COMPARATIVE ANALYSIS OF SYNTACTIC COMPLEXITY IN INDONESIAN GRADUATE-LEVEL WRITING: A CORPUS STUDY OF MASTER'S THESES VERSUS DOCTORAL DISSERTATIONS

Dhini Anjarwati
English Language Education, Faculty of Letters, Universitas Negeri Malang, Indonesia
Email: dhini.anjarwati.2202218@students.um.ac.id

Yazid Basthomi
English Language Education, Faculty of Letters, Universitas Negeri Malang, Indonesia
Email:ybasthomi@um.ac.id

Evynurul Lailiy Zen
English Language Education, Faculty of Letters, Universitas Negeri Malang, Indonesia
Email: evynurul.laily.fs@um.ac.id

Abstract: Syntactic complexity is a crucial determinant quality of academic writing made by English as a Foreign Language (EFL) students. Prior studies used large-grained measures to compare and examine the syntactic complexity of published research articles. On the other hand, graduate-level research on syntactic complexity is receiving less attention. To close the gap, this study used fine-grained measures to compare and analyze the syntactic complexity of doctorate dissertations and master theses submitted by Indonesian students. This study employed a corpus-based method in a quantitative design. The corpus data had two sub-corpora chosen using a stratified sample technique based on years. Those were 52 doctoral dissertation abstracts and 74 master thesis abstracts from English Language Education of Universitas Negeri Malang. The twelve measures of the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC) tool developed by Kyle (2016) were utilized to measure the phrasal complexity. The finding showed the characteristics of Indonesian graduate-level writing that they intended to utilize more dependents per direct object (standard deviation) measures and fewer dependents per nominal (standard deviation) measures in their writing. It also showed doctoral dissertations have higher quality than master theses which reflected significant difference in most of phrasal complexity measures. These outcomes offer syntactic complexity insights into the Indonesian context. The study highlighted the need for more attention to syntactic complexity at graduate-level writing to improve writing quality.
Keywords: *EFL, phrasal measures; academic writing; academic proficiency.*

INTRODUCTION

Syntactic complexity has gained much attention from scholars because it is beneficial for academic settings. It frequently investigated academic studies (Nasseri, 2021). Furthermore, it is necessary for students in higher-level education (Esfandiari & Ahmadi, 2021). Its role is to become the indicator of students' writing quality. Likewise, it is considered a measure of language performance, language maturity, and proficiency of L2/EFL writing quality (Casal & Lee, 2019; Yin, Gao & Lu, 2021). Therefore, syntactic complexity may help students to write academic writing as an indicator of their writing quality.

Syntactic complexity is a multidimensional construct. Alshalanee and Jaganathan (2023) stated that it consists of various degrees of sophistication, including global, clausal, and phrasal degrees. Syntactic complexity can be analyzed through syntactic complexity measures. The measures are separated into two primary categories: large-grained and fine-grained (Alsahlanee & Jaganathan, 2023). Thus, syntactic complexity has several dimensions that can be assessed through some measures.

Large-grained measures can analyze syntactic complexity. It also can be called traditional complexity measures (Zhang & Lu, 2021). Furthermore, these focus on dependent clauses for the necessary indicator (Esfandiari & Ahmadi,

2021). These measures were used in many previous studies. Thi and Nikolov (2023) determined syntactic complexity through large-grained measures. Likewise, However, the use of large-grained measures is still debatable. It is insufficient for operationalizing the linguistic features in academic writing since these cannot capture important non-clausal features in academic writing (Zaein & Golparvar, 2022). Furthermore, there is a lack of relationship between writing quality and clausal complexity (Esfiandiari & Ahmadi, 2021). Therefore, the weakness of large-grained measures can be addressed through syntactic complexity measures that focus on non-clausal features.

