

The Interplay of Linguistic Complexity Measures and Their Relation to the Quality of L2 Argumentative Writing

Wang Yanxin^{1,a,*}

¹*School of International Studies, Zhejiang University, Hangzhou, 310058, China*

a. yanxin_wang@zju.edu.cn

**corresponding author*

Abstract: Previous research revealed that syntactic and lexical complexity is crucial in predicting writing quality. However, measuring these features is difficult because they often interact in unpredictable ways. To address this challenge, measurement practices need to capture a more holistic view. This study aims to explore the integration of the complexity metrics and examines how they relate to L2 writing. A sample of 199 argumentative essays from an in-house English proficiency test was scrutinized using the Syntactic Complexity Analyzer and the Lexical Complexity Analyzer. Factor analysis and Pearson correlation were used for data analysis. The findings indicated that, syntactically, a higher frequency and richness of coordinate phrases were associated with superior L2 writing quality ($r=.27, p<.01$). Lexically, verb variation ($r=.186$), transformations of the number of different words measures ($r=.241$) and noun & mod variation ($r=.233$) were found to be valid indicators of L2 writing quality. However, different from previous research on narrative writing, verb sophistication is found to have no significant correlation with argumentative writing scores. It indicates that the genre of L2 production is a variable that may affect the validity of verb sophistication as a descriptor of L2 performance.

Keywords: syntactic and lexical complexity, second language writing, writing performance, argumentative essays

1. Introduction

Linguistic complexity can be regarded as a valid descriptor of second language (L2) performance, which can measure learners' proficiency and language development [1]. Bulté and Housen proposed that there are two major sub-components that make up linguistic complexity, lexical complexity and grammatical complexity. Grammatical complexity can be further divided into syntactic and morphological complexity [2]. According to O'Leary and Steinkrauss, however, morphological complexity is more noticeable in the early stages of L2 development than in English writing by advanced students [3]. Therefore, the current research will focus on syntactic and lexical complexity as the target measures to forecast the level of L2 English writing quality.

The variety and complexity of syntactic structures are referred to as syntactic complexity, which is an important component of learning a second language [2, 4, 5]. Lexical variety, lexical density, and lexical sophistication are the three primary markers of lexical complexity, often known as lexical richness [6]. Prior studies have demonstrated the significance of linguistic traits, such as lexical and syntactic complexity, in predicting writing quality. Noun phrase modification, mean sentence length,

and mean T-unit length are all important predictors of writing quality in argumentative essays written by non-native English speakers, according to studies by Taguchi et al. and Yang et al. [7, 8]. Engber examined the relationships between various lexical measures of writings and their overall scores and found a significant association between the score and error-free lexical variety [9]. According to Laufer and Nation's research, the Lexical Frequency Profile, a metric for lexical richness in writing, may be beneficial for identifying the elements that influence assessments of writing quality [10]. Lexical sophistication and overall writing scores were found to be highly associated with Maamujav's research [11].

The developmental character of complexity measures and the difficulties of measurement have limited earlier studies on the relationship between independent syntactic and lexical complexity measures and writing proficiency. These studies lack stability and are hindered by the arbitrary selection of complexity measures. To overcome these challenges, measurement practices should evolve to capture a more integrated view of complexity, rather than focusing on isolated metrics. This study aims to reduce the dimensionality of syntactic and lexical features and examine how the integrated linguistic features relate to L2 English argumentative writing performance.

The subsequent section will focus on the following specific research questions: 1) Which combination of syntactic complexity measures can best predict the quality of L2 argumentative English writing? 2) Which combination of lexical complexity measures can best predict the quality of L2 argumentative English writing? 3) To what extent do factor variables of the complexity measures predict second language writing performance?

2. Methodology

2.1. Samples

In this paper, 200 argumentative essays written by Chinese English learners at Zhejiang University were selected from the writing database of the Zhejiang University English Proficiency Test (ZJU EPT). The average word length of the essays was 190, with the shortest and longest totalling 116 and 284 words, respectively. The articles were scored by two automatic scoring systems, i-write and Pigaiwang, respectively. Both scoring systems had a full score of 100 points, and the final scores were averaged. One outlier was deleted based on the normality test (N=199).

2.2. Instrument

This research applied L2 Syntactic Complexity Analyzer (L2SCA), the Lexical Complexity Analyzer and SPSS (version 27) as analyzing tools [12-14].

