A Study on the Relationship between Syntactic Complexity and Quality of English Argumentative Writing Based on Syntactic Analysis Tool TAASSC

Yue Dong[1], Xianwei Pu[2,*]

dongyue@njust.edu.cn[1], phil667@njust.edu.cn[2]

School of Foreign Studies, Nanjing University of Science and Technology, Nanjing, China

Abstract. The study explored the relationship between 22 syntactic complexity indices generated by TAASSC (Tool for the Automatic Analysis of Syntactic Sophistication and Complexity) and the quality of argumentative essays and the effects of task type (independent and source-based writing task) on the relationship. 126 argumentative written texts by sophomore English majors from Nanjing University of Science and Technology were collected during regular writing classes. The research results are as follows: (1) there was a statistically significant correlation between syntactic complexity and the quality of the argumentative writings; (2) T (normed frequency of T-units) and CN/C (complex nominals per clause) were able to predict the writing quality of the independent and source-based writing tasks, respectively; (3) the predictive power of TAASSC indices for the quality of independent writing (5.7%) was much lower than that for the source-based writing task (27.3%).

Keywords. Argumentative writing quality, Syntactic complexity, Task types, TAASSC.

1   Introduction

1.1   Writing Quality, Syntactic Complexity, and Task Types

Writing well is of great significance to the current study and future language acquisition but a huge challenge for L2 learners and has attracted researchers' attention since the 1960s, covering investigation in multifaceted factors, such as language structures, writing process, individual difference, correction, feedback, etc.

Syntactic complexity, or grammatical complexity, is widely employed as a measurement of language proficiency but owns little consensus about its concept. For example, Wolfe-Quintero et al. [1] defined it as different types of accessible structure both basic and complex, while Ortega [2] defined it as the range of forms surfacing in language production and their level of complexity. As an important part of a written text, syntactic complexity can effectively reflect the level of L2 writing and is frequently used to assess learners' proficiency [3].

A balanced and appropriate level of syntactic complexity can enhance the overall readability and engagement of the text, which allows for the expression of nuanced thoughts, logical connections, and sophisticated arguments. Friginal and Weigle [4] argued that the study of their relationship helps understand the importance of syntactic complexity in writing from the

perspectives of education and assessment. Previous studies have investigated this relationship under the influence of various factors including genre, time limit, and writing strategy, among which the types of writing tasks have aroused the interest of several researchers [5, 6], especially after the change in standardized language exams like TOEFL and IELTS that multiple writing tasks or source-based task has gradually replaced the traditional model of a single, independent writing task.

In China, more and more language exams are adding different types of writing tasks, such as TEM (Test for English Majors) and NMET (National Matriculation English Test). Given the large scale of exams, however, fewer researchers in China have discussed the differences among task types, let alone studies about syntactic complexity [7]. Therefore, more investigation will help to obtain a full view of influences induced by task types to prepare students and instructors.

1.2 Software TAASSC

TAASSC (Tool for the Automatic Analysis of Syntactic Sophistication and Complexity) is a free syntactic analysis tool developed by Kyle [8] that provides various indices measuring syntactic structures. The software is available from the website which provides versions for different computer systems. It has been widely used for its effective indices and convenience compared with previous automatic text analysis tools of syntactic complexity like the Biber Tagger.

Table 1. L2SCA variables adapted from Lu's study [9].

Variable	Code	Definition
Type 1: Length of production unit		
Mean length of clause	MLC	# of words / # of clauses
Mean length of sentence	MLS	# of words / # of sentences
Mean length of T-unit	MLT	# of words / # of T-units
Type 2: Sentence complexity		
Clauses per sentence	C/S	# of clauses / # of sentences
Type 3: Subordination		
Clauses per T-unit	C/T	# of clauses / # of T-units
Complex T-units per T-unit	CT/T	# of complex T-units / # of T-units
Dependent clause per clause	DC/C	# of dependent clauses / # of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses / # of T-units
Type 4: Coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases / # of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases / # of T-units
T-units per sentence	T/S	# of T-units / # of sentences
Type 5: Particular structures		
Complex nominals per clause	CN/C	# of complex nominals / # of clauses
Complex nominals per T-unit	CN/T	# of complex nominals / # of T-units
Verb phrases per T-unit	VP/T	# of verb phrases / # of T-units

Among the indices provided by TAASSC, this study used the section called L2SCA (L2 Syntactic Complexity Analyzer) variables proposed using the research of Wolfe-Quintero et al.

