

Nominal Modifiers in Argumentative Essays as Discriminators for Writing Course Placement Decisions

Hyunwoo Kim*

Kim, Hyunwoo. (2020). Nominal modifiers in argumentative essays as discriminators for writing course placement decisions. *English Teaching*, 75(3), 3-24.

Grammatical complexity constitutes an indispensable subconstruct of L2 writing ability. Nonetheless, rating scale descriptors for grammatical complexity have heavily relied on degree modifiers or adjectives. Thus, this study attempts to explore the potential for the use of nominal modifiers as discriminators for adjacent levels of L2 writers in the context of an English Placement Test. This study analyzed 374 argumentative essays written by international undergraduates. Seven nominal modifiers in the developmental stages of grammatical complexity were examined and a cumulative ordinal logistic regression model with proportional odds was fitted to explore the relative effects of those grammar features on placement decisions. Four nominal modifiers were found to be positively associated with placement decisions, after adjusting for the effects of other variables. One educational implication is for the use of nominal modifiers to be incorporated into rating scale descriptors as discriminators for L2 writers who are not advanced enough to be exempted from an English Placement Test.

Key words: nominal modifiers, grammar developmental stages, ordinal logistic regression

*Author: Hyunwoo Kim, Lecturer, Department of English Language Education, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, 08826, Korea; Email: kim01@snu.ac.kr

Received 30 June 2020; Reviewed 4 August 2020; Accepted 5 September 2020

1. INTRODUCTION

Grammatical competence of L2 writers has been found to play a pivotal role in affecting raters' judgments of both spoken and written language in performance tests (Alderson & Huhta, 2005; McNamara, 1990). Nonetheless, rating scale descriptors for grammatical competence in L2 writing assessment have not been specific enough to capture the developmental aspects of grammatical complexity. To illustrate, the explicit relationship between rating scale descriptors and test takers' performance has been found to be often indiscernible (Carr, 2011; North & Schneider, 1998); the adjacent scale descriptors tend to differentiate from one other, mainly relying upon adjectives, such as *simple* and *complex*, without clearly defining what precise differences exist between the descriptors concerning language features in written texts. For example, the following rating scale descriptors for grammatical complexity in the widely implemented analytic rating scales in the context of ESL do not appear to explicitly define simple and complex constructions (Jacobs, Zingraf, Wormuth, Hartfiel, & Hughey, 1981).

EXCELLENT TO VERY GOOD: effective complex constructions

GOOD TO AVERAGE: effective but simple constructions

FAIR TO POOR: major problems in simple/complex constructions

VERY POOR: virtually no mastery of sentence construction rules

Ambiguous rating scale descriptors could pose a threat to the validity of ratings as they often lead to undesirable rating behaviors, such as the halo effect, regardless of the amount of rater training (Lumley, 2002). In order to address ambiguous rating scale descriptors, there have been fruitful attempts to empirically derive rating scale descriptors (Fulcher, 1996; Knoch, 2009; Turner & Upshur, 2002). However, few studies have been done to fully explore the potential for the use of salient grammar features exhibited by L2 writers as rating scale descriptors to differentiate between advanced and less advanced L2 writers.

Regarding complex grammar features exhibited by L2 writers, Biber, Gray, and Poonpon (2011) suggested the developmental stages of grammatical complexity, where advanced L2 writers are assumed to produce more complex noun phrases. Later empirical studies further supported their notion, as the collective use of nominal modifiers was found to be a reliable yardstick for distinguishing between advanced L2 writers and less advanced ones (Ansarifard, Shahriari, & Pishghadam, 2018; Kyle & Crossley, 2018; Lan, Lucas, & Sun, 2019; Musgrave & Parkinson, 2014; Taguchi, Crawford, & Wetzel, 2013).

Despite a body of research on the validity of the developmental stages of grammatical complexity, few studies have been conducted to explore the extent to which the developmental stages of grammatical complexity (Biber et al., 2011) hold true for L2

writing in an English Placement Test (hereinafter, EPT). Additionally, few studies have implemented a cumulative odds logistic regression to examine the validity of the developmental stages of grammatical complexity.

Thus, this study attempts to investigate whether nominal modifiers selected from the developmental stages are effective in differentiating between adjacent remedial academic writing course placement decisions (hereinafter, placement decisions) on essays written by L2 writers in the context of an EPT. This study also aims to examine the extent to which the relative strength of the effects of nominal modifiers on placement decisions aligns with their corresponding developmental stages of grammatical complexity (Biber et al., 2011).