2.2.1. Syntactic Analysis

To evaluate these writings' syntactic elements, this paper used the computational tool L2SCA because measures in L2SCA were generated directly from the study of the syntactic complexity of bilingual writing, which makes it an ideal tool for the research subject [13].

Lu's analyzer obtains 9 frequency indicators and 14 rate indicators. This paper excluded some indicators that do not have a strong correlation with writing quality based on previous research [12]. Specifically, clauses per sentence (C/S) is considered as clausal embeddings measure that takes both subordination and coordination into account. However, subordination and coordination are two different constructs that are usually measured by clauses per T-unit (C/T) and T-unit per sentence (T/S) respectively [15]. C/T and T/S were thus excluded.

As a result, the study's target L2SCA measurements comprised 21 measures. These 21 measures are clause (C), T-unit (T), complex T-unit (CT), dependent clause (DC), clause per sentence (C/S),

verb phrase (VP), verb phrase per T-unit (VP/T), dependent clause per clause (DC/C), dependent clause per T-unit (DC/T), complex T-unit ratio (CT/T), coordinate phrase (CP), coordinate phrase per T-unit (CP/T), coordinate phrase per clause (CP/C), word count (W), sentence (S), mean length of sentences (MLS), mean length of T-unit (MLT), mean length of clause (MLC), complex nominal (CN), complex nominal per T-unit (CN/T), complex nominal per clause (CN/C). The precise formula is documented in the research article by Lu in 2010 [13].

2.2.2. Lexical Analysis

The Lexical Complexity Analyzer (LCA) was utilized as the tool to assess lexical complexity in this study [12,14]. This analyzer provides 25 distinct measures of lexical density, variation and sophistication. According to previous research, some metrics may not be valid descriptors of L2 performance and were therefore eliminated from the analysis. Lexical density (LD) was shown to not significantly correlate with overall L2 writing evaluations [1, 16]. A number of different words (NDW) and Type–token ratio (TTR) was found both sensitive to the size of the sample [14]. Mean segmental TTR (MSTTR) might cause a waste of data since only texts of equal length could be used for analysis [17]. Four transformations of TTR measures showed unsatisfactory validity and reliability in L2 performance [18]. Besides, lexical word variation, adjective variation and adverb variation were eliminated to avoid redundant calculation. Therefore, a total of 14 measures generated from LCA were taken into account. These 14 measures include two lexical sophistication measures which are labelled as LS1 and LS2, two verb sophistication measures which are labelled as VS1 and VS2, corrected VS1 (CVS1), three NDW measures including NDW (first 50 words), NDW (expected random 50) and NDW (expected sequence 50) and four verb variation factors including verb variation-I (VV1), squared VV1 (SVV1), corrected VV1 (CVV1), verb variation-II (VV2), noun variation (NV) and modifier variation (ModV). The precise formula is documented in the research article by Lu in 2012 [14].

2.2.3. Data Analysis

To evaluate the relationship between syntactic and lexical variables and their effects on writing quality, data analysis was done using SPSS (version 27). Initially, a factor analysis was performed on the syntactic complexity and lexical complexity measures to reduce the dimensionality of the metrics. This process was essential for identifying the important factors that influenced writing performance and thereby simplifying the interpretation of the results. The degree and direction of the association between the integrated factors and the averaged writing scores were subsequently investigated using Pearson Correlation tests.

3. Results

3.1. Writing Performance

In this study, two online automated writing assessment platforms, i-write and Pigaiwang, were employed to evaluate students' essays (N=200). The correlations between the scores assessed by these two platforms were found to be statistically significant ($r=.72$, $p<.001$), indicating that the scores obtained from both platforms were valid and reliable for evaluating student writing performance. Consequently, the final writing grade for each student was derived as the average of the ratings provided by the two platforms ($M=69.6$; $SD=5.4$). After conducting the normality test and excluding the outliers, one sample was deleted (N=199).

3.2. Syntactic and Lexical Features in Students' Argumentative Essays

3.2.1. Syntactic Features

A factor analysis was conducted to examine the selected 21 syntactic complexity measures. The KMO value of 0.632 and a significant result of Bartlett's test

$$(\chi^2(153) = 7379.9, p < .001) \quad (1)$$

satisfied the assumptions for factor analysis [19]. The factor analysis process led to the removal of three indicators with the smallest loadings (VP, CN, CN/C) and combined the remaining 18 indicators into four factors, which accounted for 87.2% of the total variance.

