2022-12-15

Analyzing Patterns of Complexity in Pre-University L2 English Writing

Zachary M. Lambert
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Arts and Humanities Commons

Analyzing Patterns of Complexity in Pre-University L2 English Writing

Zachary M. Lambert

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Grant T. Eckstein, Chair
Brett J. Hashimoto
Benjamin L. McMurry

Department of Linguistics

Brigham Young University

ABSTRACT

Analyzing Patterns of Complexity in Pre-University L2 English Writing

Zachary M. Lambert
Department of Linguistics, BYU
Master of Arts

The present study involves the creation and analysis of a corpus containing 7747 samples of timed, pre-matriculated L2 English writing from an IEP. The focus of the analysis is on three phrasal complexity measures with time and proficiency as fixed effects, examining their impact on each measure. Results of the analysis suggest that this is true for some phrasal measures, such as nominalizations and attributive adjectives, while others, such as noun-noun phrases, may indicate a lower level of proficiency or lesser allotted writing time. Nominalizations in particular demonstrated a strong relationship with both allotted time and proficiency, further suggesting that certain phrasal measures may be more useful than others when examining academic writing, and therefore may merit additional focus and time spent on related structures in IEPs and other pre-matriculated ESL/EFL classrooms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Studies of syntactic complexity over the years have revealed numerous differences between L1 and L2 texts. Syntactic complexity is defined by Ai and Lu (2013) as "the range and degree of sophistication of syntactic structures that surface in language production" (p.249). Silva (1993) summarized research indicating that L2 writers produced a greater number of T-units per sentence than their L1 counterparts, with each L2 T-unit being significantly shorter on average than in L1 writing. Hinkel (2003) demonstrated that L2 writers used simple be-copula forms in main verbs significantly more frequently than L1 writers, leading to observations of L2 texts as more simplistic. More recently, computational approaches have expanded the possibilities for complexity measures. Lu and Ai (2015) identified significant differences between L1 writers and a heterogeneous group of L2 writers on measures of clause length and complex nominals per clause and T-unit. Ai and Lu (2013) found additional differences when comparing the syntactic complexity of L1 writers with Chinese L2 writers. They identified differences in length of written production as well as amounts of subordination, coordination, and phrasal sophistication. Together, these findings portray L2 writing as less syntactically complex than L1 writing (Staples & Reppen, 2016).

Researchers have further studied how students develop complexity as their language skills increase. Olinghouse & Wilson (2013) evaluated the role of vocabulary in fifth grade L1 writing across three genres, noting that in each genre, vocabulary was implemented in a manner distinct from the other genres. Duran et al. (2004) proposed a mathematical approach to measuring lexical diversity and used their model to demonstrate lexical development in young L1 English speakers. Johansson (2008) proposed using text characteristics, including lexical density and grammatical structures, to evaluate student writing to avoid the subjectivity of rubrics. Crossley & McNamara (2014) examined college-level L2 writing, searching for patterns

between syntactic features used and perceived writing quality, and noting that raters may be scoring inherently nonacademic structures higher than academic structures.

Evidence from these studies shows that students indeed develop complexity in writing as their proficiency increases, though the improvement appears smaller for foreign language students compared to second language students studying in an English-speaking environment, as the former tend to demonstrate more lexical gains (Kim, 2021). Furthermore, evidence of syntactic improvement seems to be best observed after a full year of proficiency development (Crossley and McNamara, 2014). This makes it difficult to see and measure syntactic complexity by merely sampling student writing at the beginning and end of a given semester. Additionally, while both academic writing and spoken registers display complex grammatical structures (Biber & Gray, 2010), rhetorical purpose appears to influence complexity since more complex academic writing tends to have greater *phrasal* complexity while interpersonal communication tends to have greater *clausal* complexity (Biber et al., 2011; Staples et al., 2016).

These findings suggest that T-unit measures are inadequate for evaluating academic writing, and that phrasal measures generally act as more accurate indicators of said writing (Biber et al., 2013; Casal & Lee, 2019). A study by Taguchi et al. (2013) reporting that "number of T-units [do] not distinguish [the] written performance" of nonnative English speakers further supports this notion (p.420). Accordingly, many more recent studies have begun implementing phrasal measures in addition to typical clausal measures. Atak and Saricaoglu (2021) and Saricaoglu and Atak (2022) included phrasal measures involving prepositional phrases, attributive adjectives, noun-noun phrases, and possessives, and demonstrated that the former two features were implemented more frequently in argumentative L2 writing. Qin and Zhang (2022) noted that phrasal noun modifiers are more associated with register flexibility in writing than

clausal modifiers, while Li et al. (2022) reported greater phrasal complexity and less clausal complexity among Chinese EFL students as they progressed further into the program.

This pattern of phrasal versus clausal complexity is quite meaningful, but requires an understanding of what each of these terms entail to fully appreciate. Clausal complexity involves the embedding or attaching of subordinate clauses to the main clause via the use of conjunctions or other subordinators. An example of clausal complexity:

'We wanted to measure changes **that we observed** in **what was making up the soil**.'

Phrasal complexity, on the other hand, focuses on the noun phrase. Instead of using subordinators to connect clauses together, the desired information is instead compressed into a single noun phrase. An example of phrasal complexity:

'We were interested in **changes in soil composition**.'

As can be seen in these examples, phrasally complex language is typically more succinct when compared to clausally complex language communicating similar information (hence, the term 'compression' is often associated with phrasal complexity). Additionally, some phrasally complex noun structures, such as nominalizations, are considered to be more abstract than their clausal counterparts (e.g., 'soil composition' vs. 'what was making up the soil'; Gray et al., 2019). These factors may help explain the pattern noted by Biber et al. (2011), considering both word limits and high register language are commonplace in academia.

A corpus-based study conducted by Staples et al. (2016) also supports Biber et al.'s claim regarding academic writing, showing that the use of various phrasal complexity features increases with academic level, while the use of many clausal complexity features, particularly finite dependent clauses, decreases as academic level increases. Using a separate corpus, Staples

and Reppen (2016) found that language ratings of academic writing were related to both lexical and grammatical features. Of particular note from this study is the reliance found of both L2 groups (one L1 Arabic and one L1 Chinese) on repetition of key phrases for essay cohesion, with L1 Arabic writers using noun-noun and adjective-noun sequences and L1 Chinese writers using premodifying nouns in this manner. However, these studies all focus on university-level writing.

Pre-Matriculated Writers

Past studies of phrasal complexity have focused almost exclusively on college-level writers. Recently, however, there has been an increase in studies considering pre-matriculated English writers in a variety of topics, though much work remains to be done in this area. Kitajroonchai et al. (2022) explored process writing approaches for pre-university EFL students, though did not consider complexity in their study. Hartshorn et al. (2010) examined the effects of dynamic written corrective feedback (DWCF) on the complexity, accuracy, and fluency of pre-matriculated learners attending an IEP. However, the researchers chose mean length of T-units as the sole complexity measure for analysis, limiting the generalizability of their complexity findings. Zhang and Cheng (2021) studied the effects of written corrective feedback (WCF) on the complexity, accuracy, and fluency of Chinese EAP students, but again found that T-unit-based syntactic complexity was not enhanced by its usage. Martinez (2017) and Martinez (2018) evaluated phrasal and clausal complexity among secondary-level English writers, comparing differences across L1 and L2 writers and proficiency levels, respectively. Both studies utilized the uncommon phrasal measure of mean length of noun phrase and found that said measure seemed to increase as grades and proficiency increased, respectively.

As many students are expected to produce moderately academic writing in their first year of university, pre-matriculated writers attending an IEP or college preparation course should be taught the relevant skills needed to do so. Such skills include competent usage of attributive

adjectives and nominalizations (Hinkel, 2002). Additionally, beginning L2 learners often undergo rapid language development, forming and solidifying language habits that may or may not prove helpful in their future studies (Vyatkina et al., 2015). This study focuses on such learners, evaluating their implementation of phrasal complexity in their writing to help determine whether they are being adequately prepared for university-level writing tasks.