2. REVIEW OF THE LITERATURE

2.1. Grammatical Complexity in L2 Writing

By definition, grammatical complexity refers to “the range of forms that surface in language production and the degree of sophistication of such form” (Ortega, 2003, p. 492). The degree of elaboration and sophistication of language depends on both learners’ willingness and readiness to try more challenging forms of expression and a wide variety of grammar features. In light of this aspect of grammatical complexity, it is assumed that advanced L2 learners can quickly produce “a wide variety of both basic and sophisticated structures” (Wolfe-Quintero, Inagaki, & Kim, 1998, p. 69).

By investigating the distributional patterns for 28 grammatical features across spoken and written corpora of L1 learners (i.e., face-to-face conversation and academic writing), Biber et al. (2011) demonstrated that “the [grammatical] complexity of conversation is clausal, whereas the [grammatical] complexity of academic writing is phrasal” (p. 22). Based on the results of the statistical comparisons between two corpora, phrasal nominal modifiers (e.g., prepositional phrases as nominal postmodifiers, attributive adjectives as nominal premodifiers) were found to be statistically salient features of academic writing. Hypothesizing that these grammar features of academic writing are acquired later than those of conversation, the authors called for empirical studies of L2 learners to identify if the grammatical developmental stages for L1 learners hold valid for L2 learners.

A body of empirical research has been conducted to examine the extent to which the developmental stages of grammatical complexity proposed by Biber et al. (2011) hold true. Directly motivated by their notion, Parkinson and Musgrave (2014) examined essays of L2 graduate writers to identify the degree of utilization of nominal modifiers. The researchers found that less advanced L2 graduate writers favored adjectives as nominal premodifiers than advanced L2 graduate writers, and that the statistically significant difference between

advanced and less advanced L2 writers was only found in the use of prepositional phrases as nominal postmodifiers with abstract meanings. Thus, the researchers concluded that “the complexity of the noun group will increase as proficiency level of academic writers increases” (p. 51) in research articles, partially validating the developmental stages of grammatical complexity proposed by Biber et al. (2011). However, Parkinson and Musgrave’s study is limited in that two groups of L2 graduate writers were engaged in two different genres of academic writing; the less advanced group worked on argumentative essays, while the more advanced group completed literature reviews and discussions. Furthermore, the participants in both groups were able to receive assistance when revising their works. Due to these confounding variables, the findings of Parkinson and Musgrave’s study may not be generalized to other L2 writers in a different context.

Following the same research trajectory, Taguchi, Crawford, and Wetzel (2013) analyzed argumentative essays written by L2 writers in a timed placement test, a circumstance which eliminated the possibility of receiving assistance in the revision. Additionally, all of the test takers were engaged in producing an argumentative essay from the same prompt, where they were asked to argue for one view and against the other. As a result, the effects of the genre on grammatical complexity were minimized in Taguchi et al.’s study. Then, whether the placement decision was related to grammar features in the essays was examined. Among six clausal-level and nine phrasal-level grammar features studied, the most remarkable difference between high-rated essays and low-rated essays was found in the use of adjectives as nominal premodifiers, prepositional phrases as nominal postmodifier and verb *that* complements, which occurred more frequently in the high-rated essays. Despite the unavailability of inferential statistics in their study, Taguchi et al.’s findings appeared to be consistent with previous studies where the use of adjectives and prepositional phrases as nominal modifiers was positively correlated with L2 writing proficiency (Grant & Ginther, 2000).

It is important to note that regarding a large-scale standardized test (e.g., TOEFL iBT™), a group of co-occurring grammar features was a “more robust [predictor] of task types and score-level differences” (Biber, Gray, & Staples, 2016, p. 639) than individual grammar features in the developmental stages. More specifically, Dimension 1 (e.g., nominal modifiers, pronouns), one of the co-occurring grammar patterns derived from factor analysis of grammar features, was a significant predictor of human ratings.

Lan et al. (2019) also attempted to examine the extent to which the number of nominal modifiers utilized by first-year Chinese undergraduates was related to writing proficiency levels. The frequencies of 11 nominal modifiers in the essays were found to be positively correlated with writing proficiency levels. More specifically, the frequencies of four grammar features, including premodifying adjectives, premodifying nouns, relative clauses, and prepositional phrases *of* as nominal postmodifiers, exhibited the most pronounced

differences between two proficiency levels. Surprisingly, the advanced features of nominal modifiers in the developmental stages of grammatical complexity were not found to show marked differences between proficiency levels, presumably due to the small sample size in the study, and the scarcity of such advanced features in the essays written by L2 writers, an observation that was also made in another study (Kyle & Crossley, 2018).