[1] and Ortega [2] as a basis by Lu [9] since they were proved to have a positive correlation with growth in L2 proficiency in previous studies.

These indices are categorized into five types listed in the table by measuring nine kinds of syntactic structures (i.e. word, sentence, clause, dependent clause, T-unit, complex T-unit, coordinate phrase, complex nominal, and verb phrase), and can correspond with the four Norris and Ortega-proposed [10] dimensions of syntactic complexity. The reliability of 14 L2SCA variables (see Table 1) was reported by several studies [9, 11, 12].

1.3 The Present Study

This study aims to explore the relationship between syntactic complexity and the quality of L2 writing from source-based (in this study, only the reading-to-write task) and independent tasks. Several automated syntactic indices were employed to evaluate syntactic differences in the compositions and to find out how effectively these measures may predict the holistic variance in the human rating quality of them, statistical analyses were performed. By investing in this relationship, this study hopes to enrich related empirical studies to reach a further understanding of it and give suggestions for L2 learning and writing assessing or scoring techniques. The following questions were addressed:

(1) What is the relationship between syntactic complexity and the quality of independent English argumentative writing?

(2) What is the relationship between syntactic complexity and quality of source-based English argumentative writing?

2 Methodology

2.1 Data Collection and Procedure

126 Chinese English majors from Nanjing University of Science and Technology participated in this study. Among the 126 participants, 51 of them were students of the year 2015, and 75 were students admitted in 2016. All participants were sophomores when they wrote the argumentative compositions. As English majors, they have relatively high-level language proficiency and ability in L2 writing compared with other majors.

For the data collection, the participants were asked to take an exam during the class, which was completed in accordance with the requirements of TEM-4. All the students were required to finish the composition within a time limit of 45 minutes independently.

The compositions were first typed into the computer, with spelling and punctuation mistakes corrected to prevent disturbance to automatic analysis. After selecting the input folder which contains all text files of compositions, TAASSC would output a comma separated values (.csv) spreadsheet that listed the raw data of L2SCA indices of each text according to the frequency counts of units and structured retrieved. Despite the 14 L2SCA ratio indices directly generated by TAASSC, this paper added another 8 normed frequency indices calculated according to the following formula (1) based on the raw data of nine syntactic structures:

$$S/C/DC/T/CT/CP/CN/VP \text{ (normed indices)} = \frac{S/C/DC/T/CT/CP/CN/VP}{W} \times 100 \qquad (1)$$

Two experienced TEM-4 raters scored each text using the scoring rubric of TEM-4. Each composition's score was calculated as the average of the ratings from raters. Table 2 provides details on the descriptive statistics of the composition scores.

Table 2. Participants at each score level and the writing scores' descriptive statistics.

Score (Level)	Independent	Source-Based
18-20 (5)	0	1
15-17.5 (4)	18	17
12-14.5 (3)	51	20
8-11.5 (2)	6	13
1-7.5 (1)	0	0
Maximum/Minimum	16/10	18.5/9
Mean (SD)	13.753 (1.367)	13.520 (2.377)

The data of 22 indices and holistic human scores of compositions were input into SPSS 22.0. According to task types, they were grouped into two sets and analyzed separately and identically. Pearson correlations were used to analyze relationships between the 22 indicators and the human judgment scores for each task group. All variables were tested for multicollinearity before being included in the next step of the analysis to ensure that they were not measuring the same or similar constructs of syntactic complexity. When indices exhibited a strong correlation ($r > .9$) with holistic scores, the most strongly correlated index was retained for the regression analysis, while the remaining variables were removed. After that, a stepwise regression analysis was run using the chosen indices against the writing's scores.

3   Results and Discussion

3.1   Independent Argumentative Writing

Table 3. Correlations between holistic independent argumentative writing scores and L2SCA variables entered into regression model.

