

IT'S COMPLICATED: THE RELATIONSHIP BETWEEN LEXIS, SYNTAX, AND PROFICIENCY

Christopher Williams

Abstract

This paper explores the relationship between lexical and syntactic complexity measures and proficiency in L2 English argumentative essays written by L1 Czech high school students. Syntactic complexity is generally understood as referring to the “range and sophistication” (Ortega 2015) of grammatical constructions, whereas lexical complexity can refer to the range and frequency of the words used. The research used 100 essays written by final year high school students. Lexical complexity was analysed using the Lexical Complexity Analyzer (Ai & Lu 2010, Lu 2012), syntactic complexity using the L2 Syntactic Complexity Analyzer (Lu 2010, 2014) and Biber et al.’s (2011) hypothesised developmental stages for complexity framework. Despite a large number of measurements failing to produce any significant patterns, positive correlations were found between lexical diversity measures and vocabulary scores. Similarly, Mean Length of Clause (MLC) and Complex nominals per clause (CN/C) showed weak positive associations with grammar scores, as did Stage 5 of the developmental stages. The findings provide an insight into the kinds of complexity features that can be given more focus during instruction and underscore the potential of these measures as determinants of proficiency.

Keywords

syntactic complexity, lexical complexity, writing proficiency

1 Introduction

Research on second language (L2) complexity, accuracy, and fluency (CAF) production can (depending on the indices used) reveal information about a learner’s level of proficiency in the target language (Khushik & Huhta 2019). CAF-based research is a crucial factor that serves as an indicator, diagnostic, and major parameter for L2 learning, teaching, and research (Wolfe-Quintero et al. 1998, Bulté & Housen 2014). The complexity branch of the CAF triad includes syntactic complexity (SC) and lexical complexity (LC). Both SC and LC are multidimensional constructs, with the former encompassing measures of length, subordination, and coordination, and the latter lexical density, diversity, and sophistication (Lu 2012). Numerous L2 academic writing studies have explored the extent to which these measures can serve as reliable and valid determiners of learners’ general language proficiency, particularly the quality of their writing performance (Bulté & Housen 2014, Mazgutova & Kormos 2015). For example,

with regards to SC, length-based measurements have shown that essays with longer sentences tend to score higher, and essays that score higher use less frequent words (Crossley et al. 2011, McNamara et al. 2013).

Though writing well in a second language can be challenging for learners, the writing process can be made less so through a better understanding of what makes for good writing. This study aims to contribute both towards research in complexity and towards better comprehension of what good writing is by analysing the syntactic and lexical complexity of timed independent argumentative essays written by L1-Czech L2-English highschool students in their final year of studies, in order to investigate whether there are proficiency-related differences in the complexity features used by the students. A further incentive for this study is that few studies have been conducted on the complexity of the writings of Czech EFL learners at a highschool level.

1.1 Syntactic complexity

The relationship of syntactic complexity (SC) to second language (L2) writing quality has long been the focus of L2 writing research (e.g. Biber et al. 2016, Kyle & Crossley 2018, Casal & Lee 2019). The general assumption is that more complex syntactic structures within a written text are an indication of more advanced writing abilities (Yang et al. 2015, Cossley 2020) and that a student who is able to produce more complex structures is more likely to do so, particularly in a high-stakes context, such as when under assessment.

On both conceptual and methodological fronts, researchers aim to pinpoint indices that not only capture the multifaceted essence of SC but that also exhibit robust correlations with L2 writing quality evaluations (e.g. Biber et al. 2020). From an instructional perspective, a better understanding of this interplay can provide valuable guidance as to which aspects of SC should receive greater attention for different learners, genres, and writing tasks.

Syntactic complexity has been defined as the range and the sophistication of grammatical resources used in language production that can also be referred to as variety, diversity, and elaborateness (Ortega 2015), and more recently as the addition of structural elements to ‘simple’ phrases and clauses (Biber et al. 2020). SC can be affected by L1 (Lu & Ai 2015) and task genre (Beers & Nagy 2009, Lu 2011, Yang et al. 2015, Qin & Uccelli 2016, Yoon 2017).

1.1.1 Syntactic complexity measures

The L2 Syntactic Complexity Analyzer (L2SCA) (Lu 2010) paired with the Stanford Parser (Klein & Manning 2003) uses 14 indices (see Table 1) to assess the syntactic complexity of sentences within a text. The L2SCA quantifies various

linguistic units, such as words, sentences, clauses, t-units, complex nominals, and nominal phrases (see Lu 2010: 479 for a more detailed description). The measures chosen by Lu were predicated on their established correlation with proficiency in previous research, a decision that was subsequently corroborated by further investigation.

Area	Label	Description
Length of production unit	MLC	Mean length of clause
	MLS	Mean length of sentence
	MLT	Mean length of T-unit
Amount of subordination	C/T	Clauses per T-unit
	CT/T	Complex T-unit
	DC/C	Dependent clause per clause
Amount of coordination	DC/T	Dependent clause per T-unit
	CP/C	Coordinate phrases per clause
	CP/T	Coordinate phrases per T-unit
Degree of phrasal sophistication	T/S	T-units per sentence
	CN/C	Complex nominals per clause
	CN/T	Complex nominals per T-unit
Overall sentence complexity	VP/T	Verb phrases per T-unit
	C/S	Clauses per sentence

Table 1: Syntactic complexity measures in the L2 Syntactic Complexity Analyzer (adapted from Lu 2017)

Since the development of the L2SCA, emergent research has revealed that length-based measures hold predictive potential for L2 writing quality (Kyle & Crossley 2018). Specific measures such as mean length of t-unit (MLT) have demonstrated a correlation with the writing proficiency of tertiary-level English as a second language (ESL) learners (Kim 2014, Yang et al. 2015, Casal & Lee 2019). Similarly, mean length of sentence (MLS) (Taguchi et al. 2013, Chen et al. 2014, Yang et al. 2015) and MLC have also been linked with the proficiency of argumentative essays (Chen et al. 2014, Qin & Uccelli 2016).