Syntactic complexity also can be analyzed through fine-grained measures. These measures focus on differentiating subtypes of phrase complexity level (Zhang & Lu, 2021). These measures can be accessed through TAASSC tool that was created by Kyle (2016). Furthermore, the fine-grained measures of TAASSC tool have stronger predictive power than large-grained measures (Zhang & Lu, 2021). Likewise, Phrase complexity has been determined to be the ideal criterion for identifying advanced proficient writing (Casal & Lee, 2019). Thus, fine-grained measures can explore syntactic complexity better in non-clausal features of academic writing.

Some previous studies investigated the comparison of syntactic complexity. Many of them focused on comparing the syntactic complexity of L1 writing to L2/EFL writing such as the study of Nasseri (2021). Some research also compared published research articles such as the study of Alsahlanee and Jaganathan (2023). However, few studies have compared the syntactic complexity of academic writing in higher education contexts except undergraduate studies (Esfiandiari & Ahmadi, 2021). Likewise, investigating unpublished academic writing written by master and doctoral students has less attention from researchers (Dong, Hao, & Buckingham, 2022). Therefore, scholars still pay less attention to a comparison study of syntactic complexity in graduate students' writing.

Academic writing is an essential requirement for graduate students. Esfiandiari and Ahmadi (2021) mentioned that it could be published research articles, dissertations, or theses. Furthermore, these are considered the last evaluations of the student's coursework (Wisker, 2019). Likewise, it is also treated as writing proficiency to fulfill graduation requirements (Huang, 2024). Furthermore, it enables students to

further enhance their academic skills (Tuononen & Parpala, 2021). In addition, it requires writing insight during composing academic writing which brings benefit to students' academic competence (Ahsanuddin et al., 2022; Rofiqoh et al., 2021). Thus, it has an important role for graduate students.

Besides the benefits of academic writing, some previous studies highlighted the difficulties of EFL graduate students in fulfilling writing proficiency. Likewise, writing proficiency is a significant challenge for constructing academic writing (Subandowo & Sardi, 2023). Indonesian students who use EFL are not an exception. Indonesian students at tertiary institutions face challenges in constructing academic writing (Sahan, Saridewi, Wabang, & Nabung, 2024). Furthermore, Subandowo and Sardi (2023) stated that Indonesian graduate students have a challenge to produce high-quality academic writing. They also mentioned that Indonesian students need linguistic resources to guide them in creating high-quality academic writing for the graduate level. However, there were limited studies that explored EFL graduate students' academic writing (Huang, 2024). Furthermore, these challenges can cause poor quality and low productivity of academic writing (Alsahlanee & Jaganathan, 2023; Casal & Lee, 2019). These challenges highlight the need for further research into linguistic characteristics, particularly syntactic complexity, as a representation of writing quality or competency in the Indonesian setting.

Syntactic complexity in graduate-level academic papers has been studied in some prior studies. Esfiandiari and Ahmadi (2021) conducted a comparative analysis of three types of Iranian academic writing: published research articles, doctorate dissertations, and master's theses. The outcomes proved that the prediction of academic writing was significantly improved by fine-grained measures of phrase complexity level. Furthermore, Nasseri (2021) investigated syntactic complexity in master theses of EFL, ESL, and English L1. She claimed the findings of earlier research on syntactic complexity were inconsistent. She also argued that it can be impacted by the writers' linguistic proficiency and English language backgrounds. Thus, the results of syntactic complexity investigation may be different at graduate-level.

In the prior research and literature, there is an empirical gap. The study on comparing syntactic complexity at the graduate level of academic

writing gained less attention from scholars (Esfiandiari & Ahmadi, 2021). To fill the gap, the current study conducted a corpus study and compared the syntactic complexity of master thesis abstracts and doctorate dissertation abstracts produced by Indonesian writers using twelve fine-grained measures created by Kyle (2016) and that were shown to be trustworthy indices (Kyle & Crossley, 2018). Furthermore, a corpus study is necessary to identify the most common syntactic patterns in various kinds of academic writing (Nasseri, 2021). Doing comparative analysis helps us to see the difference in phrasal features utilized by master students and doctoral students.