Variable	Pearson Correlation	Sig. (2-tailed)
Mean length of clause	.173	>. 05
Clauses per sentence	.073	>. 05
Verb phrases per T-unit	.188	>. 05
Dependent clause per clause	.189	>. 05
T-units per sentence	-.035	>. 05
Complex T-units per T-unit	.190	>. 05
Complex nominals per T-unit	.201	>. 05
Complex nominals per clause	.200	>. 05
Normed frequency of sentences	-.231*	.047
Normed frequency of verb phrases	-.039	>. 05
Normed frequency of T-units	-.239*	.039
Normed frequency of complex T-units	.106	>. 05
Normed frequency of coordinate phrase	-.107	>. 05
Normed frequency of complex nominals	.129	>. 05

Of the 22 indices, 13 variables showed multicollinearity: MLS, MLC, MLT, C/T, DC/C, DC/T, CP/T, CP/C, *S*, *T*, *C*, *DC* and *CP* (.901 ≤ |*r*| ≤ .980). After checking their correlation with holistic scores, 5 indices (MLC, DC/C, *S*, *CP* and *T*) remained. Therefore, 14 indices were selected, and Table 3 presents the correlations among these indices and holistic scores.

Stepwise multiple regression was used to analyze the remaining 14 indices, and one variable was included in the model: *T*. The result showed that the model could explain 5.7% of the variance in holistic scores of source-based compositions (Table 4 presents a summary of the model). Normed T-unit frequency was demonstrated to be an indicator of TEM-4 independent composition quality in this study.

Table 4. Summary of stepwise multiple regression model for independent argumentative writing.

Entry	Predictor included	*r*	*R²*	*R² change*	*B*	β	*SE*
1	*T*	.239	.057	0.044	-.239	-.261	1.336

Note. Estimated constant term = 10.897, β = unstandardized beta, *SE* = standard error, *B* = standardized beta.

3.2 Source-Based Argumentative Writing

Of the 22 indices considered, 13 variables showed multicollinearity: MLS, MLT, MLC, C/S, C/T, DC/T, DC/C, CP/T, CP/C, *S*, *T*, *C*, and *CP* (.903 ≤ |*r*| ≤ .980). After checking their correlation with holistic scores, 5 indices (MLC, C/T, DC/C, CP/T, and *T*) remained. Therefore, 14 indices were selected, and Table 5 presents the correlations among these indices and holistic scores.

Table 5. Correlations between holistic source-based argumentative writing scores and L2SCA variables entered into regression model.

Variable	Pearson Correlation	Sig. (2-tailed)
Mean length of clause	.500**	< .001
Verb phrases per T-unit	.157	>. 05
Clauses per T-unit	-.052	>. 05
Dependent clause per clause	.056	>. 05
T-units per sentence	-.006	>. 05
Complex T-units per T-unit	-.061	>. 05
Coordinate phrases per T-unit	.366*	.008
Complex nominals per T-unit	.376*	.007
Complex nominals per clause	.523**	< .001
Normed frequency of verb phrases	-.388*	.005
Normed frequency of T-units	-.377*	.006
Normed frequency of dependent clauses	-.191	>. 05
Normed frequency of complex T-units	-.293*	.037
Normed frequency of complex nominals	.238	>. 05

Table 6. Summary of stepwise multiple regression model for source-based argumentative writing.

Entry	Predictor included	r	R^2	R^2 change	B	β	SE
1	CN/C	.523	.273	.258	.523	4.679	2.047

Note. Estimated constant term = 7.528, β = unstandardized beta, SE = standard error, B = standardized beta.

Stepwise multiple regression with the selected 14 indices was conducted, and one variable was included in the model: CN/C. The result showed that the model could explain 27.3% of the variance in holistic scores of source-based compositions (Table 6 gives a summary of the model). The complex nominal per clause was demonstrated to be an indicator of TEM-4 source-based composition quality in this study.