While length-based measurements have shown some effectiveness in indicating proficiency levels, their use has not been without criticism. One concern is that while these measures can distinguish among Basic User, Independent User and Proficient User (according to the CEFR), they are not as effective when it comes to determining differences within user bands – for example, whether a learner is B1 or B2. This limitation becomes more pronounced higher up the

proficiency scale. Paquot (2018) found that length-based measures such as MLC, MLS, and MLT fell short in adequately differentiating between high-level and higher-level learners (B2, C1, and C2), claiming that sentence length tends to stabilise as learners begin expressing complexity in different ways.

The main problem, however, with length-based measures for descriptive purposes is that although they have been used to provide a general idea of a text's proficiency level, the fact remains that they are unable to distinguish among the wide range of structural/syntactic complexity features in English (Biber et al. 2020: 9).

Housen et al. (2012: 4) provide a more specific metric for measuring complexity, suggesting it concerns "the extent to which learners use syntactic embedding and subordinate clauses, relative to the total number of clauses produced". However, as indicated by Biber et al. (2011), more advanced writers in an academic context are more likely to demonstrate complexity through embedding rather than subordination.

1.2 Lexical complexity

Studies exploring the correlation between L2 writing performance (and/or L2 writing development) and lexical complexity have mainly centred on two factors: (a) measures within the text, such as lexical density – calculated as the proportion of lexical (or content) words to total words, and lexical diversity (also referred to as lexical variation) – the ratio of unique words to the total number of words; and (b) external measures such as lexical sophistication, usually calculated as the relative frequency or infrequency with which the L2 writers' lexis appears in the target language (Lu 2012).

The following sections are an overview of the indices of lexical complexity as interpreted by the LCA. For the full list, including formulae, please refer to Table 2.

Index	Label	Calculation	Explanation
Lexical density	LD	N_{lex}/N	Lexical words to the number of words
Lexical sophistication-I	LS1	N_{slex}/N_{lex}	Sophisticated lexical words to the total number of lexical words
Lexical sophistication-II	LS2	Ts/T	Sophisticated word types to the total number of word types
Verb sophistication-I	VS1	T_{sverb}/N_{verb}	Number of sophisticated verb types to the total number of verbs
Verb sophistication-II	VS2	T^2_{sverb}/N_{verb}	Variations (corrections) of VS1 measure
Corrected VS1	CVS1	$T_{sverb}/\sqrt{2N_{verb}}$	
Number of different words	NDW	T	Number of different words used in a language sample
NDW (first 50 words)	NDWZ-50	T in the first 50 words of sample	
NDW (expected random 50)	NDW-ER50	Mean T of 10 random 50-word samples	
NDW (expected sequence 50)	NDW-ES50	Mean T of 10 random 50-word sequences	
Type/Token ratio	TTR	T/N	Number of word types to the number of words in a text
Mean Segmental TTR (50)	MSTTR-50	<i>Mean segmental TTR - 50-word non-overlapping segments</i>	
Corrected TTR	CTTR	$T_{sverb}/\sqrt{2N}$	
Root TTR	RTTR	T_{sverb}/\sqrt{N}	
Bilogarithmic TTR	logTTR	$LogT/LogN$	
Uber Index	Uber	$Log^2 N / Log(N/T)$	
Verb variation-I	VV1	T_{verb}/N_{verb}	Variation of specific classes of words
Squared VV1	SVV1	T^2_{verb}/N_{verb}	
Corrected VV1	CVV1	$T_{verb}/\sqrt{2N_{verb}}$	
Lexical word variation	LV	T_{lex}/N_{lex}	
Verb variation-II	VV2	T_{sverb}/N_{lex}	
Noun variation	NV	T_{noun}/N_{lex}	
Adjective variation	AdjV	T_{adj}/N_{lex}	
Adverb variation	AdvV	T_{adv}/N_{lex}	
Modifier variation	ModV	$(T_{adj} + T_{adv})/N_{lex}$	

Table 2: Lexical complexity measures in the Lexical Complexity Analyzer (Lu 2012)

1.2.1 Lexical density (LD)

Lexical density refers to the number of content words (nouns, verbs, adjectives, and adverbs) to the total number of words in a text (Johansson 2008), though Lu (2012) does not include modal verbs in the LCA. Johansson posits that analysing lexical density can explain the concept of information packaging, in that a text dense with lexical words communicates more information than one primarily containing function words (such as prepositions, articles, conjunctions, and pronouns).