Timing in Writing

This study also seeks to explore the effects of timing on phrasal complexity. While past complexity studies have implemented timing in their research, no prior study has compared complexity across different time frames. As noted previously, Hartshorn et al. (2010) utilized timed written responses as data sources for the studies. Gray et al. (2019) evaluated a mixture of clausal and phrasal complexity features in both spoken and written TOEFL test responses, all of which are timed. Fathi and Rahimi (2022) followed a contemporary Complexity Accuracy Fluency (CAF) model when examining 50-minute timed essays written by L2 university students. It is notable that the only area lacking statistical significance in this study was complexity, where, similar to Hartshorn et al. (2010), Fathi and Rahimi utilized the model first outlined by Skehan (1998), implementing measures based on T-units in lieu of more in-depth measures. As Biber et al. (2011) have presented compelling evidence that phrasal measures more accurately describe complexity patterns found in academic writing, the present study includes timing as a factor to test whether measuring phrasal complexity instead of clausal may show significance in a timed environment.

The ability to produce quality writing within a given timeframe is a metric commonly used by both universities and employers when evaluating candidates (e.g., the GRE, TOEFL, and other standardized exams), and is therefore frequently practiced and tested by IEPs and other language programs. Examining the writing produced by pre-matriculated learners under these

different timing conditions, such as 10- or 30-minute tasks, is likely to reveal different levels of writing competence among learners via patterns in their response, such as the grammatical structures and features that are prioritized or sacrificed when limited by time.

Corpus Linguistics and Writing Complexity

Researchers have recently turned to corpus linguistics methods in an attempt to understand how writing complexity develops as a function of proficiency and is further mitigated by rhetorical purpose and learning context. Corpora are large samples of text which can be analyzed computationally to make observations about features of those texts. Some researchers have recently begun developing a multi-institutional corpus and repository (CROW) of learner texts that can be analyzed for complexity development (Staples & Dilger, 2018). The CROW project includes thousands of deidentified essays from primarily L1 students taking first-year English composition classes at several major universities. The CROW resources are ideal for many types of analyses, but because they are limited almost exclusively to first-year L1 writing, it is difficult to examine developmental language or writing done in short, timed environments.

To aid in this, we have compiled thousands of timed writing samples authored by ESL writers of various L1 backgrounds into a new corpus for analysis. The samples were provided by the English Language Center, an Intensive English Program (IEP) run by Brigham Young University. These written responses reflect developmental writing by pre-matriculated students at levels below first-year composition.

For our specific use of the data, we aim to answer the following research question:

1. In ESL academic writing, to what extent do learners implement attributive adjectives, noun-noun phrases, and nominalizations differently across two different timing conditions (10-minute and 30-minute written responses)?

2. To what extent do ELLs of different proficiency levels differ in their use of these three phrasal complexity measures in their academic writing?

Methods

Essays

The corpus consists of 7747 files timed responses written by students in the IEP that have been archived since 2016. The students come from various L1 backgrounds and English proficiency levels. The responses were written as part of the IEP's in-house final proficiency exam which evaluates students' English proficiency across reading, writing, listening, speaking, and grammar. The written portion is separated into 10-minute and 30-minute essays, with the 10-minute prompts[1] designed to elicit Novice or Intermediate-Low level responses (based on the ACTFL guidelines, e.g., paragraph-level texts focusing on concrete elaborations and personal anecdotes) and the 30-minute essays requiring Intermediate-High to Advanced level responses (e.g., multi-paragraph essays with clear formatting, abstract elaborations, and complete ideas).

Prior to analyzing the corpus, all samples with less than 100 words were removed from the analysis. All measure frequencies were normalized to 100 words for analysis, and responses shorter than this largely failed to demonstrate any diversity in syntactic structure. Thus, the data risked being heavily skewed and potentially invalid without the cutoff. Of the 7747 initial files, 724 were removed from consideration based on this cutoff, leaving 7,023 files of sufficient length.

Demographics

---

[1] As the English Language Center is still using these prompts as part of their final exams, we have been asked to not go into specific detail about their content so as to not compromise their testing process.

Demographic information, including assigned proficiency scores, was provided to us by the IEP based on information gathered by the host university. However, as the IEP had previously assigned students different IDs than the host university, the demographics initially did not match with any of the responses. A script was written to match student demographics and student responses using available records. Of the 7,023 written responses still in consideration, 146 did not correspond to any available demographic information, while 6,877 were successfully matched with their corresponding demographics and scores. Said demographics show that the 6,877 responses were written by 1,332 different students, with an average of 5.16 responses per student (min 1, max 18). As most students remain at the IEP for between two to four semesters and write two responses each semester, this is expected.

As seen in Table 1, of the 6,877 responses, 3,696 were written by females and 3,181 by males. As the learners are pre-matriculated, age is skewed towards the younger side, with nearly 83% of responses written by students below age 31. In terms of L1, Spanish is by far the most common, with Portuguese, Japanese, Chinese, and Korean following. A total of 33 different L1s were recorded across the 1,332 learners, including Arabic, Farsi, Tajik, Haitian Creole, Hungarian, Mongolian, Bambara, Turkmen, and others.

Table 1

Demographics

L1	Essays	L1	Essays		Age	Essays
Spanish	4064	Mongolian	100		18-24	3982
Portuguese	760	French	91		25-30	1716
Japanese	554	Russian	89		31-40	872
Chinese	544	Creole	80		40+	307
Korean	357	Thai	42			
Other (less than 30 responses)	196				**Gender**	**Essays**
					Female	3696
Total Languages	33				Male	3181

IEP Rating System

To determine the writing proficiency of the students, their essay scores were considered. The IEP scores exam responses based on a modified ACTFL scale, providing a numerical score from 0-7 based on student performance and demonstrated ability. Accordingly, the given scores roughly correspond to the full range of standard proficiencies from Novice-Low to Advanced-High. The IEP's ratings are performed by its teachers, who have a variety of teaching and L1 backgrounds. These teachers primarily come from L1 English backgrounds and include TESOL MA students, undergraduate students in the TESOL minor, recent graduates, and other teachers. Their experience ranges from one semester of student teaching to several years of teaching in a variety of programs and contexts. With regard to assessment, these teachers similarly vary from little to no previous training or experience to years of assessment development and practice.

Each semester, teachers are assigned to rate either written or spoken prompts. Because there are almost always different or new raters each semester, all raters participate in training on the modified scale through individual rating practice and follow-up calibration meetings to ensure consistency across ratings. The individual practice has the teachers rate between five and seven responses written by previous students which have already been double-rated and scored. After each teacher submits the ratings to a supervisor, all ratings for each prompt are compiled and presented, maintaining anonymity. Any major discrepancies are discussed so that individual raters can identify rating biases such as central tendency.

Once said biases are addressed, raters are assigned a number of student responses to review and rate. The responses are all double-rated, and any significant discrepancy between official ratings leads to a third rating. Each student is given one combined writing score for their responses across both prompts. Because raters vary so much in experience and not all their biases

can reasonably be expected to be controlled while rating, a polytomous Rasch model is used to account for potential variation in rater severity. This model also considers the difficulty of the writing task in conjunction with the students' language abilities. The proficiency data used in this study are those reported after the Rasch model was applied.

Corpus Development

Before tagging the essays, the corpus underwent a deidentification process. Students often included personal identifying information (i.e., names of individuals, cities of residence or birth, and important dates such as birthdays or graduation dates) in their responses, particularly in the 10-minute essays. Accordingly, each file was processed through a de-identification tool provided by the creators of the CROW Corpus and then further examined by trained project members, with said information removed in order to protect the privacy of each writer.

All corpus files were then tagged for parts of speech using the Biber Tagger (Biber, 1988), which uses a combination of contextual rules, large-scale dictionaries, and probabilistic information to annotate each word in a text with morphological, syntactic, and some semantic information (see Biber, 1988). A comprehensive accuracy analysis of the tagger was previously carried out on the TOEFL iBT PublicUse data set (Biber & Gray, 2013) and the tagger demonstrated high accuracy, with most features demonstrating precision and recall rates greater than .90, and many greater than .95. The corpus used in this study also consists of ESL learner language created in a testing environment, and therefore it can be expected that such tagging accuracy will be maintained in this project.

