWL level presented in diversity level with the minimum score of tokens 408 and types 273 to the highest score of tokens 1607 and types 10,003. The mean was observed at 78.80 with a standard deviation of 12.79. In addition, the mid score of AWLs was found in 779 in tokens and 435 in types, which demonstrates that most of the participants performed at academic writing level from mid to high competencies.

However, answering the second research question (RQ2. Does students' lexical diversity correlate with their use of academic words in an online academic discussion and writing assignment environment?), the student's academic writing competencies do not provide significant interconnection with their lexical diversity in the postsecondary setting. Table 2 above, it was presented that participant number 4 with the highest lexical diversity did not score the highest AWL level (n = 1119; 6.85%). Participant 16 with the lowest lexical diversity also did not score the lowest AWL level (n = 803; 4.92%). In addition, participant number 18 was found to have the lowest level of AWL (n = 408; 2.50%) although they scored pretty high lexical diversity ($D = 75.25$). Furthermore, participant number 5 was found to be presented with the highest level of AWL (n = 1670; 10.23%) regardless of their lexical diversity ($D = 87.92$). Hence, neither students' lexical diversity nor their academic writing competency determines each other values in academic writing.

Furthermore, to enhance the interpretation, a simple regression linear model was performed to test the relationship between the student's lexical diversity and their academic writing competencies as presented in Table 5 below. In order to run the regression analyses, the
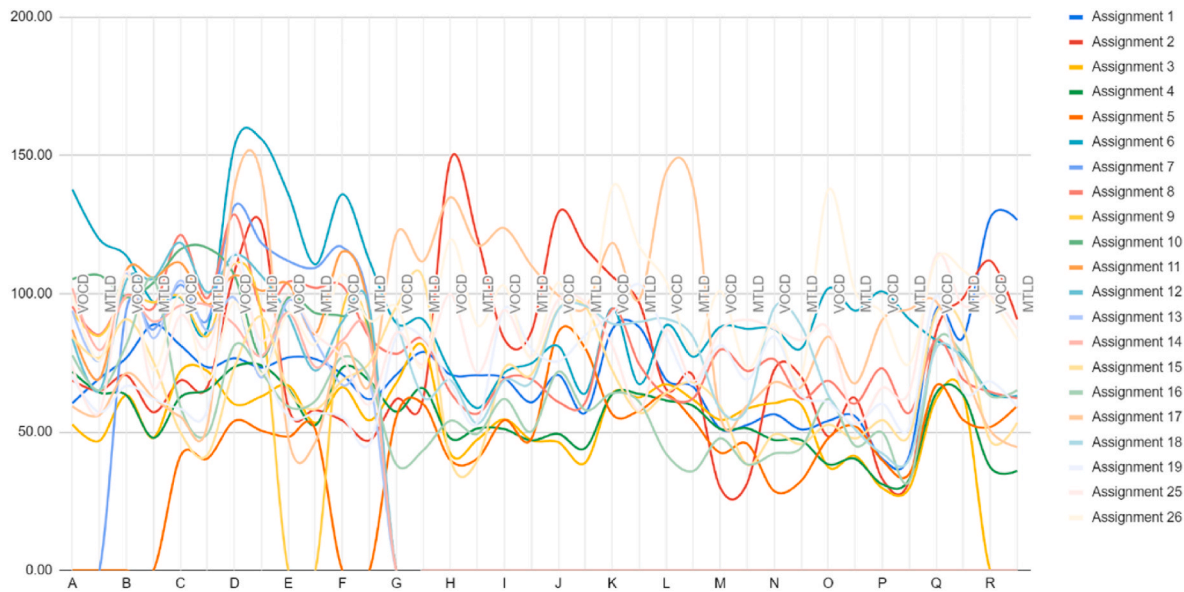
Fig. 2. D values of each participant for each session.

Table 3
The academic word value presented in the participant's writing.

Participant	Types		Tokens	
	Total	%	Total	%
A	756	7.67%	1260	7.72%
B	904	9.17%	1394	8.54%
C	1003	10.18%	1659	10.16%
D	732	7.43%	1119	6.85%
E	982	9.97%	1670	10.23%
F	908	9.21%	1431	8.76%
G	464	4.71%	792	4.85%
H	390	3.96%	687	4.21%
I	461	4.68%	765	4.68%
J	403	4.09%	607	3.72%
K	312	3.17%	464	2.84%
L	454	4.61%	760	4.65%
M	331	3.36%	614	3.76%
N	287	2.91%	470	2.88%
O	415	4.21%	873	5.35%
P	415	4.21%	803	4.92%
Q	364	3.69%	555	3.40%
R	273	2.77%	408	2.50%
Σ	9854		16,331	

Table 4
Summary of descriptive statistic of AWL.