1.2.2 Lexical diversity

Research examining lexical diversity (sometimes referred to as lexical variation) has identified a consistent positive relationship between lexical diversity, irrespective of the measurement method, and performance in L2 writing (Grant & Ginther 2000, Jarvis et al. 2003, Crossley & McNamara 2012, Kim 2014). Type-token ratio (TTR) is a standard metric for assessing lexical diversity, however, its reliability has been the subject of debate as it can be influenced by the length of the text being analysed. To counter this, refined measures such as the corrected TTR (CTTR) (Carroll 1964) and root TTR (RTTR) (Guiraud 1960, as cited in Torruella & Capsada 2013) were introduced, although some concerns persist (Vermeer 2000). Nevertheless, Lu (2012) and Daller et al. (2003) found meaningful correlations between using TTR and RTTR as measures of lexical variation and language proficiency.

This seems to apply across genres, though the measurement does reveal differences as to how diversity is realised. Yoon (2017) noted that variations in lexical diversity may occur when the same writer is composing in different genres, noting that when writing argumentative essays both native and non-native English language writers opted for a narrower range of vocabulary, however, a more extensive lexical range was used for narrative purposes. That being said, measurements of lexical sophistication (discussed in the following section) revealed that both L1 and L2 writers used less frequent but more sophisticated vocabulary in their argumentative essays as opposed to narrative essays.

1.2.3 Lexical sophistication

The vocabulary of advanced L2 writers has been analysed by comparing vocabulary usage against English corpora and academic word lists (Nation 2006, Davies 2008). Investigations into L2 writing have revealed that the use of a wider range of low-frequency words is an indication of L2 writing development and

performance (Johnson et al. 2013, 2016). However, as touched upon in the previous section, Yoon and Polio (2016) have indicated that L1 and L2 argumentative writing incorporates less frequent and more sophisticated vocabulary than L1 and L2 narrative writing, implying that the higher informational nature of argumentative genres is realised through the use of less frequent lexis.

1.3 Hypothesised Developmental Stages for Complexity Features

The Hypothesised Developmental Stages for Complexity Features (hereafter developmental stages – DS) (see Table 3) is a sequence of stages for student writing development proposed by Biber et al. (2011) based on the findings of a large-scale corpus-based analysis. In their analysis, Biber et al. suggest that the developmental progression of language acquisition begins with conversation, which is acquired first, and then progresses to the grammar of writing. The early stages contain syntactic structures that are readily acquired and frequently produced in conversation and so do not represent a high degree of complexity. The higher stages contain types of complex phrasal embedding that are only produced in specialised formal writing contexts. These styles are not acquired as naturally, as with the lower stages – given that many native speakers of English rarely produce such structures. As such, Biber et al. consider these structures as representing a higher degree of complexity.

The Biber et al. (2011) study challenges the traditional measures of grammatical complexity, such as t-units and clausal subordination, which have been frequently used in assessing language proficiency, arguing that the grammatical complexities of academic writing are fundamentally different from those of conversation and non-academic writing, in addition to being neither effective discriminators of language proficiency nor well motivated from a linguistic perspective. The study was also unique in its consideration of lexicogrammatical factors – rather than separating lexical and syntactic indices, the co-occurrence of factors was considered.

**IT'S COMPLICATED:
THE RELATIONSHIP BETWEEN LEXIS, SYNTAX, AND PROFICIENCY**

Stage	Grammatical structure
1	Finite complements (<i>that</i> and WH) controlled by very common verbs (e.g. <i>think, know, say</i>)
2	Finite complements controlled by a wider set of verbs Non-finite complements controlled by very common verbs
3	Finite complements controlled by adjectives Non-finite complements controlled by a wider set of verbs <i>That</i> relative clauses
4	Non-finite complements controlled by adjectives Non-finite relative clauses
5	Preposition + non-finite complements Complements controlled by nouns

Table 3: Hypothesised Developmental Stages for Complexity Features (adapted from Biber et al. 2011). Note that common verbs refer to those identified in Biber et al. (1999). For a full list of verbs considered common in this study, see Appendix 1.

1.4 Research aims

Much of the previous linguistic complexity in writing research has focused on advanced or university level L2 English users, the results of which may not be applicable to intermediate secondary school users (Lee et al. 2021). This focus leaves a gap in our understanding of intermediate L2 English users in a secondary school context, in particular, the relationship of complexity measurements as applied to a common task – the argumentative essay. Being able to write an argumentative essay is an important skill for second language learners given that they form a key element in numerous exams or other such proficiency evaluations. Such essays are also viewed as preliminary demonstrations of academic writing (Kyle 2016).

The use of a comprehensive suite of complexity measurements (lexical and syntactic) further enhances the ability to capture the multidimensional nature of complexity, certainly so with the blend of lexicogrammatical features as outlined in the developmental stages. This will allow for the identification of the current level of complexity students are achieving in order to more effectively help them improve their level of English.

This study focuses on the relationship of syntactic and lexical complexity to the expert raters' judgments of the quality of argumentative essays produced by Czech learners of English in their final year of highschool studies. Specifically, I address the following research questions:

1. How do the different measures of syntactic complexity correspond to the raters' judgments of essay quality?
2. How do the different measures of lexical complexity correspond to the raters' judgments of essay quality?

The definition of essay quality in this case refers to the respective grammar and vocabulary score assigned to each essay by a pair of trained raters.