That said, as should be expected when tagging learner language, not all tags given by the program were fully accurate. To compensate for this, additional research personnel were hired and trained to carry out fixtagging, correcting any mistakes made by the tagger when tagging misspelt or misused words typical of learner language (e.g., "espesiary" to "especially" and "my

passion is ingenerd" to "my passion is engineering"). This fixtagging process involved opening each file using the Biber Tag Checking Tool, an open-source tool developed by the CROW team that is publicly available on the GitHub Repository (Picoral et al., 2018-). Regular expressions were used to identify tokens with specific tags, and tokens that appear with incorrect or incomplete tags were corrected. Corrected files were then saved into a new 'tagchecked' version of the corpus, ensuring that we retained the original files while having access to a cleaned-up version of the data. Multiple passes of fixtagging occurred, with particular emphasis on tokens marked as problematic by the tagger, tokens labelled as nominalizations, and tokens ending *-ed* and *-ing*. All fixes were recorded and tracked throughout the process.

Features Considered

Historically, clausal complexity measurements, specifically those involving T-units, have been used to evaluate academic writing samples from both L1 and L2 learners (Wolfe-Quintero et al, 1998). However, recent studies have expressed skepticism at their overall effectiveness (Biber et al. 2011; Staples et al. 2016; Staples & Reppen 2016). Biber et al. (2011) report that "most clausal subordination measures are actually more common in conversation than academic writing. In contrast, fundamentally different kinds of grammatical complexity are common in academic writing: complex noun phrase constituents (rather than clause constituents) and complex phrases (rather than clauses)."

Staples et al. (2016) highlight a related trend among university-level L1 writing, with data from the British Academic Written English corpus (BAWE) showing that "as academic level increases, the use of phrasal complexity features in writing also increases. On the other hand, the use of clausal complexity features in student writing…decreases as academic level increases" (p.17). This is in line with a parallel study by Staples and Reppen (2016), which finds that while "first-year writers, both L1 and L2, use lexical and grammatical patterns in particular

ways that are not always consistent with the trends found in published academic writing," these writers "have identified these important [phrasal] features of academic writing and are using them in their texts" (p.31), though the L2 writers appear to overuse many of these features at the intermediate level.

Many researchers have thus incorporated phrasal measures from these and other studies into their analyses. Biber & Gray (2013) and Gray et al. (2019) both evaluated the phrasal and clausal complexity of speaking and writing tasks on the TOEFL, implementing numerous phrasal measures including nominalizations, attributive adjectives, and adverbials. Biber et al. (2016) used similar measures when comparing the grammatical structures used in conversation and academic writing. Taguchi et al. (2013) examined which measures were indicative of writing quality at the university level, and included attributive adjectives, adverbials, and prepositional phrases in their analysis. As no other corpus study thus far has consisted of academic writing from pre-university ESL students, this is a great opportunity to examine how phrasal complexity develops alongside student language proficiency and whether it is consistent with the pattern noted by Biber et al. (2011), that is, that complex phrases and phrase constituents are common in academic writing.

In considering which measures to use, various studies were examined for statistically significant results returned when evaluating phrasal complexity measures. Both nouns as noun premodifiers and attributive adjectives have been consistently significant and informative measures across numerous studies (Biber & Gray 2013; Biber et al. 2016; Parkinson & Musgrave 2014). Furthermore, use of nominalizations is an indicator of abstract language (Gray et al. 2019), the production of which is considered an 'Advanced' or 'Superior' skill according to the ACTFL scale and is frequently required in college-level writing tasks. Taking these results

into consideration, this paper will utilize all three of these phrasal measures (noun-noun phrases, attributive adjectives, and nominalizations) in its analysis.

Analysis

The corpus data was analyzed in R using linear mixed effects models (LMMs), with Participant as a random effect and Time (10-minute responses compared to 30-minute responses) and Proficiency (whether the essay was rated as a Novice-, Intermediate-, or Advanced-level work) as fixed effects, with $p < 0.05$ as the threshold for significance. The interaction between time and proficiency could not be considered using the data provided by the IEP as the proficiency scores were assigned to students, not to responses, i.e., each student received one combined score based on both the 10- and 30-minute response submitted.

To account for potential errors, multiple models—both LMMs and regressions—were created and tested for best fit using the models' R-squared values. In all cases, the LMMs with participant as a random effect returned better fits than models without it or with no random effects. Following this, iterations of various random effect combinations occurred, considering age, L1, and gender alongside the aforementioned participant. Gender did not improve fit in any model and was discarded, participant and L1 improved fit for all models, and age improved fit for attributive adjectives and noun-noun phrases but had an effect of <.001 on fit for nominalizations. Furthermore, all models attempting to evaluate random slopes failed to converge, meaning all models used consider random intercepts only.

The LMMs used follow three assumptions:

- Linearity
- Homogeneity of variance
- Normality of error

These assumptions were checked in each of the three models, and all models displayed similar patterns for each assumption.

The first assumption is tested by plotting the residuals against the feature being considered. If the resulting distribution appears random in nature, the linearity assumption is met.

Figure 1
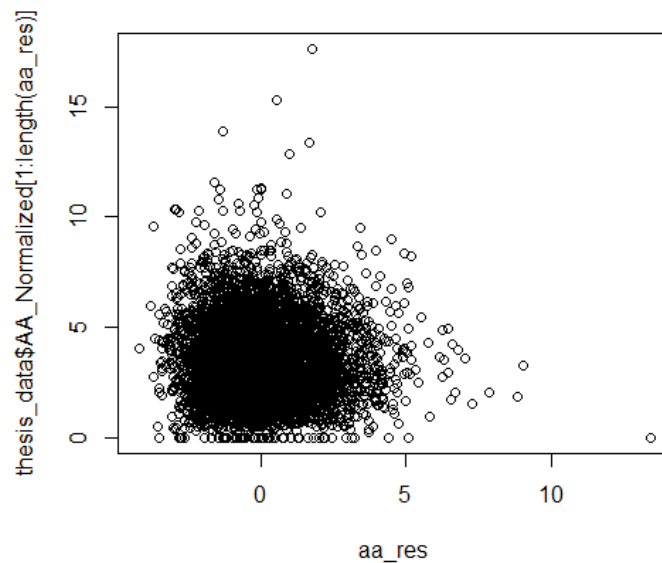
Linearity of Attributive Adjective Model

Figure 2
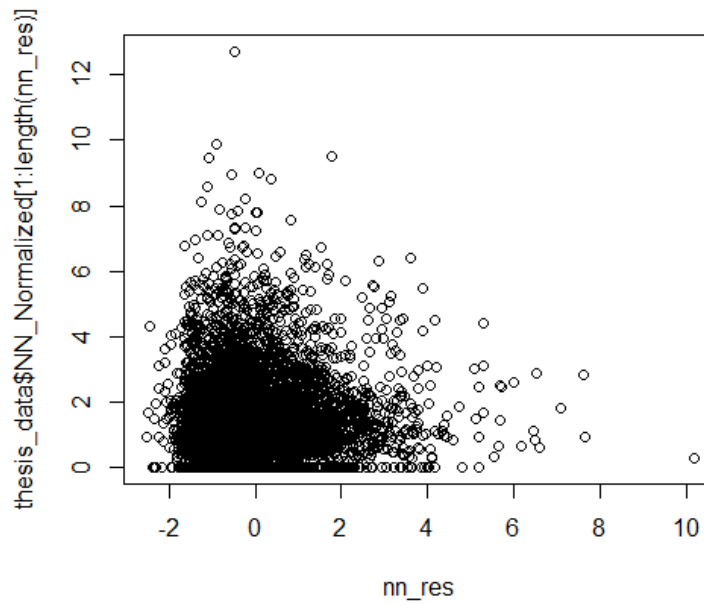
Linearity of Noun-Noun Phrase Model