Examining the different skills of people from various backgrounds is not a surprising fact, but it creates the questions of what metrics would be most useful in explaining the differences (Rahayu, Utomo & Setyowati, 2021). Therefore, this study has three primary aims as follows: (1) To what extent do master students and doctoral students utilize phrasal measures? (2) Is there a significant difference in phrasal syntactic complexity measures between master theses and doctoral dissertations? (3) What are the significant phrasal complexity measures utilized by master students and doctoral students?

METHOD
This study compared the syntactic complexity of master's thesis abstracts and doctoral dissertation abstracts using a comparative quantitative design of a corpus-based analysis. Two sub-corpora comprise the corpus data of this study, including master thesis abstracts and doctoral dissertation abstracts. The data were gathered from English Language Education Department at Universitas Negeri Malang and written by Indonesian students. This abstract section was chosen because it is an important part that should be considered by authors (Arianto, et al., 2021; Budiyono & Fadhly, 2023). Thus, the corpus data were abstract compilations written by Indonesian graduate students.

The corpus data were obtained from the library website database at Universitas Negeri Malang. Some criteria were employed to select data to keep the homogeneous data. Furthermore, academic language patterns may change over time (Yin, Gao, & Lu, 2023). So, this study applied some criteria for obtaining data. Because of various focuses, this study classified focus data into English Language Teaching, Assessment, English Specific Purposes, Second Language

Acquisition, and Applied Linguistics. Furthermore, all the texts were also published between 2018 and 2022. These texts should also follow IMRAD (Introduction-Method-Results-Discussion) format. Therefore, this study obtained homogeneous data using some standards.

Furthermore, this study tried to create a good corpus. data selection was conducted through a stratified sampling technique. It is a better method for selecting data (Zufferey, 2022, p. 145). The abstracts were divided based on years of publication. The data were classified into five groups: 2018, 2019, 2020, 2021, 2022. Then, selecting data for each stratum employed simple random sampling (see Table 1). Furthermore, Zufferey (2022, p. 145) mentioned that there is no perfect corpus size. However, it should have phenomena of representative language. Thus, this study conducted representative standards based on Carradini and Swarts (2023, p. 65), including diversity, balancedness, and saturation.

This study followed diversity, balancedness, and saturation aspect. Diversity refers to the data that is obtained from various sources. This study tried to follow diversity by obtaining data from five different publication years. Furthermore, balancedness aspect is about how data is sampled from sources. This study followed this aspect by using systematic stratified sampling to select data from each stratum. In addition, saturation refers to the equal size of tokens. This study tried to build an equal number of the total tokens. Therefore, this research built a corpus based on three aspects of the representative language phenomena.

Table 1. *Corpus data*

Year (Stratum)	Type	Number of abstracts	Means of abstract words	Number of tokens
2018	S2	17	443	7531
2019		18	441	7938
2020		17	445	7565
2021		17	443	7531
2022		5	447	2235
Total		74	443	32800
2018	S3	12	629	7548
2019		9	632	5688
2020		12	631	7572
2021		12	629	7548
2022		7	634	4438
Total		52	630	32794

Note. S3: Doctoral dissertation; S2: Master Thesis

This study applied TAASSC tool to evaluate syntactic complexity of abstract sections in doctoral dissertations and master theses. Some fine-grained measures of TAASSC tool are

confirmed that it can analyze academic writing quality. Zhang and Lu (2021) claimed that some phrasal complexity measures from this tool are correlated to writing quality. Therefore, this study utilized twelve phrasal complexity measures of TAASSC tool from Alsahlanee and Jaghanathan (2023). These measures can predict the writing quality (Kyle & Crossley, 2018). These measures can be seen in Table 2.