	Types		Tokens	
Mean	547.44	5.56%	907.28	5.56%
Standard Deviation	255.99	2.60%	411.62	2.52%
Median	435	4.41%	779	4.77%
Min	273	2.77%	408	2.50%
Max	1003	10.18%	1670	10.23%

Table 5
Regression model analysis.

Regression Model Analysis			
F = 10,5	p = 0.185	R = 0.630	R^2 = 0.397
VIF = 1.00		Shapiro-Wilk = 0.929	
Predictor	β	SE	t
Intercept		498.37	−1.38
AWL	0.630	6.52	3.24

data was first examined for the assumption of its linearity and normality. The data was found to normally distributed (Shapiro-Wilk $p > 0.05$) and there is no collinearity was found (VIF <5).

The regression analysis indicated a low connection between the students' lexical diversity and their academic writing competencies (F = 10,5, $p = 0.185$, R^2 = 0.397). Therefore, emphasizing the answer to the second research question, it was revealed that there is no significant correlation between students' lexical diversity and their use of academic vocabulary. A high level of lexical diversity does not determine the students' academic writing competencies. Likewise, excellent academic writing competencies do not require a high-level lexical diversity.

In sum, the second research question was answered by JAMOVI and simple regression analyses proved lexical diversity *D* had a negative correlation with academic vocabulary use statistically (Wang, 2014). It was also noted that lexical diversity does not predict students' academic writing competencies. A high level of lexical diversity does not guarantee a sufficient level of academic writing competency. Likewise, excellent academic writing competencies do not require a high-level lexical diversity. Although it was claimed that the more frequently the AWL appears in the text, the more reliable the text as an academic writing text (Coxhead, 2000), it was found that there is no significant correlation between students' lexical diversity and their use of academic vocabulary. The frequent appearances of academic writing vocabulary do not determine the academic writing quality.

The above finding aligns with an earlier study conducted by Hassanzadeh et al. (2021), which suggests that while specific interventions may enhance learners' lexical diversity, this improvement does not necessarily lead to an overall enhancement in writing quality. Although, it should be emphasised that the relationship between students' lexical diversity and writing quality may become more pronounced when considering additional external variables, such as English Language Proficiency (ELP) designation (i.e., English Learner vs. Reclassified Fluent English Proficient) and gender as control variables (Maamuujav, 2021). It is essential to take these factors into account, as they can significantly influence writing performance. For instance, students classified as English Learners may have varying levels of exposure to academic language compared to their Reclassified Fluent English Proficient peers, which could affect their writing outcomes. More proficient writers tend to use a greater variety of vocabulary, which can contribute to more accurate and precise expression of ideas during the writing process (see Kim, 2014; Nasseri and Thompson, 2021).

4. Conclusion and recommendation

This article has examined (1) the level of lexical diversity among students in online discussion and academic writing assignments, and (2) the correlation between a student's lexicon diversity and the use of academic vocabulary in postsecondary settings. The analysis adopted parameter D to investigate the lexical diversity degree and AWL examiner to analyse the academic writing competencies using the *Text Inspector* software. The *Text Inspector* tools facilitate the researcher in measuring the students' lexical diversity from their online written discussion by applying the parameter D formula. The analysis of the sample from students' online written assignments data proved that lexical diversity *D* had a negative correlation with academic vocabulary use statistically. It was also noted that lexical diversity does not predict students' academic writing competencies. A high level of lexical diversity does not guarantee sufficient level of academic writing competency. The current study findings follow a study by Wang (2014), that claimed there was no statistically significant relationship between lexical diversity and the participants' academic vocabulary scores. The current study thus concludes that although lexical diversity has had positive relation to student writing, when there is a tendency toward academic quality of writing, it turns into negative.

However, as such lexical diversity has become an influential tool in measuring learners' language ability, The language learning activity used in this study was students' online written assignments in EFL class. Consequently, several external factors affect the linguistic diversity between students in this regard, e.g., topic selected, language proficiency background and experience in writing. In addition, the analysis process has been influenced by limited sample material collected during this study due to a large number of outliers. Therefore, to be able to fully examine the relationship between lexical diversity and academic writing performance more closely, it is appropriate to carry out an additional study with a deeper analysis involving more detail investigation units as well as larger numbers of samples.