2 Methodology

2.1 Data collection and analysis

Three local high schools took part in the research, specifically, the students in the final year of their studies. Each participating student completed the same task in the same conditions, responding to the statement *Some people think that teachers should be paid according to how much their students learn* within 45-minutes and using between 160-180 words. The time and word limits were determined by circumstances of the collection context – the written data collection had to happen with the students’ regular 45-minute school lesson. Though more time would have allowed the students more opportunity for planning their response, we can say that the task does mimic a typical school writing assignment, and so is able to add to the generalizability. Essays that were longer than 180 words (and so having exceeded the word limit) were trimmed to the nearest sentence near the word count in order to ensure that the essays were all within a similar range. Essays that were shorter than 160 words were removed from the process. This was a necessary step as the length of a text will have an impact on some of the complexity measurements. Also removed were essays written by non-native Czech speakers.

Two raters were recruited for the study, Rater A and Rater B. Rater A has 11 years of English language teaching experience, and Rater B has 8. Both raters have extensive experience preparing students for Cambridge suite exams – including the B2 Cambridge First, on which the marking rubric is based.

While other studies have either used or adapted rubrics from TOEFL (Kyle 2016), TOEFL-IBT (Biber et al. 2016), IELTS (Shadloo et al. 2019), and the Cambridge PET (Bi & Jiang 2020), an adapted Cambridge First (formerly FCE) rubric is used in this study. Cambridge First participants are marked according to B2 on the CEFR and so is a suitable choice as the participating schools assess their final year students at the same level. Furthermore, it was a tool that the two raters were already familiar with. It was necessary to adapt the rubric as in the existing rubric grammar and vocabulary are included together as part of the language ‘section’. With this study focusing on lexical complexity, it was necessary for the grammar and vocabulary-based descriptors to be separated. The separation of vocabulary and grammar means that each element has a ‘clean’ score – neither is affected by the other. This has the added bonus of allowing for future research on lexical and syntactic complexity and task fulfilment using the same set of data.

The raters were introduced to the adapted rubric and were then guided through an instruction manual to clarify terminology and to synchronize marking. The raters then entered their results into a prepared recording document. The rater agreement window was a difference no greater than one. In the instances that a score difference was greater than one, a third rater determined the score. Using Cohen's kappa, inter-rater reliability for vocabulary was ≈ 0.946 (or 97.06%), and for grammar $\kappa \approx 0.952$ (or 97.14%). Both values are very close to 1, indicating a high level of agreement between the raters. The essays were then processed using all indices on Lu's L2SCA and LCA (see Tables 1 and 2).

3 Results

3.1 Syntactic complexity

Each rater scored grammar out of five – with five being the highest possible score and one the lowest. These individual scores were then combined (see Table 4). The majority of the participants scored between six and eight points, following a similar pattern as the vocabulary scores.

Number of essays: 100	
Grammar score (/10)	Frequency
10	2
9	10
8	21
7	26
6	31
5	8
4	2

Table 4: Combined grammar score and frequency of occurrence

Of the 14 indices used by the L2SCA, 12 reported no significant correlation with grammar score (MLS, MLT, T/S, C/S, VP/T, C/T, DC/C, DC/T, CT/T, CP/T, CP/C, CN/T). Only MLC and CN/C showed a weak positive correlation ($r\ 0.245$, $p\ 0.014$ and $r\ 0.202$, $p\ 0.044$ respectively), indicating that as proficiency increases, so too does average clause length and use of complex nominals. That CN/C correlated yet CN/T did not is indicative of the nature of the measurements, the former being more granular (or 'fine-grained') and the latter offering a broader view. A criticism of the t-unit as a measure of complexity is that it can overshadow the effects of fine-grained features – it is simply too broad to be able to capture nuance. This is demonstrated in Appendix 3, where both a high

and low scoring essay and a comparison of their complexity measurements are presented. In the appendix, note that the lower scoring essay scores higher in all t-unit based measurements than the higher scoring essay.

Though previous studies have found L2SCA measurements capable of useful proficiency indicators, the general lack of correlation (positive or negative) in this study between L2SCA indices and grammar score is not too surprising, in particular the length-based measurements, given the homogeneity of the participant group. Appendix 3 shows that the lower scoring essay generally scores higher or similar to the higher scoring essay, serving as an indication of the weak or lack of correlation between grammar score and L2SCA indices. Another interpretation of these results could be that while each of the indices can contribute to the overall syntactic complexity of a text, they were not the best indicators of proficiency in the context of this study.

3.2 Lexical complexity

One hundred essays were processed. Each rater scored vocabulary out of five – with five being the highest possible score and one the lowest. These individual scores were then combined (see Table 5). The majority of the participants scored between six and eight points, with, interestingly, a similar pattern of outliers on either side.

Number of essays: 100	
Vocabulary score (/10)	Frequency
10	1
9	8
8	24
7	37
6	21
5	8
4	1

Table 5: Combined vocabulary score and frequency of occurrence

Tables 6 and 7 report only on the indices demonstrating a correlation with vocabulary score. Of the 25 indices, eleven reported no correlation with vocabulary score (LD, LS2, NDWZ, NDW-ER, NDW-ES, TTR, MSTTR, logTTR, AdjV, AdvV, and ModV). Of the results shown in Table 6, NDW (r 0.205), VV1 (r 0.215), LV (r 0.221), and NV (r 0.216) have a weak positive correlation. CTTR (r 0.262), RTTR (r 0.262), Uber (r 0.270), SVV1 (r 0.318), CVV1 (r 0.309), and VV2 (r 0.260) show a moderate positive relationship. The results show that

**IT'S COMPLICATED:
THE RELATIONSHIP BETWEEN LEXIS, SYNTAX, AND PROFICIENCY**

higher-evaluated essays were written with a greater lexical diversity. Overall, the indices that demonstrated the largest difference across the proficiency levels among the 25 lexical complexity measures were the verb indices and the root and corrected TTR measurements, having shown the stronger positive correlation to vocabulary score. This aligns with the idea that more sophisticated and varied language usage is characteristic of higher proficiency writers.