Table 2. *Phrasal complexity measures in TAASSC tool*

Phrasal measures/ measure names	Functions
dependents per nominal/ av_nominal_deps	calculating modifiers that transform a noun or noun phrase
dependents per direct object/ av_dobj_deps	addressing dependents in the direct object compositions
dependents per object of the preposition/ av_pobj_deps	evaluating modifiers that make up the preposition's object
dependents per nominal (standard deviation)/ nominal_deps_stdev	evaluating the variety of modifiers for producing noun phrases
dependents per nominal subject (standard deviation)/ nsubj_stdev	evaluating a wide variety of dependents that comprise a nominal subject
dependents per direct object (standard deviation)/ dobj_stdev	assessing the range of modifiers in direct object phrase
dependents per object of the preposition (standard deviation)/ pobj_stdev	counting the various kinds of dependents for creating objects of the preposition
determiners per nominal/ det_all_nominal_deps_st ruct	examining determiners in noun phrase modifications
prepositions per nominal/ prep_all_nominal_deps_ struct	exploring prepositions that modify noun phrase or noun
adjectival modifiers per object of the preposition/ amod_pobj_deps_struct	counting adjective dependents that compose the object of the preposition
adjectival modifiers per direct object/ amod_dobj_deps_struct	analyzing adjectival modifiers in direct object constructions
prepositions per object of the preposition/ prep_pobj_deps_struct	evaluating prepositions in objects of the preposition compositions

This study used the Statistical Package for the Social Sciences (SPSS) program to perform descriptive and inferential statistics for addressing three research questions. The initial research question was addressed using descriptive analysis to explore the extent to which master and doctoral students utilized phrasal features. The mean values of twelve measures provided data for master theses and doctoral dissertations. To answer the second research question, this study applied one-way multivariate analysis of variance (MANOVA) of inferential statistics to discover whether there are significant distinctions between master theses and doctoral dissertations regarding the 12 phrasal complexity measures. To answer the third research question, this study employed Analysis of Variance (ANOVA) using the Bonferroni method. Before conducting MANOVA and ANOVA, homogeneity and normality tests were performed to examine the data. The analysis demonstrated that every measure of this study produced homogeneity and normal distribution findings.

RESULTS AND DISCUSSION
This section explores the syntactic complexity in abstracts of master's theses and doctorate dissertations.

Characteristics of Indonesian graduate-level writing
This study revealed postgraduate students' traits based on 12 phrasal complexity measures. These measures were analyzed using the TAASSC tool. Then, the SPSS tool was applied to perform descriptive statistical analysis (see Table 3).

Table 3. *Descriptive statistics of phrasal complexity measures*

Measures name	N	Mean	
		S3	S2
av_nominal_deps	124	1.67	1.44
av_dobj_deps	124	1.2	1.12
av_pobj_deps	124	1.61	1.49
nominal_deps_stdev	124	0.24	0.21
nsubj_stdev	124	1.14	1.09
dobj_stdev	124	1.78	1.57
pobj_stdev	124	0.38	0.31
det_all_nominal_deps_ struct	124	0.33	0.34
prep_all_nominal_deps _struct	124	1.22	1.16
amod_pobj_deps_struct	124	0.29	0.24
amod_dobj_deps_struct	124	0.38	0.31
prep_pobj_deps_struct	124	1.09	1.03

Note: S2= Master theses; S3= Doctoral dissertations

Both master theses and doctoral dissertations utilized a wide range of modifiers for modifying

direct objects. It can be seen that the highest mean value of measures is dependents per direct object (standard deviation). This measure determines the frequency of modifiers of direct objects. The example of dependents per direct object (standard deviation) measure shows a sentence (see Figure 1). The direct object is *data* with one modifier type which is *the*. The high mean value indicates that Indonesian graduate students employed various dependents to compose direct object phrases in their academic writing. It can be concluded that many types of modifiers for composing direct objects are crucial phrase features for graduate-level writing.

Figure 1. *Dependents per direct object (standard deviation)*

the	researcher	analyzed	the	data
	{det}	{dobj}		

Note: det= determiner; dobj= direct object

Furthermore, the high mean value of dependents per direct object (standard deviation) outcome supports the finding of Alsahlanee and Jaghanathan (2023). Their study investigated the Iraqi writing and L1 writing through twelve phrasal complexity measures. They found that the mean value of dependent per direct object (standard deviation) was also high. This means that various modifiers for composing direct are essential features in constructing arguments of academic writing. Thus, graduate students should master a wider range of dependents in composing process of direct objects.