Figure 3

Linearity of Nominalization Model
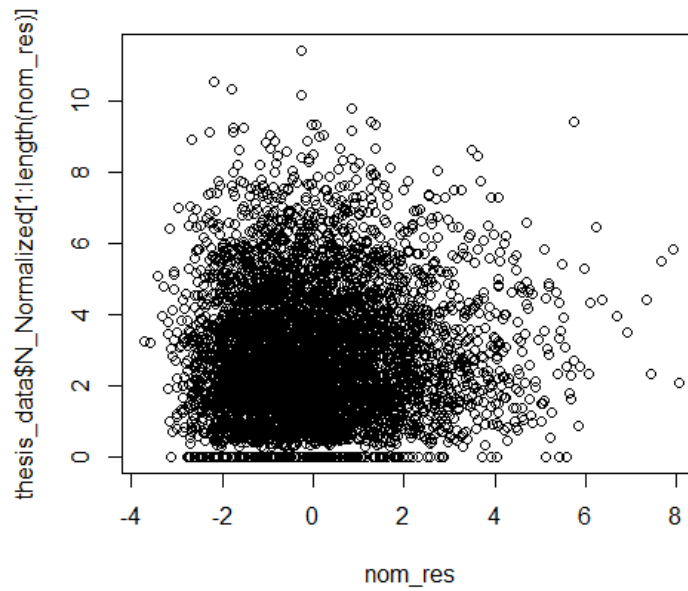
As can be seen, each model conforms to the linearity assumption.

The assumption of homogeneity of variance is tested using a fitted versus residual plot. If the points are distributed approximately evenly across y = 0, the assumption is met.

Figure 4

Homogeneity of Attributive Adjective Model
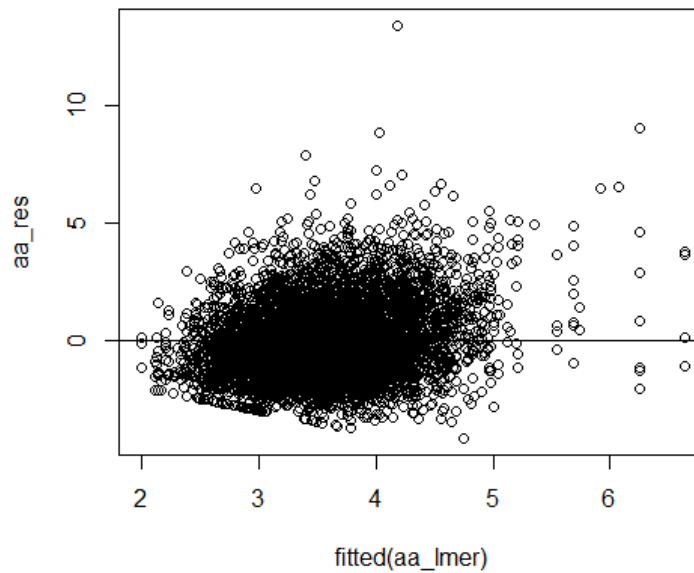
Figure 5

Homogeneity of Noun-Noun Phrase Model
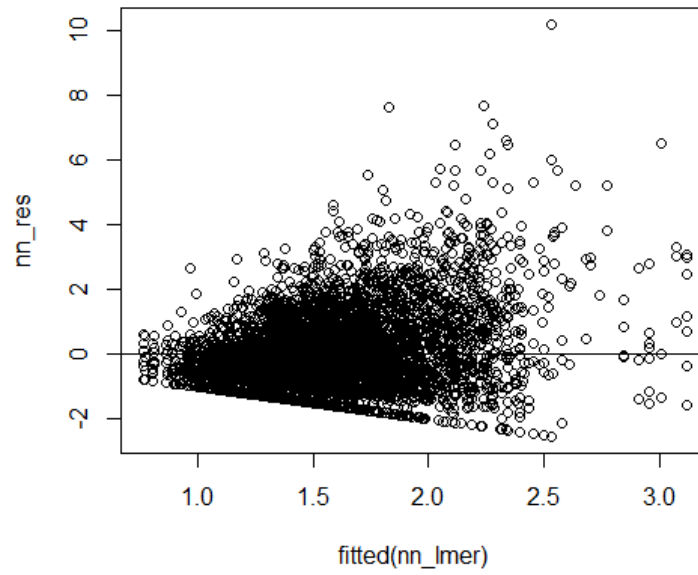
Figure 6

Homogeneity of Nominalization Model
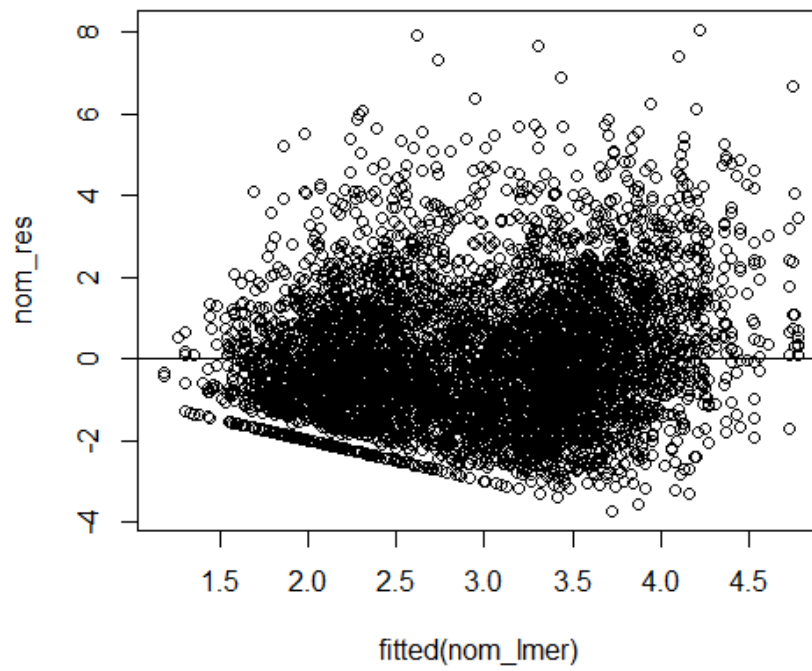
Again, all models meet the requirements of homogeneity of variance.

The final assumption of normality of error is tested by using QQ plots to view how residuals match up to normal quantiles. The straighter the line is, the greater the normality. Extreme convex or concave lines are indicative of a lack of normality.

Figure 7

Normality of Attributive Adjective Model

Figure 8

Normality of Noun-Noun Phrase Model

Figure 9

Normality of Nominalization Model

All three measures have concave curvatures, with noun-noun phrases having the most

obvious curve. These curves are significant enough to suggest that the models—particularly that

of noun-noun phrases—may not follow the third assumption. Logarithmic transformations did not improve the normality, either. This questionable state of the models indicates that the results and subsequent discussion may be unreliable.