Table 7 shows that of the five lexical sophistication measures, LS1 (r 0.215), VS1 (r 0.242), and VS2 (r 0.219) showed a weak positive correlation, with CVS1 (r 0.263) showing a moderate positive relationship. These results show that as in McNamara et al. (2010) higher proficiency writers make more use of less frequently occurring, or, more sophisticated, language.

Index	Pearson's r p -value
NDW (number of different words)	0.205 0.041
Type-token ratio	
CTTR	0.262 0.008
RTTR	0.262 0.009
Uber	0.270 0.007
Verb diversity	
VV1	0.215 0.032
SVV1	0.318 0.001
CVV1	0.309 0.002
Lexical word diversity	
LV (lexical word variation)	0.221 0.027
VV2	0.260 0.009
NV	0.216 0.031

Table 6: Lexical variation measures

Index	Pearson's r p -value
LS1	0.216 0.031
VS1	0.242 0.0015
VS2	0.219 0.029
CVS1	0.263 0.008

Table 7: Lexical sophistication measures

A possible reason for these results is that less proficient writers use a writing style similar to that of spoken/conversational language. This corresponds with

Biber et al. (2011), who suggested that early developmental stages of complexity features are more likely to contain verbs commonly used in conversation. This also suggests that more capable writers are able to access and use a broader range and more specialised vocabulary, resulting in a better quality essay. It is also clear that though positive trends exist between some of the lexical complexity measurements and vocabulary scores, these trends are mostly weak – or at best, moderate. This would suggest that, while having some useful implications, lexical complexity alone is not a sufficient indicator of proficiency.

3.3 Developmental stages

A strong positive relation exists between grammar score and the participants that used structures found at developmental Stage 5 (r 0.411, p <.001), and also a weak negative relationship between Stage 1 and Stage 2 (r 0.172, p 0.043). Though a weak trend across scores, Stage 1 features appeared with greater frequency in essays that scored a grammar total of 4 (the lowest score in the data set) than any other score. Regarding the correlation between the developmental stages and vocabulary, again, only Stage 5 has a correlation (r 0.287, p 0.004), though one of weak-to-medium significance.

4 Discussion

The results of the syntactic complexity measurements showed that measures like MLC and CN/C had positive correlations with grammar scores. Longer clauses and the use of complex nominals can indicate a writer's ability to construct more intricate and layered sentences, reflecting a deeper understanding and command of the language.

With regards to the lexical complexity measurements, several practical classroom implications are suggested. Given the positive correlation between the sophistication and variation measures and vocabulary scores, it seems that students would benefit from teaching strategies designed to enhance their range and use of more sophisticated vocabulary beyond the 2,000 most common words, with particular attention being given to lexical items that are more likely to occur in academic writing than those that feature more heavily in speech.

Of particular note was the prominence of Stage 5 features in essays that scored higher in both vocabulary and grammar, suggesting that preposition and noun-controlled complements are indicative of grammatical proficiency, and that the ability to form such constructions is complemented with a wider lexical range. On reflection, this should not be too surprising, as Stage 1 is the use of complement clauses controlled by very common verbs – which would affect lexical complexity measures. It also implies that while the correct use of

foundational Stage 1 structures are essential, students should be encouraged to explore and incorporate more complex structures and vocabulary as they advance in their writing skills.

Appendices 2 and 3 show a comparison between the complexity measurements of a high and low scoring essay. It is interesting to observe how a straightforward comparison between two essays can reveal something about the predictive power of the complexity measurements. For example, the limitations of length-based measurements are revealed – the lower scoring essay has longer t-units and sentences than the higher. The difference in verb measurements are particularly telling, with the higher rated essay having a much higher degree of variation and sophistication.

Eleven lexical complexity indices reported no correlation; the lack of correlation suggests that these particular measurements, even though they are related to lexical complexity, might not be pivotal indicators of vocabulary proficiency. The absence of correlation for measures like LD (Lexical Density) might indicate that the overall density of lexical items in an essay is not as indicative of vocabulary proficiency as the sophistication or variety of words used. Similar comments can be made on the lack of correlation with most of the syntactic complexity measurements as used by the L2SCA. A possible reason for this absence is that while those features may be present in the essays, they are not necessarily determinants of proficiency scores. It could also indicate that the range of variation in these measures was insufficient in establishing a clear relationship.

Students aside, the findings are also of benefit to teachers or those working in test development. Complexity should be taken into consideration when setting a task, with regards to the type of language a student is expected to produce in response, and also during assessment – particularly when constructing an assessment rubric. By incorporating insights from linguistic complexity research, teachers and assessors can create a more targeted and specific evaluation tool.