As shown in Table 1, Factor 1 loaded six measures (MLT, MLS, VP/T, DC/T, CN/T, and C/S) together. This factor was subsequently labelled as the dependent structure ratio factor, representing the proportion of dependent structures in the text. Factor 2 comprised four measures (CT, DC, CT/T, DC/C) and was named the subordinate clause factor, reflecting the presence and frequency of subordinate clauses in the writing. Factor 3 included four frequency measures (C, T, W, S) and was designated as the productivity factor, representing the overall production of linguistic units within the text. CP/T was found to be acceptable for both Factor 1 and Factor 4. To facilitate the interpretation of the factors, CP/T was categorized under Factor 4. As a result, four measures (CP/T, CP, CP/C, MLC) loaded on Factor 4, which was named the coordinate phrase factor, indicating the frequency and distribution of coordinate phrases within the text.

Table 1: Rotated factor loadings for syntactic complexity measures.

Measures	Factors			
	1	2	3	4
MLT	.961			
VP/T	.961			
DC/T	.942			
CN/T	.933			
C/S	.896			
MLS	.868			
CP/T	.696			.594
DC/C		.916		
CT/T		.874		
DC		.865		
CT		.732		
C			.820	
T			.768	
W			.737	
S			.684	
CP/C				.952
CP				.903
MLC				.642
Eigenvalues	6.508	3.580	2.902	2.705
% of Variance	36.158	19.888	16.120	15.027

Note. Extraction method: principal component analysis; Rotation method: Varimax with Kaiser Normalization. A Rotation converged in 5 iterations.

3.2.2. Lexical Features

The KMO value of 0.712 and a significant result of Bartlett’s test

$$(\chi^2(91) = 3496.2, p < .001) \quad (2)$$

satisfied the assumptions for factor analysis [19]. Five variables were identified via factor analysis, which together accounted for 84.3% of the total variance.

As shown in Table 2, the results demonstrated that four measures (VV1, SVV1, CVV1, VV2) loaded together on Factor 1 as the verb variation factor. This factor represents the diversity of verb usage. Factor 2 consisted of three measures (VS1, VS2, CVS1) and was named the verb sophistication factor, reflecting the complexity and advanced nature of verb usage in the writing. Factor 3 incorporated three transformations of NDW (NDW–50, NDW–ER50, NDW–ES50) and was named the NDW factor, indicating the variety of words used in the text. Two lexical sophistication measures were included in factor 4, which was referred to as the “lexical sophistication factor,” and they provided insight into the sophisticated and complicated nature of word usage in writing. Factor 5 comprised noun variation and modifier variation and was labeled as the noun & modifier variation factor, which highlights the diversity of noun and modifier usage within the text.

Table 2: Rotated factor loadings for lexical complexity measures.

Measure	Factor				
	1	2	3	4	5
LS1				.906	
LS2				.895	
VS1		.952			
VS2		.931			
CVS1		.949			
NDWZ50			.816		
NDWER50			.847		
NDWES50			.811		
VV1	.726				
SVV1	.939				
CVV1	.947				
VV2	.898				
NV					.690
MODV					.840
Eigenvalues	3.272	2.977	2.204	1.982	1.421
% of Variance	23.370	21.262	15.740	13.774	10.151

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. A Rotation converged in 6 iterations.

3.3. Correlation Between Integrated Syntactic & Lexical Complexity and Writing Performance

In this section, the relationship between linguistic features and writing quality is examined. At the syntactic level, correlations between the writing quality and coordinate phrase factor were significant in the selected sample ($r=.27$, $p<.01$). At the lexical level, three lexical factors (verb variation factor, NDW factor and noun & mod variation factor) displayed statistically significant correlations with the holistic score. As shown in Table 3 and Table 4, no significant correlation was found between writing quality and other metrics measuring syntactic and lexical complexity.

Table 3: Correlation matrix: Integrated syntactic factors and holistic score of writing quality.

	M	SD	1	2	3	4	5
Syntactic							
1. writing score	69.64	5.26	--				
2. dependent structure ratio factor	-.0009	1.0024	-.012	--			
3. subordinate clause factor	-.0053	.9996	.084	-.001	--		
4. productivity factor	-.0010	1.0024	-.045	.000	-.001	--	
5. coordinate phrase factor	-.0006	1.0024	.268 **	.000	-.001	.000	--
N=199, ** $p < 0.01$ *** <0.001							

Table 4: Correlation matrix: Integrated lexical factors and holistic score of writing quality.