Results

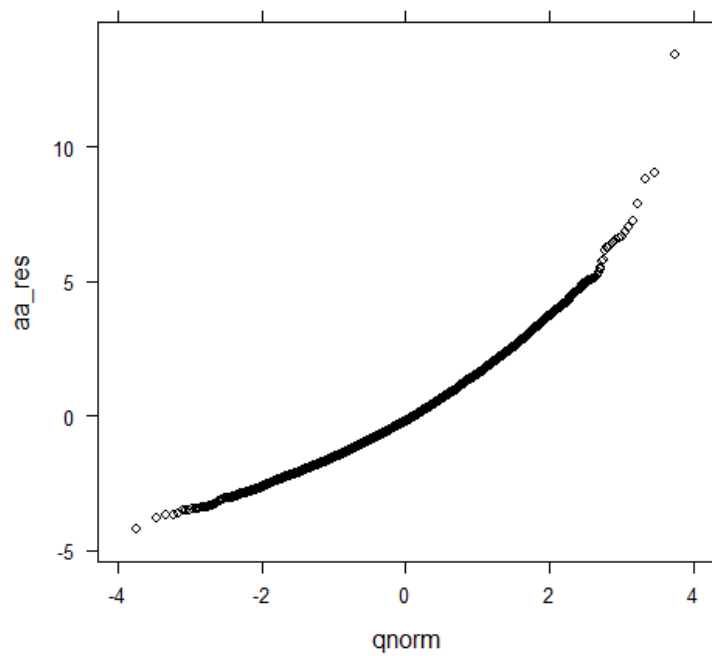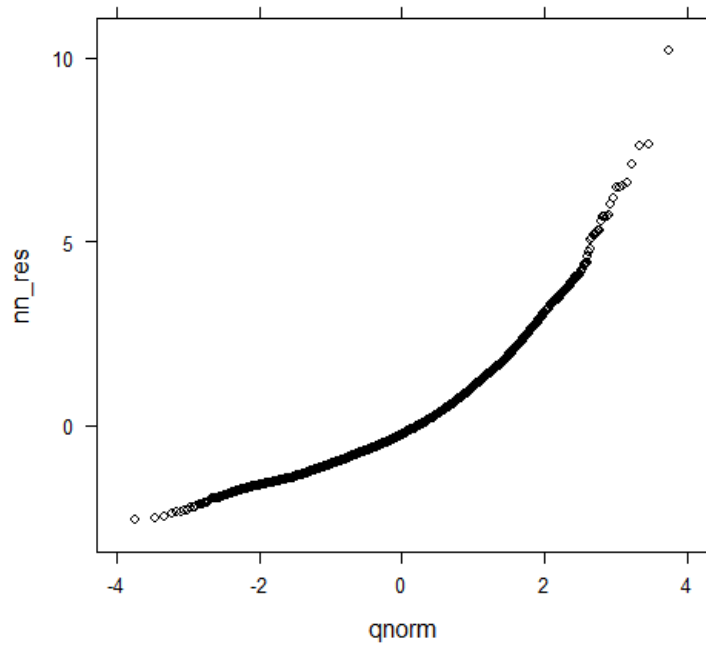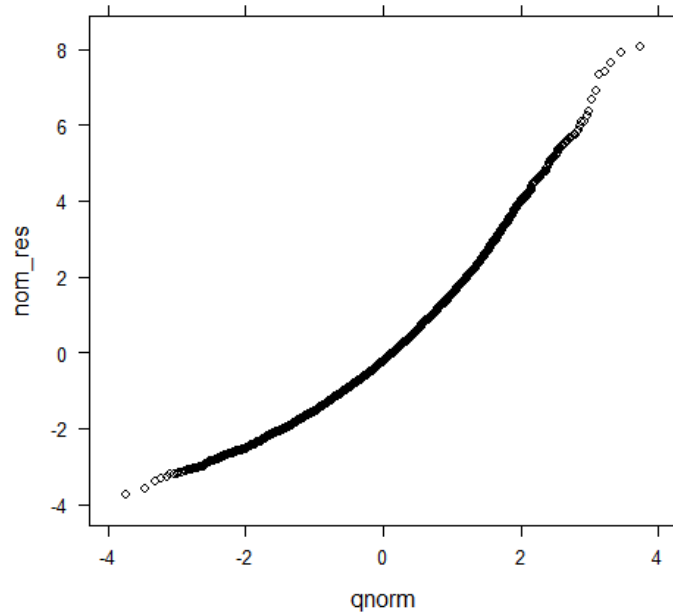The purpose of this study was to examine the effects of time and proficiency on phrasal complexity. First, we looked at the main effect of time, then the main effect of proficiency. Again, there was no interaction between time and proficiency, so that will not be discussed.

Effects of Time on Phrasal Complexity

There were main effects for all three variables measured. Additionally, each variable showed a positive correlation to time. That is, as allotted writing time increased, so did the frequency of each phrasal measure in learners' writing.

Attributive Adjectives

For attributive adjectives, normalized counts increased by more than 17% on average between 10- and 30-minute responses (see Figure 10). The increase is statistically significant and has an effect size of 0.37 (see Table 2), indicating that allotted writing time has a small relative effect on the number of attributive adjectives per 100 words in student responses.

Table 2

Time and Attributive Adjectives

	Mean	SD	Estimate	SE	df	t	*p*	Cohen's d
10-Minute	3.27	1.88	0.5666	0.04562	4526	12.418	< .001	0.37
30-Minute	3.84	1.74						

Figure 10

Time and Attributive Adjectives
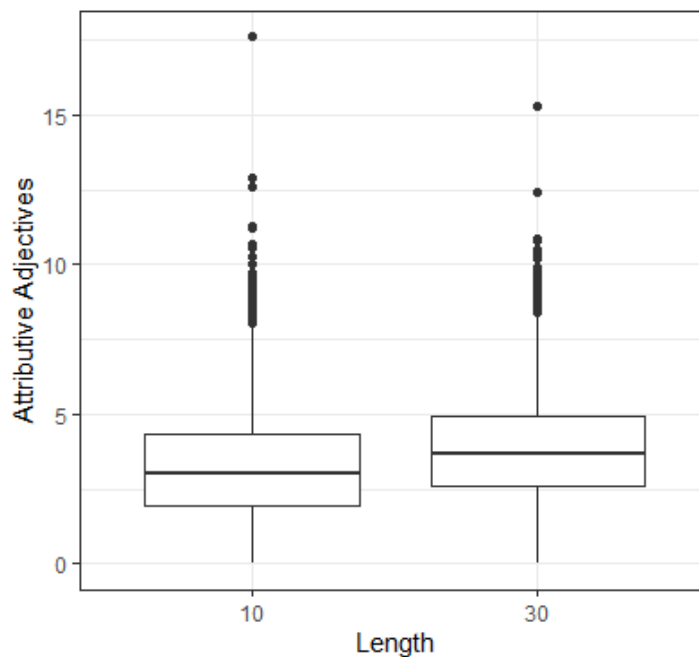
Noun-Noun Phrases

Means show that noun-noun phrases decreased when shifting from 10- to 30-minute responses, with more than a 9% decrease on average (see Figure 11). The effect of time on noun-noun phrases per 100 words is significant, but with an effect size of -0.14 (see Table 3). This suggests that the actual effect of this inverse relationship is negligible.

Table 3

Time and Noun-Noun Phrases

	Mean	SD	Estimate	SE	df	t	*p*	Cohen's d
10-Minute	1.66	1.52	-0.1625	0.03357	4520	-4.841	< .001	-0.14
30-Minute	1.51	1.15						

Figure 11

Time and Noun-Noun Phrases
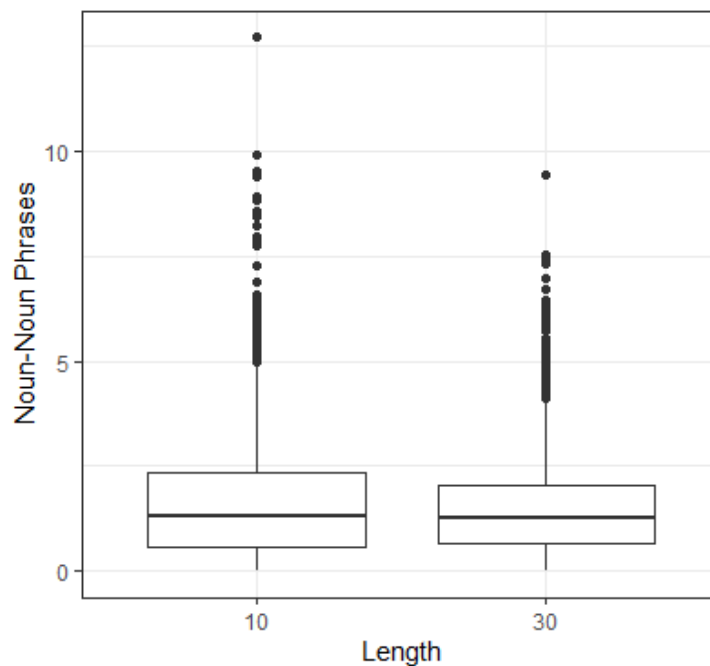
Nominalizations

      As with attributive adjectives, the frequency of nominalizations per 100 words increased significantly as allotted time increases. However, this increase was much larger for nominalizations, with an impressive jump of nearly 53% between 10- and 30-minute responses (see Figure 12). Additionally, the effect size of 0.80 is the largest in the study, and suggests that time has a similarly large relative effect on normalized nominalizations in pre-matriculated L2 writing (see Table 4).