Funding

CRediT authorship contribution statement

Beauty Sholeha Raufi: Writing – original draft, Validation, Resources, Investigation, Formal analysis, Data curation, Conceptualization. **Herri Mulyono:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Hamzah Puadi Ilyas:** Writing – review & editing, Validation, Supervision, Resources, Funding acquisition. **Siti Zulaiha:** Writing – review & editing, Validation, Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Akbari, N., 2017. Lexical diversity and the use of academic and lower frequency words in the academic writing of EFL students. Aust. Rev. Appl. Ling. 40 (1), 3–18. https://doi.org/10.1075/aral.40.1.02akb.

Alshehri, A., 2022. Writing proficiency of Saudi EFL learners: examining the impact of lexical diversity. Journal of Language and Linguistic Studies 18, 519–529. http://www.jlls.org.

Biber, D., Conrad, S., Reppen, R., 1994. Corpus-based approaches to issues in applied linguistics. Appl. Ling. 15 (2) https://doi.org/10.1093/applin/15.2.169.

Coxhead, A., 2000. A new academic word list. Tesol Q. 34 (2), 213. https://doi.org/10.2307/3587951.

Durán, P., Malvern, D., Richards, B., Chipere, N., Duran, P., 2004. Developmental trends in lexical diversity. Appl. Ling. 25 (2), 220–242. https://doi.org/10.1093/applin/25.2.220.

Hammou, B.A., Larouz, M., Fagroud, M., Ait Hammou, B., Larouz, M., Fagroud, M., 2021. Word frequency, range and lexical diversity: picking out changes in lexical proficiency among university learners in an EFL context. International Journal of Linguistics and Translation Studies 2 (2), 22–38. https://doi.org/10.36892/ijlts.v2i2.131.

Hassanzadeh, M., Saffari, E., Rezaei, S., 2021. The impact of computer-aided concept mapping on EFL learners' lexical diversity: a process writing experiment. ReCALL 33 (3), 214–228. https://doi.org/10.1017/S095834402100001X.

Kim, J., 2014. Predicting L2 writing proficiency using linguistic complexity measures: a corpus-based study. Engl. Teach. 69 (4).

Lim, M., 2017. Freedom to hate: social media, algorithmic enclaves, and the rise of tribal nationalism in Indonesia. Crit. Asian Stud. 49 (3), 411–427.

Maamuujav, U., 2021. Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. Assess. Writ. 49, 100540 https://doi.org/10.1016/j.asw.2021.100540.

Malvern, D., Richards, B., Chipere, N., Durán, P., 2004. Lexical Diversity and Language Development Quantification and Assessment, first ed. Palgrave Macmillan. https://doi.org/10.1057/9780230511804.

McCarthy, P., 2005. An Assessment of the Range and Usefulness of Lexical Diversity Measures and the Potential of the Measure of Textual, Lexical Diversity (MTLD). University of Memphis.

McCarthy, P.M., Jarvis, S., 2007. vocd: a theoretical and empirical evaluation. Lang. Test. 24 (4), 459–488. https://doi.org/10.1177/0265532207080767.

Nasseri, M., Thompson, P., 2021. Lexical density and diversity in dissertation abstracts: revisiting English L1 vs. L2 text differences. Assess. Writ. 47 (100511) https://doi.org/10.1016/j.asw.2020.100511.

Ritonga, R., Syahputra, I., 2019. Citizen journalism and public participation in the era of new media in Indonesia: from street to tweet. Media Commun. 7 (3), 79–90.

Shibata, N., 2022. The Effectiveness of theme-based instruction on Japanese university students' lexical diversity in writing. Nagoya JALT Journal 3 (1), 78–105. https://doi.org/10.37546/jaltchap.nagoya3.1-4.

Treffers-Daller, J., Parslow, P., Williams, S., 2018. Back to basics: how measures of lexical diversity can help discriminate between CEFR levels. Appl. Ling. 39 (3), 302–327. https://doi.org/10.1093/applin/amw009.

Vidal, K., Jarvis, S., 2020. Effects of English-medium instruction on Spanish students' proficiency and lexical diversity in English. Lang. Teach. Res. 24 (5), 568–587. https://doi.org/10.1177/1362168818817945.

Wang, X., 2014. The relationship between lexical diversity and EFL writing proficiency. University of Sydney Papers in TESOL 9, 65–88. http://www.lognostics.co.uk.

Yu, G., 2010. Lexical diversity in writing and speaking task performances. Appl. Ling. 31 (2), 236–259. https://doi.org/10.1093/applin/amp024.