5 Conclusion

This study investigated the relationship of 14 measures of syntactic complexity (Table 1), 25 measures of lexical complexity (Table 2) and the five stages of a hypothesised complexity developmental route (Table 3) to the grammar and vocabulary scores of L1-Czech L2-English high school student argumentative essays, using analytical software and approaches to identify patterns between grammar and vocabulary scores and complexity measures, and so providing insights into features of students' writing abilities at different levels of proficiency. The data suggests a link between lexical and syntactic complexity

measurements, in that essays with a richer vocabulary and more advanced syntactic structures generally received higher vocabulary and grammar scores respectively.

These findings can aid teachers in designing writing courses and materials to enhance the writing skills of more advanced students while also addressing the needs of less skilled students. In addition, the findings can also support rater training as well as incorporating automated tools as part of assessment and evaluation procedures. Complexity should be taken into consideration when setting a task, with regards to the type of language a student is expected to produce in response, and also during assessment – particularly when constructing an assessment rubric. By incorporating insights from linguistic complexity research, teachers and assessors can create a more targeted and specific evaluation tool.

Though offering such insights, this study was not without limitations. First, this research was completed with the use of a rather homogenous group of participants – they were all of a similar age and academic background. Future studies would benefit from the inclusion of different age groups and/or a wider range of English proficiency. A wider participant range may more starkly reveal correlations between the complexity measurements and proficiency, potentially leading to more generalisable results.

The essay length and 45-minute time limit were necessary conditions of the collection context; however, future research can investigate the effects of longer time allowance. On constraints, it is likely that the word limit affected the output – a higher word count, or even removing the limit, would have possibly allowed some participants to write much more rather than feeling confined.

The relationship between syntactic and lexical complexity remains worthy of investigation, as no single independent measure can truly capture and inform on the quality of writing as a whole – a good essay is much more than a diverse range of sophisticated vocabulary. It is also worth reiterating that the quality of writing is determined by several factors, including accuracy and syntactic complexity, as well as task type, genre, and sociolinguistic factors – all factors that can be used to guide future research.

Finally, effective writing skills are important for the clear communication of ideas. As students gain a deeper understanding of the components that make up quality writing, they will be more able to effectively incorporate them into their writing.

References

- Beers, S. and Nagy, W. (2009) 'Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre?' *Reading and Writing* 22(2), 185-200. <https://doi.org/10.1007/s11145-007-9107-5>
- Bi, P. and Jiang, J. (2020) 'Syntactic complexity in assessing young adolescent EFL learners' writings: Syntactic elaboration and diversity.' *System* 91, 102248. <https://doi.org/10.1016/j.system.2020.102248>
- Biber, D., Gray, B. and Poonpon, K. (2011) 'Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?' *TESOL Quarterly* 45(1), 5-35. <https://doi.org/10.5054/tq.2011.244483>
- Biber, D., Gray, B. and Staples, S. (2016) 'Predicting patterns of grammatical complexity across language exam task types and proficiency levels.' *Applied Linguistics* 37(5), 639-668. <https://doi.org/10.1093/applin/amu059>
- Biber, D., Gray, B., Staples, S. and Egbert, J. (2020) 'Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement.' *Journal of English for Academic Purposes* 46, 100869. <https://doi.org/10.1016/j.jeap.2020.100869>
- Bulté, B. and Housen, A. (2014) 'Conceptualizing and measuring short-term changes in L2 writing complexity.' *Journal of Second Language Writing* 26, 42-65. <https://doi.org/10.1016/j.jslw.2014.09.005>
- Carroll, J. B. (1964) *Language and Thought*. Englewood Cliffs: Prentice-Hall.
- Casal, E. J. and 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>
- Chen, H., Xu, J. and He, B. (2014) 'Automated essay scoring by capturing relative writing quality.' *The Computer Journal* 57(9), 1318-1330. <https://doi.org/10.1093/comjnl/bxt117>
- Crossley, S. A., Salsbury, T., McNamara, D. S. and Jarvis, S. (2011) 'Predicting lexical proficiency in language learner texts using computational indices.' *Language Testing* 28(4), 561-580. <https://doi.org/10.1177/02655322103780>
- Crossley, S. A., Cai, Z. and McNamara, D. (2012) 'Syntagmatic, paradigmatic, and automatic n-gram approaches to assessing essay quality.' In: McCarthy, P. M. and Youngblood, G. M. (eds) *Proceedings of the 25th International Florida Artificial Intelligence Research Society (FLAIRS) Conference*. AAAI Press. 214-219.
- Crossley, S. A. and McNamara, D. S. (2014) 'Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners.' *Journal of Second Language Writing* 28, 66-79. <https://doi.org/10.1016/j.jslw.2014.09.006>
- Crossley, S. A. (2020) 'Linguistic features in writing quality and development: An overview.' *Journal of Writing Research* 11(3), 415-443. <https://doi.org/10.17239/jowr-2020.11.03.01>
- Daller, H., Van Hout, R. and Treffers-Daller, J. (2003) 'Lexical richness in the spontaneous speech of bilinguals.' *Applied Linguistics* 24(2), 197-222. <https://doi.org/10.1093/applin/24.2.197>
- Davies, M. (2008) *The Corpus of Contemporary American English: 425 million words, 1990-present*. corpus.byu.edu/coca/
- Grant, L. and Ginther, A. (2000) 'Using computer-tagged linguistic features to describe L2 writing differences.' *Journal of Second Language Writing* 9(2), 123-145. [https://doi.org/10.1016/S1060-3743\(00\)00019-9](https://doi.org/10.1016/S1060-3743(00)00019-9)