Table 4

Time and Nominalizations

	Mean	SD	Estimate	SE	df	t	*p*	Cohen's d
10-Minute	2.31	1.61	1.218	0.0450	4623	27.05	< .001	0.80
30-Minute	3.53	1.88						

Figure 12

Time and Nominalizations
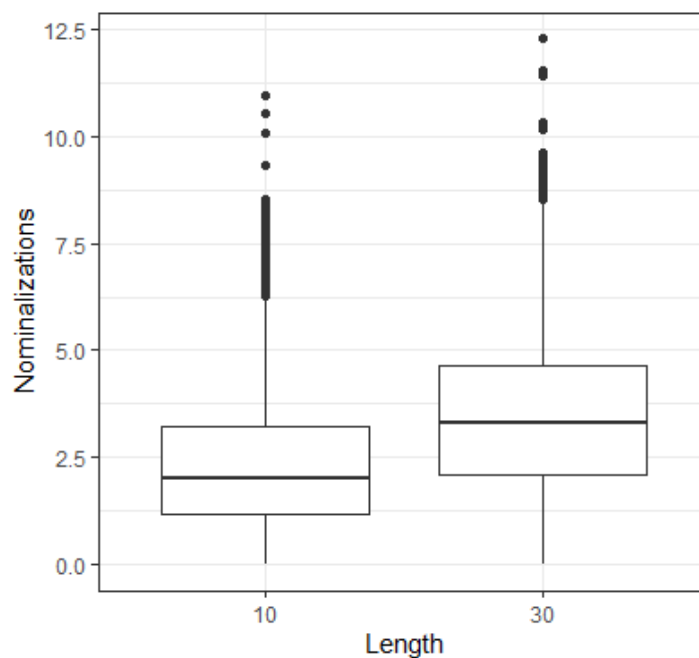
Effects of Proficiency on Phrasal Complexity

Again, there were main effects for all three variables measured, and again, both attributive adjectives and nominalizations positively correlated with proficiency, while noun-noun phrases negatively correlated with proficiency. For each variable, all levels showed significant differences from one another, with the sole exception of Intermediate by Advanced for noun-noun phrases.

Attributive Adjectives

For attributive adjectives, the overall model is significant ($p < .001$; see Table 5). There was a roughly 11% average increase in their use between Novice and Intermediate proficiencies with a Bonferroni-adjusted p-value of $< .001$, and just above a 4% increase between Intermediate and Advanced with an adjusted p-value of .0425 (see Figure 13). The effect size of 0.32 is similar to the effect size of time, again indicating a small relative effect.

Table 5

Proficiency and Attributive Adjectives

	Mean	SD	Estimate	SE	df	t	*p*	Cohen's d
Novice	3.22	1.88	0.1672	0.03206	1070	5.215	<.001	0.32
Intermediate	3.58	1.84						
Advanced	3.73	1.79						

Figure 13

Proficiency and Attributive Adjectives
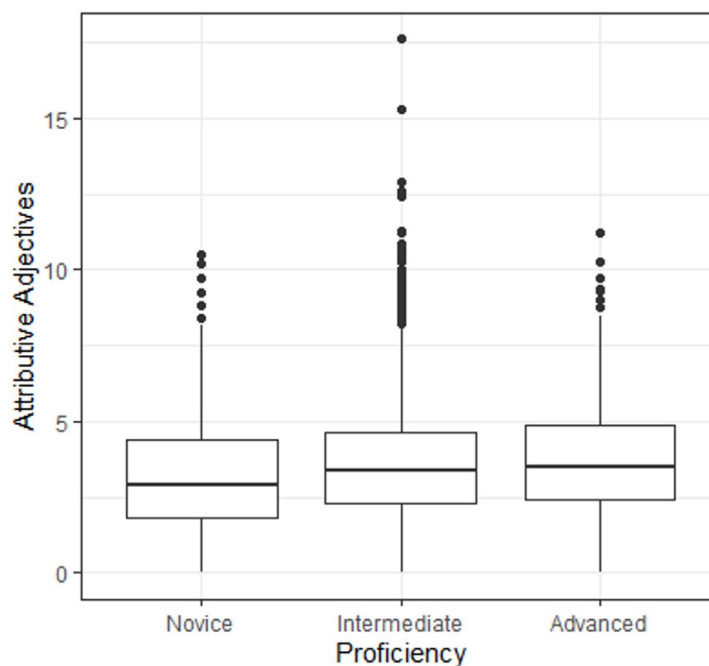
Noun-Noun Phrases

The noun-noun phrase model also shows significance (*p* = .022; see Table 6). There was an average decrease of roughly 12.5% from Novice to Intermediate with an adjusted *p*-value of 0.0014 (see Figure 14). The difference between Intermediate and Advanced trended downwards, but is not statistically significant, even before Bonferroni adjustment (*p* = .478). Like with time, the effect size of -0.14 suggests that the actual effect of proficiency on the normalized frequency of noun-noun phrases is negligible.

Table 6

Proficiency and Noun-Noun Phrases

	Mean	SD	Estimate	SE	df	t	*p*	Cohen's d
Novice	1.78	1.56	-0.0531	0.02317	1071	-2.289	.022	-0.18
Intermediate	1.56	1.32						
Advanced	1.53	1.26						

Figure 14

Proficiency and Noun-Noun Phrases
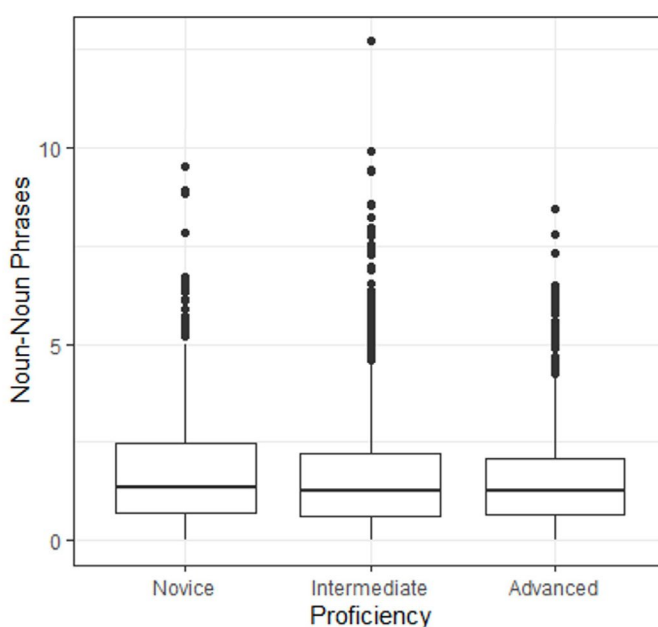
Nominalizations

Once again mirroring attributive adjectives, the frequency of nominalizations per 100 words increased significantly with proficiency across all levels (see Figure 15). There was a roughly 10% average increase across each proficiency level. Novice by Intermediate gave an adjusted *p*-value of .01, and Intermediate by Advanced one of < .001. This, combined with a medium effect size of 0.43 (see Table 7), supports the idea that higher proficiency learners were more likely to use nominalizations in their writing.