	M	SD	1	2	3	4	5	6
1. writing score	69.64	5.26	--					
2. verb variation factor	.0041	1.0008	.186* *	--				
3. verb sophistication factor	-.0143	.9818	.045	.012	--			
4. NDW factor	.0014	1.0023	.241* *	-.001	.004	--		
5. lexical sophistication factor	-.0018	1.0022	-.028	.001	-.005	.000	--	
6. noun & mod variation factor	.006	.9989	.233* *	-.005	.018	-.002	.002	--
N=199, ** $p < 0.01$ *** <0.001								

4. Discussion

This study comprehensively analyzed the syntactic and lexical complexity measures of L2 students' argumentative essays to explore how the integrated factors predict their writing quality.

At the syntactic level, the findings suggest that essays exhibiting a higher frequency of coordinate phrases and a richer variety of such phrases tend to demonstrate improved writing quality. To be more precise, essays that received better scores had more coordinate phrases (CP), as well as greater proportions of CP per T-unit and CP per clause. The syntactic results were consistent with what Uzun contended that the variables obtaining the most insight into L2 writing ability were the number of

coordinate phrases, CP per sentence, and CP per T-unit [20]. In a similar vein, Jiang claimed that students demonstrating higher writing proficiency are inclined to produce a greater quantity of coordinate phrases [21].

At the lexical level, the verb variation factor, NDW factor and noun & mod variation factor all have a significant correlation with the holistic scores of writing quality, which indicates that verb variation, transformations of NDW measures and noun & mod variation can be valid descriptors of L2 writing quality. The finding is consistent with Lu's study, in which he suggested NDW-ER50 and NDW-ES50 as suitable substitutes when the sample size effect needs to be properly managed [14]. The finding is also consistent with some results of McClure's study, in which she discovered significant differences in her noun and modifier variation measures [22].

Meanwhile, verb sophistication and lexical sophistication have no significant correlation with writing scores. The latter result is consistent with previous research [23], while the former outcome is distinct from those that have been observed in other previous L2 writing studies [14, 22, 24]. According to Lu's research, verb sophistication measures were strongly correlated with test takers' scores. One possible reason for the different relationship between verb sophistication and L2 performance lies in the different genres of the two corpora. The corpus employed in the current study consists of argumentative essays, while the corpus used in Lu's research was descriptive oral data [14]. Previous research shows that L2 learners tend to use many nominalizations in argumentative essays, which might lower the level of verb sophistication [25]. Therefore, whether the verb sophistication factor can be a valid descriptor of L2 performance may be influenced by the form or genre of L2 production.

5. Conclusion

The results of this study show that L2 students need to be trained in a number of linguistic features, hence an important pedagogical question is how teachers might assist L2 students in becoming skilled academic writers. Teachers should pay more attention to the cultivation of specific language skills, especially linguistic features like verb variation, noun & modifier variation, and coordinate phrases. The study also emphasizes how crucial it is to take the genre of the writing into account when determining whether or not particular linguistic features are accurate identifiers of writing quality. In terms of L2 writing automatic rating, this study may provide a reference for the design of the rating system. To better measure students' performance, the automatic rating system should be more sensitive to the linguistic features in the material to achieve better reliability and validity. The valid descriptors of the writing quality should carry more weight in the rating system.

Although the research findings add to the understanding of how linguistic features relate to L2 argumentative writing performance, there are a number of limitations that call for debate. Firstly, the study's small sample size (N=199) and limited participant pool from one university in China may limit the generalizability of the findings. Secondly, using two automated writing assessment systems with slightly different scoring constructs may introduce scoring errors. The use of human raters could enhance consistency in marking criteria. Thirdly, the study only analyzed language in L2 argumentative writing, but genre differences can affect L2 written production. To further understand the connection between linguistic traits and writing performance, future research might look at data from various writing genres.

Acknowledgements

The author wishes to express sincere gratitude to the College English Teaching center of Zhejiang University which provides the Zhejiang University English Proficiency Test (ZJU EPT) that served

as the foundation for this research. The writing materials provided by the ZJU EPT database were a valuable resource, enabling the study to be conducted effectively.