Figure 2. *Dependents per nominal (standard deviation)*

an	argumentative	writing	prompt
{det}	{amod}	{nn}	{N}

Note: amod= adjective as modifier; nn= noun as modifier; N= nominal

Besides the highest value of measures, master's and doctoral students employed fewer dependents per nominal (standard deviation). This is because the least mean value of the twelve measures is dependents per nominal (standard deviation). This measure counts various dependent types for creating noun or noun phrases. This study provides an example of this measure (see Figure 2). The nominal is *prompt* with three modifiers, including determiner (*an*), adjective as modifier (*argumentative*), and noun as modifier (*writing*). It can be interpreted that master's and doctoral students barely apply many modifier types in noun phrases or nouns. Therefore, master theses and doctoral dissertations barely need various kinds of modifiers in delivering an argument of writing.

The quality of master theses and doctoral dissertations

This study found that the level of doctoral dissertations is higher than master's theses based on phrasal complexity. The quality of writing can be determined using the complexity of phrase structures. Likewise, the high mean value of the phrasal measures illustrates the high-quality writings (Kyle & Crossley, 2018). The means of phrasal complexity in Table 3 shows that doctoral dissertations are more complex than master theses in eleven out of twelve measures. This means that doctorate dissertations have higher quality than master theses.

However, determiners per nominal measure outcomes present that the mean of master theses (0.34) is higher than doctoral dissertations (0.33). This measure explores determiners for creating noun phrases. Figure 3 explains a nominal (*study*) that has a modifier (*their*). The high value of master theses illustrates that master students utilize more determiners to modify noun phrases than doctorate students.

Figure 3. *Determiners per nominal*

their	future	study
{det}		{N}

The extensive usage of determiners in noun phrase compositions supports the findings of Alsahlanee and Jaganathan (2023) who reported that lower-proficiency writers typically utilize more determiners in noun phrase constructions than higher-proficiency writers. This is influenced by writers who have different writing proficiency levels. Similarly, Esfandiari and Ahmadi (2021) argued that writing proficiency may develop to become mastered in certain aspect that influences the utilization of syntactic features. It means that doctoral students have already mastered the determiners of applying or not applying them in noun phrase constructions. However, master students still have lower writing proficiency than doctoral students. They also employ more determiners to construct noun phrases. So that they can avoid making errors that they might not understand.

The significant difference in phrasal complexity measures in graduate-level writing

This study found a significant difference in phrasal features between abstracts of master theses and doctoral dissertations. MANOVA was performed to analyze the significance of 12

phrasal complexity. If the p-value is less than 0.05, the data value is accepted as a significant difference. Table 4 shows the significance values of multivariate test results are < 0.05. It indicates the writing proficiency gaps between graduate student types. Furthermore, it influences the production of writing quality.

Table 4. *The outcomes of multivariate tests*

Tests	Value	F	Hypothesis df	Error df	Sig.
Pillai's Trace	0.24	9.72	12.00	367.00	0.00
Wilks' Lambda	0.76	9.72	12.00	367.00	0.00
Hotelling's Trace	0.32mA	9.72	12.00	367.00	0.00
Roy's Largest Root	0.32	9.72	12.00	367.00	0.00

The significant difference of phrasal complexity measures between master thesis abstracts and doctoral dissertation abstracts is aligned with Esfandiari and Ahmadi (2021). They compared abstract sections of master thesis and doctoral dissertation that were written by Iranian students. They identified a significant distinction in the syntactic difficulty of doctorate dissertations and master's theses in academic writing. The writing proficiency of writers causes this difference. Likewise, linguistic proficiency of writers influences the difference of syntactic complexity (Nasseri, 2021). Doctoral students who have higher writing proficiency (see Table 3), have more experienced writing than master students. Likewise, the different background may influences inconsistent outcomes (Nasseri, 2021; Zen, 2020). Therefore, both master and doctoral students have different background of writing that impacts the difference in the case of syntactic complexity.