Table 7

Proficiency and Nominalizations

	Mean	SD	Estimate	SE	df	t	*p*	Cohen's d
Novice	2.65	1.91	0.2102	0.0291	1144	7.23	< .001	0.43
Intermediate	2.91	1.85						
Advanced	3.20	1.83						

Figure 15

Proficiency and Nominalizations
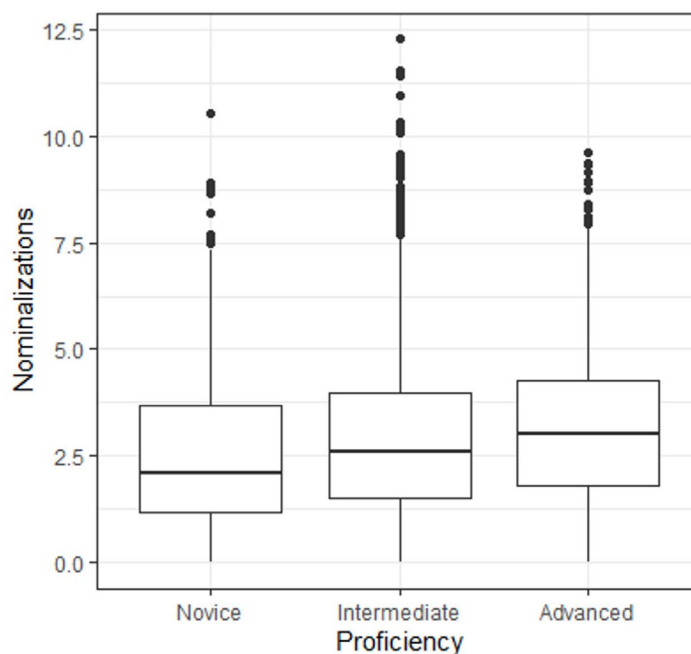
Random Effects

The random effect of participant consistently showed the largest variance of all random effects (see Table 8). L1 had low variances across all models, reaching a maximum variance of 6% with nominalizations. Age showed little-to-no variance for both noun-noun phrases and nominalizations, but did have a 9% variance for attributive adjectives.

Table 8

Random Effect Variances

	Attributive Adjectives	Noun-Noun Phrases	Nominalizations
Participant	0.40	0.20	0.25

Age	0.09	0.02	--
L1	0.02	0.03	0.06

Discussion

In this study our objective was to examine the effect of time and the effect of proficiency on the occurrence of three phrasal measures in the writing of pre-matriculated L2 English learners. These measures—attributive adjectives, noun-noun phrases, and nominalizations—were analyzed in a corpus of 10- and 30-minute student responses taken from an IEP. Our expectations were that all three measures would increase in normalized frequency as both time and proficiency increased.

Random Effects

The three random effects of participant, age, and L1 differed greatly in variance and, therefore, in how they affected the data. Participant having the largest variance is expected. Individual stylistic preferences and writing ability are obvious and impactful factors in language production. If a writer was exposed to more of a certain structure throughout their language acquisition, it is logical to assume that they might produce proportionally more of said structure in their speech and writing. Interestingly, age had the greatest effect on the use of attributive adjectives, with learners ages 35 and older averaging 1.84 attributive adjectives per 100 words and learners ages 18-22 averaging 3.38, suggesting that younger learners are nearly twice as likely to use adjectives. This finding could perhaps be explained by a cultural or generational trend, but that is beyond the scope of this study.

L1 having the least variance may be surprising to some, considering the breadth of L1s in the data and the differences in grammar across each language. That said, some interesting patterns can be noted among nominalizations. Spanish and Portuguese students used an average of 8.06 and 8.11 nominalizations per 100 words, respectively. Japanese learners had an average

of 7.58 nominalizations per 100 words, a 6% drop, while Chinese learners had an average of 8.68, an increase of more than 7%. In short, students with different L1s did vary slightly in usage of each measure, most notably with nominalizations. Overall, both age and L1 do see to have an effect on the usage of phrasal structures, but not to as extreme an extent as individual differences.

Time

The first research question examined the effect of time on phrasal complexity. The present study showed that the amount of allotted writing time was likely to have an effect on the phrasal complexity of student writing, varying from small to large depending on the measure. Specifically, timing had a small effect on attributive adjectives and noun-noun phrases, and a large effect on nominalizations. Both attributive adjectives and nominalizations increased in normalized frequency as timing increased, while noun-noun phrases decreased in frequency.

Hartshorn et al. (2010) used 10-minute responses as their data source and only considered mean length of T-units and neglected phrasal measures completely, but noted that the mean length increased significantly after their experimental treatment. Gray et al. (2019) found that certain noun-noun phrases and attributive adjectives positively correlated with scores on the 30-minute Independent Writing Task of the TOEFL, with nominalizations following a similar— though insignificant—pattern. Fathi and Rahimi (2022) failed to find significant changes in complexity across different proficiencies while using 50-minute essays as their data sources and mean length of T-units as their single complexity measure.

Hartshorn et al.'s research suggests that, in a 10-minute environment, students use complexity to a lesser degree without intervention, but with intervention have room to improve their complexity. In some features, specifically nominalizations, Gray et al. (2019) found that there was greater complexity in 30-minute writing compared to 20-minute writing. This comes with some caveats. The two timings also involved different genres, which are expected to show

differences in complexity regardless of timing. Additionally, noun-noun phrases were less frequent in the longer task.

Our findings align well with Gray et al., as nominalizations increased in normalized frequency between the 10- and 30-minute responses while noun-noun phrases decreased. Only six 10-minute responses reported 9 or more nominalizations per 100 words, with 26 of the 30-minute responses meeting this criterion—a 333% increase. In those six 10-minute responses, the most common nominalizations were 'activity,' 'university,' 'teacher,' and 'building,' with each being repeated multiple times throughout most responses, accounting for more than 65% of their total nominalizations. This can reasonably be attributed to the fact that the prompt asks students to describe a school they attended previously. None of these nominalizations are marked as especially high register or obscure words. In the six 30-minute responses with the highest frequency of nominalizations, we see a much greater variety in word choice and a generally higher register than the 10-minute responses. Some examples include 'environment,' 'violence,' 'opportunity,' 'behavior,' 'necessity,' 'rehabilitation,' 'instability,' and 'tranquility.' No single nominalization accounted for more than 10% of total nominalizations in these six responses. This notable contrast highlights the difference in writing quality reported here and by Gray et al.

While Fathi and Rahimi (2022) failed to find significance using a single clausal complexity measure, our study suggests that the pattern of increased complexity noted in Gray et al.'s research would continue, with even more phrasal complexity found in 50-minute writing than in 30-minute responses. Overall, our findings conform with past research involving timed writing, suggesting that forms such as nominalizations—and, to a lesser degree, attributive adjectives—are more linguistically complex, requiring additional time and thought to implement effectively, whereas structures like noun-noun phrases are simpler and less time-intensive.

Proficiency

The second research question examined the effect of proficiency on phrasal complexity. The results for both attributive adjectives and nominalizations fall in line with the pattern related by Biber et al. (2011) and Staples and Reppen (2016)—that is, that the frequency of phrasal complexity features in writing is directly related to overall writing proficiency. Nominalizations in particular showed a strong correlation with proficiency, demonstrating increased linguistic complexity and positive responses from raters in a testing environment.

Similar to timing, noun-noun phrases showed a negative correlation to proficiency. This could be due to a variety of factors, but one in particular draws attention. Noun-noun phrases in our data proved to be quite polarizing in terms of accuracy, register, and reoccurrence across proficiencies. For instance, one Novice response used the phrase 'my home town' 15 times in 210 words (not including two instances of 'my hometown'). Other noun-noun phrases in this entry include 'nature load' and 'home town sight.' The constant repetition of this phrase and the use of incorrect word forms (e.g., 'nature' instead of 'natural') are indicative of limited vocabulary and grammatical skills. Examining an Advanced response, only five noun-noun phrases were used in 367 words, including 'human being' and the erroneous 'self steam' that was likely meant to be 'self-esteem.' Furthermore, 45 nominalizations were used, including 'implementation,' 'necessities,' 'percentage,' and 'tranquility.' Each nominalization used was fairly high register, used accurately, and was not repeated more than three times, if at all, which corresponds with the higher proficiency of the writer.

While a phrase like 'complexity measures' may be similarly valued as high-level or academic in nature due to its rarity and the relative difficulty of usage within a timed written response, the highly specific context (i.e., the field of linguistics) makes it unlikely to be acquired by students of an IEP preparing for general college studies. Therefore, learners may be more likely to acquire 'simple,' relatively informal noun-noun phrases early in development,

and, while the acquisition of more complex noun-noun phrases may be unlikely in a general-purpose IEP, more advanced L2 writers may recognize the lower 'value' of the previously acquired phrases as they continue to increase in proficiency. This would naturally lead to noun-noun phrases being used less often until learners encounter situations where more complex phrases are needed and subsequently acquired.