3.3   Discussion

Syntactic Complexity and Writing Quality of Independent Writing Task. The results of regression analysis indicate that the relationship between syntactic complexity and holistic scores of essays for the independent task was significant but small. The multiple regression model that contains one variable T could explain only 5.7% of the variance of holistic scores. What's more, T (normed frequency of T-unit) was the only measure among those entered into regression that has a statistically significant correlation ($p < .05$) with scores, $r = -.239$, which means raters prefer to give higher scores to independent compositions with fewer T-units. The result is similar to Kyle & Crossley's research [13], which reported a model possible to explain 5.8% of the variance of scores, but their model only included MLC (mean length of clause).

In previous SLA research, T-unit-based measures have been one of the most vital and widely used indices of syntactic complexity and have been demonstrated to be strongly positively correlated with writing quality in many studies [1, 2]. However, in the present study, the normed frequency of T-units was found to have a weak negative correlation with the quality of independent writing. There are some possible explanations for the results.

First, the scores of compositions in any exams are expected to reflect multiple factors of writing proficiency, and for the independent task, raters may pay more attention to other dimensions like linguistics errors and lexical proficiency. For example, Kyle and Crossley [14] found that lexical sophistication was a significant predictor of human ratings of independent writing in their study, with their regression model explaining 36.8% of the variance. In TEM-4 independent argumentative writing, learners are expected to argue for their opinions in the whole essay, and they may focus on the choice of vocabulary to give examples and reasons to better support themselves rather than complicate their sentence structures.

Second, the proficiency level of participants can influence the results. All the participants were sophomores majoring in English, who were intermediate-level learners but still had much room for improvement. The average normed frequency of T-unit for this group was 6.81, while the number of complex T-units was 2.97; it indicates that more than half of the T-units used were simple T-units like "I drank water". The frequent use of such structure may create meaningless and redundant sentences though it brings higher levels of T-unit-based measurements, leading to the negative correlation between writing quality and frequency variable of T-unit. On the contrary, given the same length of composition, the lower frequency of T-units means fewer but

longer T-units, and they can contain more complex structures like longer clauses, which brings a higher level of syntactic complexity and rating.

Other possible explanations include the influence of the corpus size, writing topic, and the choice of measures. The corpus size of the independent task used in this study is very small, including only 75 essays on the same topic and participants from only one university. What's more, the distribution of scores (see Table 2) was likely to affect the results, since the majority of scores were concentrated at the same score level. As mentioned before, this study reports a different predictor than Kyle & Crossley's [13] research. It may be attributed to the difference in writing topics (they collected compositions of two topics) besides the size of the corpus. Concerning the indices used, it is possible that traditional indices can hardly gauge the syntactic complexity in the independent task in exams like TEM-4, and usage-based indices could offer a better interpretation of their relationship. For instance, according to Kyle & Crossley's study [13], usage-based measures explained greater variance (14.2%) than traditional indices (5.8%) of the holistic scores of compositions.

Syntactic Complexity and Writing Quality of Source-Based Writing Task. The relationship between syntactic complexity and holistic scores of compositions from the source-based task was much stronger than with scores of compositions from the independent task. Among the 14 indices that were entered into regression, 7 indices were statistically correlated ($p < .05$) and 6 had an above-medium correlation ($r > .3$) with scores. The multiple regression model could explain 27.3% of the variance of scores, which only included one predictor: CN/C, which is considered a phrasal elaboration measurement. It indicates that writings that are more complicated at the phrasal level tend to receive higher ratings. Though different measures and statistical analysis methods were selected, Biber et al. [5] also pointed out that phrasal complexity features were strongly associated with the written source-based task compared with independent and spoken tasks. In Casal and Lee's research [15], they reported that phrasal (MLC and CN/C) measures differed significantly among the three rating levels of source-based compositions, but MLT, a global measure, was also reported.

The great influence of CN/C on the quality of the source-based writings suggested in this study seems to be consistent with Biber et al.'s results [16], as well as subsequent L2 writing studies that proved the predictive power of different measurements of noun modifiers [17]. Lu [18] also concluded the best measures for distinguishing proficiency of syntactic complexity in his study across four levels, and CN/C was one of them.