To detect the exact measure that exhibits significant differences, the post hoc or ANOVA test was utilized through the Bonferroni method. The outcomes are categorized significantly differences when the p-value is < 0.05. Nine out of twelve phrasal complexity measures were significant differences between master theses and doctoral dissertations (see Table 5). Only three measures reveal no significant difference. It can be interpreted that the quality of the master thesis abstract and doctoral dissertation abstract is still significantly distinct.

Table 5. *Pairwise comparison*

Dependent variable	Academic Writing	Academic Writing	Mean difference	Std. Error	Sig
av_nominal_deps	S3	S2	.083	.024	.001
	S2	S3	-.083	.024	.001
av_dobj_deps	S3	S2	.072	.038	.064
	S2	S3	-.072	.038	.064
av_pobj_deps	S3	S2	.092	.026	.000
	S2	S3	-.092	.026	.000
nominal_deps_stdev	S3	S2	.044	.020	.029
	S2	S3	-.044	.020	.029
nsubj_stdev	S3	S2	.041	.042	.333
	S2	S3	-.041	.042	.333
dobj_stdev	S3	S2	.082	.032	.010
	S2	S3	-.082	.032	.010
pobj_stdev	S3	S2	.069	.026	.011
	S2	S3	-.069	.026	.011
det_all_nominal_deps_struct	S3	S2	-.007	.040	.864
	S2	S3	.007	.040	.864
prep_all_nominal_deps_struct	S3	S2	.034	.012	.004
	S2	S3	-.034	.012	.004
amod_pobj_deps_struct	S3	S2	.062	.019	.001
	S2	S3	-.062	.019	.001
amod_dobj_deps_struct	S3	S2	.063	.030	.035
	S2	S3	-.063	.030	.035
prep_pobj_deps_struct	S3	S2	.039	.014	.006
	S2	S3	-.039	.014	.006

The outcomes are significant at p-value< 0.05

However, three phrasal complexity measures present no significant difference between abstracts of master theses and abstracts of doctoral dissertations, including dependents per direct object, dependents per nominal subject

(standard deviation), and determiners per nominal. The p-values of the three measures show more than 0.05. Dependents per direct object is the initial measure that shows no significant difference regarding phrasal feature utilization between master students and doctoral dissertations. It counts dependents for constructing direct object phrase type. Figure 5 portrays the instance of dependents per direct object measure. There are two modifiers (*a; phenomenology*) for modifying direct object (*research*). The no significant difference value of this measure indicates that master and doctoral students already have the same understanding of modifiers for creating direct object phrases.

The	researcher	designed	a	phenomenology	research
			{det}	{nn}	{dobj}

Figure 5. *Dependents per direct object*

Furthermore, the dependents per nominal subject (standard deviation) measure reveals no significant distinct between the two graduate-level writing types. It evaluates the occurrences of dependent types to compose nominal subject phrases. The example shows a phrase with modifiers (see Figure 6). The nominal subject is *feedback* and the modifiers are *the; two; corrective*. Thus, master's and doctoral students also have similar comprehension for utilizing modifiers for direct object compositions.

Determiners per nominal measure also performs no significant difference outcome between doctoral and master students. This measure explores the occurrences of determiners for modifying nominal. Figure 7 shows the example of nominal with determiners. The nominal is *interview* and the determiner is *the*. No significant difference in this measure indicates that both graduate student levels have almost equal writing proficiency for utilizing determiners for composing nominal.

Figure 6. *Dependents per nominal subject (standard deviation)*

the	two	corrective	feedbacks	are inserted
{det}	{amod}	{amod}		

{nsubj}

Figure 7. *Determiner per nominal*

the	retrospective	interview
{det}		{N}

The outcomes of no significant difference in phrasal complexity are aligned with Alsahlanee and Jaganathan (2023). They compared Iraqis/ EFL research articles to L1 research articles that have difference writing proficiency. Their study found no significant difference in two phrasal complexity measures. The finding shows that both writer types already have similar comprehension for utilizing certain phrasal features in Iraqi contexts. Therefore, students should considerate and improve their writing proficiency gap between high-proficiency writers and lower proficiency writers.