Lu (2011) explored the usefulness of 14 grammatical complexity measures as L2 developmental indices for Chinese ESL writers. It was found that the proficiency levels, the genre of the essays, and timing conditions of the assessment exercised significant influence on most of the syntactic complexity measures. After the genre of the essays and timing condition of the assessment were controlled, it was shown that the complex nominal per T-unit (CN/T) was able to successfully differentiate between L2 writing proficiency levels for Chinese ESL university students. Additionally, Bulté and Housen (2014) examined the effects of a short-term L2 academic course on the grammatical complexity of the written texts. Although there were no significant changes in the proportion of subordinated clauses and compound-complex sentences, the mean length of noun phrases was found to have increased significantly by the end of the course, and the mean length of noun phrases also turned out to be one of the more useful predictors for human ratings. Similarly, Crossley and McNamara (2014) investigated the effects of a semester-long academic writing course on increased syntactic complexity in learners' essays and the relationship between this growth and human ratings. The researchers found that, for a semester, learners produced longer noun phrases and more words before the main verb, which was indirectly indicative of the complexity of noun phrases.

Despite the differences in research designs and statistical analyses in previous studies, the findings of empirical studies on the developmental stages of grammatical complexity (Biber et al., 2011) have consistently indicated the usefulness of nominal modifiers as effective discriminators between written essays across L2 proficiency levels.

2.2. Effects of the Essay Length and Topic of an Essay

Along with the use of the grammar features, it is necessary to examine how the length of written essays is associated with raters' judgments of the overall quality of written texts. Traditionally, the length of an essay has been considered to be a promising predictor to distinguish advanced L2 writers from less advanced ones. It was previously found that more proficient L2 writers tended to produce more extended essays in a timed writing test (Larsen-Freeman, 1978, 2006). Similarly, Grant and Ginther (2000) showed that as L2 writers produced more extended essays with longer and more varied words, the holistic scores given to the written works tended to increase, findings that were congruent with previous results (e.g., Ferris, 1994). Thus, it is suggested that the effects of the length of an

essay on placement decisions are not negligible, indicating the necessity of statistically controlling them.

Previous research into the relationship between topics for argumentative essays and linguistic features elicited in the essays also demonstrated that the degree of both clausal- and phrasal-level grammatical complexity was associated with prompts for the essays. By conducting a multidimensional analysis of linguistic features, Weigle and Friginal (2015) demonstrated that topics of argumentative essays elicited different sets of linguistic features. Additionally, complex noun phrases per clause (CNP/C) were found to exhibit a statistically significant difference between two prompts for argumentative essays (Yang, Lu, & Weigle, 2015). Therefore, the effects of topics on placement decisions need to be statistically accounted for when examining the use of nominal modifiers.

The main objective of this study is to explore the effects of nominal modifiers selected from the developmental stages of grammatical complexity on placement decisions (Biber et al., 2011), after adjusting for the effects of the length and topic of an essay. The secondary research goal is to examine the extent to which the relative strength of the effects of those nominal modifiers on placement decisions aligns with their corresponding developmental stages of grammatical complexity. The research questions for this study are as follows:

1. To what extent is the number of nominal modifiers in argumentative essays associated with placement decisions for incoming international undergraduates after adjusting for the effects of the length and topic of an essay?
2. To what extent does the relative strength of effects of nominal modifiers in argumentative essays on placement decisions align with their corresponding developmental stages of grammatical complexity proposed by Biber et al. (2011) after adjusting for the effects of the length and topic of an essay?

3. METHODOLOGY

3.1. Learner Corpus

This study analyzed a learner corpus, which consisted of 374 timed argumentative essays as the independent writing task of an EPT for incoming international undergraduates at one university in the United States. In the EPT's independent writing task, the prompts under analysis mandated that international undergraduates write argumentative essays on four topics in a time span of 30 minutes. The tests were administered prior to the class start dates. Timed argumentative essays in the context of the

EPT were analyzed because test takers shared similar motivations and spent the same amount of time on the writing. It should be noted that only those international undergraduate students who were required to take the EPT constitute the study population of this study; highly advanced L2 writers were excluded from the study population due to the exemption policy at this institution.

Prior to the study, all the essays had been transcribed and annotated using the Biber Tagger by the developer of the learner corpus. All the essays were digitally stored as text files to build a learner corpus prior to analysis. As the developer of the learner corpus had deidentified personal information of the essays (e.g., test takers' L1s and educational backgrounds) before analysis, the 374 essays were considered existing data.