There are two possible reasons to explain the results. One of the reasons may be the similarity between academic writing and source-based writing. In their study on syntactic complexity, Biber et al. [16] found that complex noun phrase constituents and complex phrases can better capture the complexities of professional academic writing by investigating 28 grammatical features. Source-based writing, which also prompts test-takers to synthesize information given by source texts and language knowledge owned by themselves, is believed to resemble the writing type of academic essay in higher education [19]. In source-based writing, various types of complex nominals are needed to accurately summarize contents and produce refined sentences, as the specialized circumstances of academic writing. Hence, compositions from this task are overwhelmingly likely to share a similar syntactic structure that contains abundant and complex phrasal constituents, with professional academic contents.

Another possible explanation is the development stages for L2 learners proposed by Biber et al. [16] that mirror the progression from the conversational ability to academic writing competence of L1 learners. Although in many L2 writing classes, students are encouraged to lengthen and complicate their sentences with longer constructions, the clarity of expression and condensed information is what is valued in higher education and its common written registers, which require complexity features usually acquired in later developmental stages. According to Biber et al. [16], the final stage requires heavy use of phrasal-dependent structures as noun phrase constituents, which is able to explain why raters prefer essays with more complex noun phrases in the source-based writings.

Compared with Casal and Lee's [15] study, MLT (mean length of T-unit) did not turn out to be a predictor in this study. The corpus size and students' English proficiency level may cause this difference, for their research selected 280 compositions written by L2 students in US universities who have obtained a high score in a TOEFL, IEP (intensive English program) test, and/or first-year writing class. This study's data, however, consists of compositions written by 75 English majors in a Chinese university.

Comparison between Results of Independent and Source-Based Compositions. In general, significant predictors for the independent and the source-based compositions are quite different. The independent essays with higher scores demonstrated less frequency of T-units, and source-based writings rated higher benefited from more complex nominals. The difference in significant predictors between the task types may indicate the notion that these two tasks are of different genres [6], and require different structures of syntactic complexity. It also presents the possibility to distinguish the essays from these two task types simply by these indices.

Besides, in both groups, indices measuring particular structures and frequency indices outperformed length measures like MLT though they were commonly used. The results are partly consistent with Biber et al.'s research [5], in which they recommended the consideration of phrasal features in future research and denied the use of holistic measures depending on overall length. As more recent studies [14, 20] challenged the linguistic interpretability of these traditional indices, these results also raise a concern for the choice of indices in future syntactic complexity studies.

In the case of the regression model's strength, though the model of source-based essays is much stronger than the one of independent essays, neither of them was particularly strong when compared with models that contained more comprehensive indices [6]. Since complexity is not a single-dimensional concept, it is illogical to believe that any single metric will accurately describe it; what's more, in addition to syntactic complexity, exams' essay grading rubrics incorporate assessments from a variety of language proficiency aspects like lexical proficiency and cohesion. However, the difference in the variance of holistic scores, which means the predictive powers of syntactic complexity measures for independent and source-based writings, may indicate the distinction in the focus of grading standards: syntactic complexity plays a relatively more important role in source-based essays' rating than in independent ones. For example, students may pay more attention to syntax when writing source-based essays, while to other aspects like vocabulary when writing independent essays.

These results to some extent align with the nature of both tasks, while independent compositions generally rely on writers' personal opinions and supporting reasons and examples, source-based essays need precise and condensed summarization of source text as well as the synthesis of

information supported by source contents and personal experience. The substantial difference in syntactic structures between the two task types, on the one hand, proves the necessity to include multiple writing tasks in major exams of English proficiency for L2 learners; on the other hand, is vital for the preparation of training and exercising methods for teachers of English writing who want to help learners grip with different linguistic features of diverse registers and students who wish to improve their performance in those tests. Moreover, future automatic second language writing assessing and scoring techniques may benefit from these findings. By focusing on the different syntactic features of writing tasks, these scoring systems can better simulate real raters and give more reliable scores, which is able to facilitate the scoring procedure in large-scale exams (to reduce a margin of error in scores caused by raters and to improve efficiency) and self-assessment for test-takers.

4   Conclusion

The paper investigated the relationship between syntactic complexity and writing quality of two different task types (independent task and reading-to-write task, a kind of source-based task) by using the automatic text analysis tools of syntactic complexity TAASSC and data analysis software SPSS. In terms of independent argumentative writing, the final model of one predictor (normed frequency of T-unit) only explained 5.7% of the variance of holistic scores, and the predictor presented a low negative correlation with scores. In terms of source-based argumentative writing, the regression model consisted of one predictor (complex nominals per clause) that could explain 27.3% of the variance of holistic scores, and this indicator showed a high positive correlation with writing scores.