In summary, there is evidence that phrasal complexity can be used as an indicator of proficiency among L2 learners' writing, with higher frequencies of phrasal structures indicating greater writing proficiency. However, this pattern seems to depend upon the accuracy with which learners implement phrasally complex structures into their writing. Furthermore, repetition of a small group of phrasal structures appears to have the inverse effect, suggesting that variety in phrasal complexity is an additional factor that requires assessing.

Conclusion

This paper evaluated phrasal complexity in the writing of pre-matriculated L2 English learners, and how said complexity was affected by time and proficiency. We found that both attributive adjectives and nominalizations increase with both effects, while noun-noun phrases decrease. While this pattern aligns with findings from past studies, there are some shortcomings of this study to consider and areas that would benefit from further research.

Limitations and Future Research

Foremost among these limitations is the inability to measure the interaction between time and proficiency. The proficiency data provided by the IEP was limited to a single score per learner, meaning that even if a learner had eight entries across four semesters, we still only had one score for that learner, the one given to the latest response. This led to our proficiency data being less accurate than ideal, and prevented any fixed effect interaction analyses. Future studies should aim to use data with an individual score for each response. Additionally, the data was

gathered from a single IEP based in Utah. Including data from other locations and programs would increase the reliability of the findings, allowing researchers and instructors to generalize the findings to a wider range of ESL/EFL writers.

Furthermore, due to resource and time limitations, only three features were considered in this study. In future studies, analyzing additional phrasal complexity features—such as prepositional phrases and possessive of-phrases as noun postmodifiers—would help further confirm which specific phrasal features are most closely associated with academic writing and which features should have a greater or lesser priority in ESL instruction.

Finally, further research is required on whether accuracy and the repetition of phrasal structures may affect whether phrasal complexity functions as an accurate indicator of proficiency. Our findings suggest that they could, but neither idea was specifically tested in this study.

Pedagogical Applications

The results of this study can be readily applied to both classrooms and assessment creation. Regarding timing, where the more 'academic' phrasal features tend to decrease in frequency as allotted time decreases, teachers may consider adjusting their rubrics to account for this. If a relatively short time is allowed for a certain task, students cannot reasonably be expected to produce numerous examples of more complex grammatical structures, nor to do so accurately. Furthermore, such short responses should not be used in summative assessments as a means of measuring proficiency in said structures. Rather, they might be more effective in evaluating simpler structures or general writing fluency.

The results for each individual feature are also informative. As noun-noun phrases displayed inverse relationships with both time and proficiency and were often used erroneously by students, teachers might consider either reducing the number of noun-noun phrases taught to

beginning levels or more regularly reviewing word forms, parts of speech, and correct usage of the noun-noun phrases used in class. With nominalizations having the greatest effect size for both main effects, students may benefit from additional direct vocabulary and usage instruction on this topic. Such instruction can help pre-matriculated students learn to better 'compress' information into complex phrasal structures typical of university-level writing.

Should an ESL/EFL writing curriculum not include explicit instruction on these phrasal structures, results from this and other studies suggest its addition will better prepare ELLs to produce quality writing as they pursue higher education in English-speaking countries. And, as more phrasal measures are evaluated in future research, curricula should be regularly updated with those that are shown to be indicative of academic writing.

References

Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Dıaz-Negrillo, N. Ballier, & P. Thompson Eds., Automatic treatment and analysis of learner corpus data, 249–264. Amsterdam, the Netherlands: John Benjamins.

Atak, N. & Saricaoglu, A. (2021). Syntactic complexity in L2 learners' argumentative writing: Developmental stages and the within-genre topic effect. Assessing Writing, 47.

Biber, D. (1988). *Variation across speech and writing*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

Biber, D. & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*(1), 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

Biber, D., & Gray, B. (2013). Discourse characteristics of writing and speaking task types on the TOEFL ibt® test: a lexico-grammatical analysis. ETS Research Report Series, 2013(1), i-128.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5-35.

Biber, D., Gray, B., & Poonpon, K. (2013). Pay attention to the phrasal structures: Going beyond T-Units—A response to WeiWei Yang. *TESOL Quarterly*, *47*(1), 192–201.

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, *37*(5), 639-668.

Biber, D., Gray, B., & Staples, S. (2016). Contrasting the grammatical complexities of conversation and academic writing: Implications for EAP writing development and teaching. *Language in Focus*, *2*(1), 1-18.

Casal, J. & Lee, J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. Journal of Second Language Writing, 44, 51-62.

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing, 26*, 66-79.

Durán, P., Malvern, D., Richards, B. & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics, 25*(2), 220-242.

Fathi, J. & Rahimi, M. (2022). Examining the impact of flipped classroom on writing complexity, accuracy, and fluency: a case of EFL students. *Computer Assisted Language Learning, 35*(7), 1668-1706.

Gray, B., Geluso, J., & Nguyen, P. (2019). The longitudinal development of grammatical complexity at the phrasal and clausal levels in spoken and written responses to the TOEFL iBT® test. *ETS Research Report Series*, *2019*(1) 1-51.

Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly, 44*(1), 84–109.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features.* Mahwah, NJ: Lawrence Erlbaum.

Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L2 and L1 academic texts. *TESOL Quarterly, 37*, 275–301.

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers in Linguistics, 53*, 61-79.

Kim, M. (2021). Exploring longitudinal changes in lexical and syntactic features in beginning-level EFL learner writing. System, 103.

Kitajroonchai, N., Kitjaroonchai, T., & Sanitchai, P. (2022). The Effects of Process Genre-Based Writing and Process Writing Approaches on Asian EFL Pre-University Students' Writing Performance. *Journal of Language Teaching and Research, 13*(4), 860-871. https://doi.org/10.17507/jltr.1304.19

Li, Y., Nikitina, L., & Riget, P. (2022). Development of syntactic complexity in Chinese university students' L2 argumentative writing. Journal of English for Academic Purposes, 56.

Lu, X. & 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

Martinez, A. (2017). Syntactic complexity in secondary-level English writing: Differences among writers enrolled on bilingual and non-bilingual programmes. Porta Linguarum, 28.

Martinez, A. (2018). Analysis of syntactic complexity in secondary education EFL writers at different proficiency levels. Assessing Writing, 35, 1-11.

Olinghouse, N. G. & Wilson. J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing, 26*, 45–65.

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.

Parkinson, J. & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, *14*, 48-59.

Picoral, A., Goulary, L. & Fullmer, M. (2018-). Biber Tag Checking Tool. *GitHub Repository*. https://github.com/writecrow/coding_tools

Qin, W. & Zhang, X. (2022). Do EFL learners use different grammatical complexity features in writing across registers? Reading and Writing.

Saricaoglu, A. & Atak, N. (2022). Syntactic complexity and lexical complexity in argumentative writing: Variation by proficiency. Novitas-ROYAL, 16(1), 56-73.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: ESL research and its implications. *TESOL Quarterly, 27*, 657–677. https://doi.org/10.2307/3587400

Skehan, P. (1998). A cognitive approach to language learning. *Oxford University Press.*

Staples, S. & Dilger, B. (2018-). *Corpus and repository of writing [Learner corpus articulated with repository]*. Available at https://crow.corporaproject.org

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, *33*(2), 149–183. https://doi.org/10.1177/0741088316631527

Staples, S. & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing, 32*, 17–35. https://doi.org/10.1016/j.jslw.2016.02.002

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.

Vyatkina, N., Hirschmann, H., & Golcher, F. (2015). Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study. Journal of Second Language Writing, 29, 28-50.

Wolfe-Quintero K., Inagaki, S., & Kim, H. (1998). Second language development in writing: Measures of fluency, accuracy, and complexity. *Studies in Second Language Acquisition, 23*(3), 423-425.

Zhang, L. & Cheng, X. (2021). Examining the effects of comprehensive written correct feedback on L2 EAP students' linguistic performance: A mixed-methods study. Journal of English for Academic Purposes, 54.