- Guiraud, P. (1960) *Problemes et Methodes de la Statistique Linguistique*. Paris: Presses universitaires de France.
- Housen, A., Kuiken, F., and Vedder, I. (2012) 'Complexity, accuracy and fluency: Definitions, measurement and research.' In: Housen, A., Kuiken, F. and Vedder, I. (eds) *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA Language Learning & Language Teaching*. Amsterdam and Philadelphia: John Benjamins.1-20.
- Jarvis, S., Grant, L., Bikowski, D. and Ferris, D. (2003) 'Exploring multiple profiles of highly rated learner compositions.' *Journal of Second Language Writing* 12, 377-403. <https://doi.org/10.1016/j.jslw.2003.09.001>
- Johansson, V. (2008) 'Lexical diversity and lexical density in speech and writing: A developmental perspective.' *Working Papers* 53, 61-79.
- Johnson, M. D., Acevedo, A. and Mercado, L. (2013) 'What vocabulary should we teach?: Lexical frequency profiles and lexical diversity in second language writing.' *Writing and Pedagogy* 5(1), 83-103. <https://doi.org/10.1558/wap.v5i1.83>
- Johnson, M. D., Acevedo, A. and Mercado, L. (2016) 'Vocabulary knowledge and vocabulary use in second language writing.' *TESOL Journal* 7(3), 700-715. <https://doi.org/10.1002/tesj.238>
- Kim, J. (2014) 'Predicting L2 writing proficiency using linguistic complexity measures: A corpus-based study.' *English Teaching* 69(4), 27-51. <https://doi.org/10.15858/engtea.69.4.201412.27>
- Klein, D. and Manning, C. D. (2003) 'Fast exact inference with a factored model for natural language parsing.' In: Becker, S., Thrun, S. and Obermayer, K. (eds) *Advances in Neural Information Processing Systems 15*. Cambridge: MIT Press. 3-10
- Kyle, K. (2016) *Measuring Syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-based Indices of Syntactic Sophistication*. (Doctoral dissertation). Online document. Retrieved on 23 November 2023 <http://scholarworks.gsu.edu/alesl_diss/35>.
- Kyle, K. and Crossley, S. A. (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>
- Khushik, G. A. and Huhta, A. (2019) 'Investigating syntactic complexity in EFL learners. Writing across Common European Framework of Reference Levels A1, A2, and B1.' *Applied Linguistics* 41(4), 506-532. <https://doi.org/10.1093/applin/amy064>
- Lee, C., Ge, H. and Chung, E. (2021) 'What linguistic features distinguish and predict L2 writing quality? A study of examination scripts written by adolescent Chinese learners of English in Hong Kong.' *System* 97, 102461. <https://doi.org/10.1016/j.system.2021.102461>
- Lu, X. (2010) 'Automatic analysis of syntactic complexity in second language writing.' *International Journal of Corpus Linguistics* 15, 474-496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- 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>
- Lu, X. (2012) 'The relationship of lexical richness to the quality of ESL learners' oral narratives.' *The Modern Language Journal* 98(2), 190-208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x

- Lu, X. and Ai, H. (2015) 'Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds.' *Journal of Second Language Writing* 29, 16-27. <https://doi.org/10.1016/j.jslw.2015.06.003>
- Lu, X. (2017) 'Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment.' *Language Testing* 34(4), 493-511. <https://doi.org/10.1177/0265532217710675>
- Mazgutova, D. and Kormos, J. (2015) 'Syntactic and lexical development in an intensive English for Academic Purposes programme.' *Journal of Second Language Writing* 29, 3-15. <https://doi.org/10.1016/j.jslw.2015.06.004>
- McNamara, D. S., Crossley, S. A. and Roscoe, R. (2013) 'Natural language processing in an intelligent writing strategy tutoring system.' *Behavior Research Methods* 45, 499-515. <https://doi.org/10.3758/s13428-012-0258-1>
- Nation, I. (2006) 'How large a vocabulary is needed for reading and listening?' *Canadian Modern Language Review* 63(1), 59-82.
- Ortega, L. (2015) 'Syntactic complexity in L2 writing: Progress and expansion.' *Journal of Second Language Writing* 29, 82-94. <https://doi.org/10.1016/j.jslw.2015.06.008>
- Paquot, M. (2018) 'Phraseological competence: A missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations.' *Language Assessment Quarterly* 15(1), 29-43. <https://doi.org/10.1080/15434303.2017.1405421>
- Qin, W. and Uccelli, P. (2016) 'Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance.' *Journal of Second Language Writing* 33, 3-17. <https://doi.org/10.1016/j.jslw.2016.06.001>
- Shadloo, F., Ahmadi, H. S. and Ghonsooly, B. (2019) 'Exploring syntactic complexity and its relationship with writing quality in EFL argumentative essays.' *Topics in Linguistics* 20(1), 68-81. <https://doi.org/10.2478/topling-2019-0005>
- Torruella, J. and Capsada, R. (2013) 'Lexical statistics and typological structures: A measure of lexical richness.' *Procedia – Social and Behavioral Sciences* 95, 447-454.
- Vermeer, A. (2000) 'Coming to grips with lexical richness in spontaneous speech data.' *Language Testing* 17(1), 65-83
- Wolfe-Quintero, K., Inagaki, S. and Kim, H. Y. (1998) *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. Honolulu, HI: University of Hawaii, Second Language Teaching & Curriculum Center.
- Yang, W., Lu, X. and Weigle, S. (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>
- Yoon, H. J. and Polio, C. (2016) 'The linguistic development of students of English as a second language in two written genres.' *Tesol Quarterly* 51(2), 275-301. <https://doi.org/10.1002/tesq.296>
- Yoon, H. (2017) 'Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality.' *System* 66, 130-141. <https://doi.org/10.1016/j.system.2017.03.007>

Appendix 1: Verbs counted as common

According to Biber et al. (1999) these verbs are more likely to occur in speech than writing, which are then categorised as common or very common. No distinction is made between common and very common in this study.