The results indicate that, first, syntactic complexity is a significant predictor of writing quality for the source-based task, but a much weaker one for the independent task. Also, the differences in models provide evidence that independent writing and source-based writing don't share the same set of syntactic constructs, i.e. their syntactic features differ. As for measure validity, this study presents to be in line with previous studies that denied traditional T-unit and length-based indices and promoted phrasal complexity indices. In addition, the low percentage of the variance of scores explained by those large-grained indices, especially in terms of independent compositions, suggests the requirement to adopt more comprehensive and valid indices in future research and development of automatic analysis tools.

Practically, the study provides evidence for the diversifying of writing task types in standardized and high-stakes language exams, which improves the measurement of test-takers' writing abilities. Since specific predictors are meaningfully related to the holistic scores of certain task types, the scoring rubrics should differ, both for human judgment and possible AES systems. With a deeper understanding of syntactic complexity, writing quality, and task types, writing instructors and students can prepare themselves with more targeted and effective lectures and exercises.

References

[1] Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y.: Second language development in writing: Measures of fluency, accuracy, & complexity. University of Hawaii Press, Honolulu (1998)

[2] Ortega, L.: Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. Applied Linguistics. pp. 492-518+558 (2003)

[3] Skehan, P.: Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. Applied Linguistics. pp. 510-532 (2009)

[4] [Friginal, E., & Weigle, S.: Exploring multiple profiles of L2 writing using multi-dimensional analysis. Journal of Second Language Writing. pp. 80-95 (2014)

[5] Biber, D., Gray, B., & Staples, S.: Predicting patterns of grammatical complexity across language exam task types and proficiency levels. Applied Linguistics. pp. 639-668 (2016)

[6] Guo, L., Crossley, S. A., & Mc Namara, D. S.: Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. Assessing Writing. pp. 218-238 (2013)

[7] Zhang, X., & Zhou, Y.: The influence of task types on Chinese English learner's writing performance. Modern Foreign Languages. pp. 548-558+585 (2014)

[8] Kyle, K.: Measuring Syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-Based Indices of Syntactic Sophistication. Dissertation, Georgia State University (2016)

[9] Lu, X.: Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics. pp. 474-496 (2010)

[10] Norris, J., & Ortega, L.: Towards an organic approach to investigating CAF in SLA: the case of complexity. Applied Linguistics. pp. 555-578 (2009)

[11] Yang, W., Lu, X., & Weigle, S. C.: Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. Journal of Second Language Writing. pp. 53-67 (2015)

[12] Yoon, H.-J., & Polio, C.: The linguistic development of students of English as a second language in two written genres. TESOL Quarterly. pp. 275-301 (2016)

[13] Kyle, K., & Crossley, S.: Assessing syntactic sophistication in L2 writing: A usage-based approach. Language Testing. pp. 513-535 (2017)

[14] Kyle, K., & Crossley, S.: The relationship between lexical sophistication and independent and source-based writing. Journal of Second Language Writing. pp. 12-24 (2016)

[15] Casal, J. E., & Lee, J. J.: Syntactic complexity and writing quality in assessed first-year L2 writing. Journal of Second Language Writing. pp. 51-62 (2019)

[16] Biber, D., Gray, B., & Poonpon, K.: Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? TESOL Quarterly. pp. 5-35 (2011)

[17] Crossley, S. A., & McNamara, D. S.: Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. Journal of Second Language Writing. pp. 66-79 (2014)

[18] Lu, X.: Computational methods for corpus annotation and analysis. Dordrecht: Springer (2014)

[19] Cumming, A., Kantor, R. Powers, D., Santos, T., & Taylor, C.: TOEFL 2000 writing framework: A working paper. TOEFL Monograph 18. Educational Testing Service, Princeton, NJ (2000)

[20] Biber, D., Gray, B., Staples, S., & Egbert, J. Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. Journal of English for Academic Purposes. 100869 (2020)