CONCLUSION
The analysis and comparison of phrasal complexity measures of Indonesian graduate-level writing were conducted in this study. The descriptive statistics showed the characteristics of academic writing written by Indonesian graduate students. They tend to use more various types of modifiers to modify direct object composition. However, they require fewer dependents to compose nominal constructions. The descriptive analysis also showed doctoral students have higher writing proficiency and writing quality than master students. However, the master students utilize more determiners to modify nominal than doctoral students.

Besides the descriptive analysis, the inferential analysis showed a significant gap between master theses and doctoral dissertations. Furthermore, nine out of 12 phrasal complexity measures showed significant differences between both writing types. However, three measures revealed no gap between both writing types, including dependents per direct object, dependents per nominal subject (standard deviation), and determiners per nominal. Thus, there are gaps and similarities in linguistic comprehension between master and doctoral students for utilizing modifiers in composing noun phrase constructions

Based on the findings, this study can contribute to the linguistics field of syntactic complexity context. It provides the academic writing characteristics of Indonesian graduate students in terms of phrasal complexity measures. This study also revealed the gap in phrasal complexity utilization between Indonesian master students and Indonesian doctoral dissertations that should be addressed to compose high-quality academic writing. Thus, these findings are

beneficial linguistic insights for the Indonesian graduate level.

While this study has investigated phrasal syntactic complexity in master theses and doctoral dissertations, this study has limitations. Due to the available data from the database, each corpora stratum of master thesis and doctoral dissertation consists of different sizes of tokens. The differences influence corpus saturation aspects. In addition, the master thesis and doctoral dissertation have different numbers of words that impact the unequal tokens of balancedness aspect between master thesis corpora and doctoral dissertation corpora. Furthermore, the representative updated data may transform because students' academic writing increases over time. It can be identified that the shortcomings of this study are linked to the corpus's representative phrasal attributes.

Based on the weakness, this study provides some recommendations for future studies. They can compare Indonesian published research articles between master's and doctoral students. An equal number of abstract words can capture the representative phrasal features to establish a quality corpus. In addition, they can investigate updated corpus data to capture current linguistic phenomena that can help graduate students create high-quality writing. The study outcome will contribute crucial linguistic insight for writing ability improvement in the Indonesian context.

REFERENCES

Ahsanuddin M., Hanafi Y., Basthomi Y., Taufiqurrahman F., Bukhori H.A., Samodra J., Widiati U., Wijayati P.H. (2022). Building a corpus-based academic vocabulary list of four languages. *Pegem Journal of Education and Instruction*, 12(1), 159-167. https://doi.org/10.47750/pegegog.12.01.15

Alsahlanee, A., & Jaganathan, P. (2023). Phrasal syntactic complexity measures in linguistics research articles written by Iraqi and English L1 writers. *3L: Language, Linguistics, Literature*, 29(1), 154–168. https://doi.org/10.17576/3L-2023-2901-11

Arianto, M. A, Sukhah, A. Basthomi, Y., Wulyani, A.N. (2021). Indonesian doctoral students', indonesian academics', and international authors' research gap strategies in ELT Research Article Abstracts and Introductions. *Journal of Language & Education,* 7(2). 25-44. https://doi.org/10.17323/jle.2021.11735

Budiyono, S., & Fadhly, F. Z. (2023). A qualitative evidence synthesis of article abstract writing in ELT and literature journals. English Review: *Journal of English Education*, 11(1),

253-262. https://doi.org/10.25134/erjee.v11i1.7753

Casal, J. E., & Lee, J. J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing*, 44, 51–62. https://doi.org/10.1016/j.jslw.2019.03.005

Carradini, S. & Swarts, J. (2023). *Text at scale: corpus analysis in technical communication.* Colorado: University Press of Colorado

Dong, J., Hao, W., & Buckingham, L. (2022). Mapping out the disciplinary variation of syntactic complexity in student academic writing. *System*, 113. 102974. https://doi.org/10.1016/j.system.2022.102974