<i>ask</i>	<i>listen</i>	<i>send</i>	<i>come</i>	<i>lose</i>	<i>stand</i>
<i>bring</i>	<i>live</i>	<i>sit</i>	<i>eat</i>	<i>love</i>	<i>stay</i>
<i>buy</i>	<i>look</i>	<i>speak</i>	<i>feel</i>	<i>make</i>	<i>stop</i>
<i>give</i>	<i>meet</i>	<i>take</i>	<i>get</i>	<i>mean</i>	<i>suppose</i>
<i>go</i>	<i>pay</i>	<i>talk</i>	<i>happen</i>	<i>pick</i>	<i>tell</i>
<i>hear</i>	<i>play</i>	<i>thank</i>	<i>keep</i>	<i>put</i>	<i>think</i>
<i>know</i>	<i>remember</i>	<i>try</i>	<i>leave</i>	<i>run</i>	<i>turn</i>
<i>let</i>	<i>say</i>	<i>want</i>	<i>like</i>	<i>see</i>	

Appendix 2: High and low-scoring essays

Two sample essays are displayed. The first essay, J4, scored highly in grammar and vocabulary. The second essay, J1, scored much lower. Following the essays, the scores and measurement results are presented side by side for comparison.

Essay J4

There are some people that think teachers' salary should be higher or lower depending on how much they are able to teach their students. The question is, should this be implemented or not?

There would possibly be many benefits if this was to be used. It could give teachers a push to start working harder and really pay attention to their students. There is no doubt some of them teach just for the sake of teaching and are forgetting their responsibilities for their students' futures. But it shouldn't be surprising considering how much they are getting paid and even when they try to work harder the result is still the same. For those who find joy in this field would raise be a motivation to keep going and it would also improve the overall level of knowledge in that class. Of course, another factor is the teacher's ability in transferring information to the students and students' will to learn. But then, a good teacher should be able to handle a whole class.

Therefore, the final conclusion is that by implementing this method, most of the teachers could be motivated.

Essay J1

For some people this topic could be sensitive, but i think it's a kinda important to disccus it, because the point of this job is to teach students something and there's a big amount of teachers who sadly can't do it.

These days there is a plenty of jobs where are people paid by the job they do or more like how good, the job they did is. For example if you are an architect and you do some project that is not really good or convenient you won't get that good. So why couldn't that be for teachers too? Why would they do the job if students don't learn anything. It's just pointless to not do anything and still get the same money as someone who is really trying and just works hard to teach students something.

I think these days there's plenty of teachers who can't really teach and yeah, I mean sometimes they don't know it, but that doesn't mean that it is okay. They can always ask students for a feedback and try to figure it out.

Combined rater assigned scores

Measure	J4	J1
Grammar	10	4
Vocabulary	9	5

Appendix 3: Complexity measurements of the high and low-scoring essays presented in appendix 2

L2SCA measures

Measure	J4	J1
MLS	19.4	24
MLT	13.9	21.3
MLC	8.4	6.9
C/S	2.3	3.5
VP/T	2.6	3.7
C/T	1.6	3.1
DC/C	0.3	0.5
DC/T	0.6	1.6
T/S	1.4	1.1
CT/T	0.4	0.8
CP/T	0.2	0.6
CP/C	0.1	0.2
CN/T	1.4	2.2
CN/C	0.9	0.7

Lexical density

Measure	J4	J1
LD	0.45	0.48

Lexical sophistication

Measure	J4	J1
LS1	0.13	0.12
LS2	0.10	0.12
VS1	0.11	0.04
VS2	0.33	0.04
CVS1	0.41	0.14

Developmental stages

Measure	J4	J1
Stage 1	1	3
Stage 2	3	3
Stage 3	2	5
Stage 4	2	2
Stage 5	6	1

Lexical variation

Measure	J4	J1
NDW	99	82
NDWZ-50	36	37
NDW-ER50	37.4	36.5
NDW-ES50	37.7	33.6
TTR	0.51	0.43
MSTTR-50	0.75	0.73
CTTR	5.03	4.18
RTTR	7.11	5.92
LOGTTR	0.87	0.84
UBER	17.91	14.11
VV1	0.85	0.48
SVV1	19.59	6.26
CVV1	3.13	1.77
LV	0.8	0.56
VV2	0.26	0.14
NV	0.74	0.56
ADJV	0.13	0.009
ADV	0.11	0.14
MODV	0.24	0.23

Christopher Williams is a doctoral student and language centre instructor at the Faculty of Education, Masaryk University, Brno. His dissertation is focused on the role of syntactic and lexical complexity in the argumentative essays of highschool students.

Address: Christopher Williams, Department of English Language and Literature, Faculty of Education, Masaryk University, Pořící 9, Brno 603 00, Czech Republic. [email: chris.williams@mail.muni.cz]