References

- [1] Verspoor, M., Lowie, W., Chan, H. P., & Vahtrick, L. (2017). *Linguistic complexity in second language development: variability and variation at advanced stages*. *Recherches En Didactique Des Langues Et Des Cultures*, 14(1). <https://doi.org/10.4000/rdlc.1450>
- [2] Bulté, B., & Housen, A. (2012). *Defining and operationalising L2 complexity*. *Dimensions of L2 Performance and Proficiency*, 21–46. <https://doi.org/10.1075/llt.32.02bul>
- [3] O'Leary, J. A., & Steinkrauss, R. (2022). *Syntactic and lexical complexity in L2 English academic writing: Development and competition*. *Ampersand*, 9, 100096. <https://doi.org/10.1016/j.amper.2022.100096>
- [4] Lu, X. (2011). *A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development*. *TESOL Quarterly*, 45(1), 36–62. <https://doi.org/10.5054/tq.2011.240859>
- [5] Ortega, L. (2003). *Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Synthesis of College-level L2 Writing*. *Applied Linguistics*, 24(4), 492–518. <https://doi.org/10.1093/applin/24.4.492>
- [6] Read, J. (2000). *Assessing Vocabulary (Cambridge Language Assessment) (1st ed.)*. Cambridge University Press.
- [7] Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). *What Linguistic Features Are Indicative of Writing Quality? A Case of Argumentative Essays in a College Composition Program*. *TESOL Quarterly*, 47(2), 420–430. <https://doi.org/10.1002/tesq.91>
- [8] Yang, W., Lu, X., & Weigle, S. C. (2015). *Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality*. *Journal of Second Language Writing*, 28, 53–67. <https://doi.org/10.1016/j.jslw.2015.02.002>
- [9] Engber, C. A. (1995). *The relationship of lexical proficiency to the quality of ESL compositions*. *Journal of Second Language Writing*, 4, 139–155.
- [10] Laufer, B., & Nation, P. (1995). *Vocabulary Size and Use: Lexical Richness in L2 Written Production*. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>
- [11] Maamuujav, U. (2021). *Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays*. *Assessing Writing*, 49, 100540. <https://doi.org/10.1016/j.asw.2021.100540>
- [12] Ai, H., & Lu, X. (2010, June 10-12). *A web-based system for automatic measurement of lexical complexity [Paper presentation]*. *The 27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10)*, Amherst, Massachusetts, United States of America. <http://dx.doi.org/10.13140/RG.2.2.16499.07208>
- [13] Lu, X. (2010). *Automatic analysis of syntactic complexity in second language writing*. *International Journal of Corpus Linguistics*, 15(4), 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- [14] Lu, X. (2012). *The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives*. *The Modern Language Journal*, 96(2), 190-208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x
- [15] Norris, J. M., & Ortega, L. (2009). *Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity*. *Applied Linguistics*, 30(4), 555–578. <https://doi.org/10.1093/applin/amp044>
- [16] Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. CWK Gleerup.
- [17] Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave MacMillan.
- [18] Vermeer, A. (2000). *Coming to grips with lexical richness in spontaneous speech data*. *Language Testing*, 17(1), 65–83. <https://doi.org/10.1177/026553220001700103>
- [19] Field, A. (2009). *Discovering Statistics Using SPSS, 3rd Edition (Introducing Statistical Methods) (3rd ed.)*. SAGE Publications Ltd.
- [20] Uzun, K. (2020). *Classification of L2 Thesis Statement Writing Performance Using Syntactic Complexity Indices*. In *Fourth International Conference Computational Linguistics in Bulgaria* (p. 42).
- [21] Jiang, J., Bi, P., & Liu, H. (2019). *Syntactic complexity development in the writings of EFL learners: Insights from a dependency syntactically-annotated corpus*. *Journal of Second Language Writing*, 46, 100666. <https://doi.org/10.1016/j.jslw.2019.100666>
- [22] McClure, E. (1991). *A comparison of lexical strategies in L1 and L2 written English narratives*. *Pragmatics and Language Learning*, 2, 141–154.
- [23] Zhang, Y., & Lee, S. (2018). *Lexical Richness, Syntactic Complexity and Writing Performances of Chinese EFL Learners*. *The Journal of Mirae English Language and Literature*, 23(1), 477–502. <https://doi.org/10.46449/mjell.2018.02.23.1.477>
- [24] Wolfe-Quintero, K. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity (Technical Report)*. National Foreign Language Resource Center.

[25] Lee, J. (2021). *Using corpus analysis to extend experimental research: Genre effects in L2 writing*. *System*, 100, 102563. <https://doi.org/10.1016/j.system.2021.102563>