The results of the writing component of the EPT were used to determine whether test takers needed to take remedial academic writing courses. As shown in Table 1, if a test taker failed the test, the individual was placed into either the low-level academic writing course (ENGL 101B) or the intermediate-level academic writing course (ENGL 101C). If a test taker passed the test, the individual was not asked to take any remedial academic writing courses.

TABLE 1
Distribution of Essays across Topics and Placement Decisions

Placement Decision	Topic 1 (n = 139)	Topic 2 (n = 96)	Topic 3 (n = 87)	Topic 4 (n = 52)	Total (n = 374)
ENGL 101B (low)	46	13	7	17	83
ENGL 101C (intermediate)	46	39	20	22	127
PASS (upper-intermediate)	47	44	60	13	164
Total number of word	38,179	25,266	24,501	15,447	103,393

Note. Topic 1 = Modern conveniences, Topic 2 = Rules and regulations, Topic 3 = Preferences for e-books, Topic 4 = Violence on TV.

The essays were rated by two experienced ESL instructors, who were either master's or doctoral students in the university's Department of English at the time of test administration; raters were required to participate in rigorous onsite rating training sessions as part of the training. A modified version of the widely implemented ESL analytic rating scales was adopted for assessing essays, categories of which include *Organization*, *Grammar & Vocabulary*, *Function*, *Comprehensibility*, and *Mechanics* (Jacobs et al., 1981). Considering combined aspects of these categories, raters holistically assessed the essays and made placement decisions for test takers. Additionally, when raters failed to reach an agreement on the placement decision of an individual test taker, an academic writing program coordinator adjudicated on their disagreement.

3.2. Linguistic Analysis

In this study, nominal modifiers are defined as phrasal nominal premodifiers and postmodifiers as constituents of a complex noun phrase. The structure of a complex noun phrase with nominal modifiers under analysis in this study is as follows (Biber, Conrad, & Leech, 2002):

Determiner + nominal premodifiers + head (noun) + nominal postmodifiers

For instance, the noun phrase, *increasing competition for mobile cellular radio services*, consists of one *-ing* participial adjective as a nominal premodifier (*increasing*), one head noun (*competition*), and one prepositional phrase other than *of* with abstract meanings as a nominal postmodifier (*for mobile cellular radio services*).

As depicted in Table 2, seven nominal modifiers in the essays were analyzed. It is important to note that nominal modifiers in the later stages were deemed characteristics of essays written by advanced L2 writers. In this regard, nominal modifiers in Stages 4 and 5 were hypothesized to be acquired later than those in Stages 2 and 3. Nominal modifiers were manually coded by the researcher with the assistance of the annotated essays, as the essays were annotated by the Biber Tagger in a learner corpus. It is important to note that those nominal modifiers with spelling errors in the essays, which could have presumably affected raters' placement decisions, were excluded from the study. Additionally, prepositional phrases other than *of* that refer to tangible or physical entities (e.g., body parts) or specific locations (e.g., *China*) were coded as concrete or locative meanings; otherwise, they were coded as abstract meanings (e.g., *a barrier in communication*). Those prepositional phrases that were equally likely to be adverbial rather than a noun modifier (e.g., *in their backyards* in the noun phrase, *fruits and vegetables in their backyards*) were excluded from the analysis as well. The following examples illustrate how a variety of nominal modifiers were orchestrated in one of the essays in the study; prepositional phrases other than *of* (abstract) and prepositional phrases *of* are underlined, and premodifying adjectives are in bold.

*Malaysia has been well known for its **great** diversity in culture, religion, and race. The Malays, Chinese and Indians interact politely with one another although we all share a **different** set of culture, religion and language¹. However, it did not turn out as a barrier in communication.*
[Rules and Regulations_Essay_34_PASS]

¹ This study focused on syntactic aspects of a complex noun phrase. However, it is worthwhile to note that *culture, religions, and language* may be perceived as the head of the complex noun phrase, *a different set of culture, religions, and language*, when considering semantic aspects of the complex noun phrase.

To check the inter-coder reliability for seven nominal modifiers, another experienced ESL instructor coded a randomly selected 10% of the occurrences of those nominal modifiers. Cohen's κ showed that there was substantial agreement between two coders, $\kappa = .82$, 95% CI [0.75, 0.89], $p < .001$.