Esfandiari, R., & Ahmadi, M. (2021). Syntactic complexity measures and academic writing proficiency: A corpus-based study of professional and students' prose. *The Journal of Asia TEFL, 18*(3), 745–763. https://10.18823/asiatefl.2021.18.3.1.745

Huang, J.C. (2024). Unveiling EFL graduate students' EAP needs and perceptions of EAP courses: A large-scale survey in Taiwan. *Journal of English for Academic Purposes*, 68, 101348. https://doi.org/10.1016/j.jeap.2024.101348

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine-grained indices of syntactic complexity and usage-based indices of syntactic sophistication.* Dissertation. Georgia State University. https://doi.org/10.57709/8501051

Kyle, K., & Crossley, S. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), 333–349. https://doi.org/10.1111/modl.12468

Nasseri, M. (2021). Is postgraduate English academic writing more clausal or phrasal? Syntactic complexification at the crossroads of genre, proficiency, and statistical modelling. *Journal of English for Academic Purposes*, 49, 100940. https://doi.org/10.1016/j.jeap.2020.100940

Rahayu, F. E. S., Utomo, A., & Setyowati, R. (2021). Syntactic and lexical complexity of undergraduate students' essays: a comparison study between l1 and l2 writings. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, 5(2), 251. https://doi.org/10.21093/ijeltal.v5i2.706.

Rofiqoh, Basthomi, Y., Widiati, U., Puspitasari, Y., Marhaban, S., & Sulistyo, T. (2022). Aspects of writing knowledge and EFL students' writing quality. *Studies in English Language and Education*, 9(1), 14-29. http://dx.doi.org/10.24815/siele.v9i1.20433

Sahan, A., Saridewi, M. P., Wabang, R. J., & Nabung, A. (2024). Implementation of android based augmented reality in English education.

English Review: Journal of English Education, 12(1), 269-282. https://doi.org/10.25134/erjee.v12i1.8595

Subandowo, D., & Sárdi, C. (2023). Academic essay writing in an English medium instruction environment: Indonesian graduate students' experiences at Hungarian universities. *Ampersand*, 11, 100158. https://doi.org/10.1016/j.amper.2023.100158

Thi, N.K., Vo, D.V., & Nikolov, M. (2023). Investigating syntactic complexity and language-related error patterns in EFL students' writing: corpus-based and epistemic network analyses. *Language Learning in Higher Education*, 13, 127-151. https://doi.org/10.1515/cercles-2023-2009

Tuononen, T., & Parpala, A. (2021). The role of academic competences and learning processes in predicting bachelor's and master's thesis grades. *Studies in Educational Evaluation*, 70, 101001. https://doi.org/10.1016/j.stueduc.2021.101001

Wisker, G. (2019). *developing scientific thinking and research skills through the research thesis or dissertation*. In: Murtonen, M., Balloo, K. (eds) Redefining scientific thinking for higher education. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-24215-2_9

Yin, S., Gao, Y., & Lu, X. (2021). Syntactic complexity of research article part-genres: Differences between emerging and expert international publication writers. *System*, 97, 102427. https://doi.org/10.1016/j.system.2020.102427

Zen, E.L. (2020). Role of regional language background and speech styles on the production of Voice Onset Time (VOT) in English among Indonesian Multilinguals. *Indonesian Journal of Applied Linguistics,* 10, 359-368. http://dx.doi.org/10.17509/ijal.v10i2.28604

Zhang, X., & Lu, X. (2022). Revisiting the predictive power of traditional vs. fine-grained syntactic complexity indices for L2 writing quality: The case of two genres. *Assessing Writing*, 51, 100597 https://doi.org/10.1016/j.asw.2021.100597

Ziaeian, E., & Golparvar, S. E. (2022). Fine-grained measures of syntactic complexity in the discussion section of research articles: The effect of discipline and language background. *Journal of English for Academic Purposes*, 57, 101116. https://doi.org/10.1016/j.jeap.2022.101116

Zufferey, S. (2020). *Introduction to Corpus Linguistics*. London: ISTE Ltd