TABLE 2
Nominal Modifiers as Study Variables

Stage	Nominal Modifier ^a	Example from Essay
2	Premodifying adjective (attributive or participial adjective ^b)	great diversity (attributive) law-abiding citizens (-ing participial) inexperienced driver (-ed participial)
3	Premodifying noun Prepositional phrase <i>of</i> as nominal postmodifier Prepositional phrase other than <i>of</i> as nominal postmodifier (concrete / locative)	CO₂ emissions the use of modern conveniences an article in the magazine
4	Prepositional phrase other than <i>of</i> as nominal postmodifier (abstract)	a barrier in communication
5	Prepositional phrase with nonfinite complement clause Multiple prepositional phrases	a purpose of bringing further conveniences the negative effect on the social life of an individual

Note. Adapted from Biber et al. (2011). Words in bold are head nouns.

^aThat relative clauses as constituents of complex noun phrases were excluded from the analysis, as they were not found to be statistically salient grammar features of academic writing (Biber et al., 2011).

^bParticipial premodifiers were not originally included in Biber et al. (2011).

3.3. Statistical Analysis

The independent variables of the study were frequencies of seven nominal modifiers in the 374 essays (see Table 2) and the dependent variable of the study comprised three placement decisions for the 374 essays. Additionally, the tokens (word counts) and topics of the essays were plugged into the current regression model in order to control for their effects on placement decisions. Regarding the topics of the essays, Topic 1 (Modern conveniences) was designated as the reference group due to its largest sample size and its approximately equal distribution of placement decisions, as shown in Table 1. Additionally, AntConc 3.5 was used to compute the tokens of the 374 essays.

In order to answer the first research question, a cumulative odds ordinal logistic regression with proportional odds was fitted using SPSS 22.0, due to the ordinal nature of placement decisions and, more importantly, the violation of the assumption of homoscedasticity in ordinary least squares (OLS) regression (Cohen, Cohen, West, &

Aiken, 2013). Then, this logistic regression model was used to determine the degree to which the frequency of nominal modifiers in the developmental stages of grammatical complexity was related to placement decisions after adjusting for the effects of essay length and topics of the essays.

Regarding the second research question, due to the fact that each predictor was measured on a different measurement scale (i.e., topic, word count, frequency of occurrence), fully standardized coefficients of each predictor were computed to estimate the relative strength of effects of predictors in the model after fitting the model (Menard, 2002, 2011); although the nominal modifiers shared the seemingly identical measurement unit (i.e., the number of occurrences), this statistical adjustment was essential. This is mainly because one addition of premodifying adjectives to an essay was qualitatively different from one addition of multiple prepositional phrases to an essay. Obviously, two grammar features were not linguistically comparable to each other from the perspective of human raters. Thus, the computation of fully standardized coefficients of the nominal modifiers in this study is believed to shed light on the scale-free effects of predictors on placement decisions.

Before fitting the cumulative ordinal logistic regression model, the assumptions of proportional odds and multicollinearity among predictors in the model were checked using SPSS 22.0. In order to check the assumption of proportional odds, a full likelihood ratio test was conducted to compare a model with invariant coefficients of predictors to a model with varying coefficients of predictors. The existence of multicollinearity among continuous and categorical predictors was checked by consulting with variance inflation factors (VIFs).

Lastly, it is essential to note that against established principles in corpus-based research, the rates of occurrence of nominal modifiers were not normalized per 1,000 words, but calculated for individual essays, following recommendations by Biber and Gray (2015). This atypical approach was deliberately taken to examine the extent to which placement decisions for individual test takers were associated with their use of nominal modifiers. Additionally, this decision is believed to facilitate the interpretation of both unstandardized and fully standardized regression coefficients of the current logistic regression model. In order to address the issue of not normalizing the rates of occurrence of the grammar structures, the length of individual essays was plugged into the logistic regression model in this study.

4. RESULTS AND DISCUSSION

4.1. Effects of Nominal Modifiers on Placement Decisions

The first research question regarded the extent to which the use of nominal modifiers in the 374 argumentative essays was associated with placement decisions for incoming international undergraduates at one university in the United States. Using SPSS 22.0, a cumulative odds ordinal logistic regression with proportional odds was fitted to determine the degree to which frequencies of seven nominal modifiers were related to placement decisions. As shown in correlations among seven nominal modifiers and tokens in Table 3, there existed weak positive correlations. Table 4 illustrates the descriptive statistics of raw frequencies of nominal modifiers along with the length of essays across placement decisions.

Two major prerequisites of the cumulative odds ordinal regression model were checked: the assumption of proportional odds and the existence of multicollinearity among categorical and continuous predictors. Firstly, the assumption of proportional odds seemed to be reasonably met, $\chi^2(11) = 9.57, p = .570$. In other words, the effects of nominal modifiers along with the length and topic of the essay on placement decisions were not statistically different across two cumulative splits of three placement decisions (i.e., ENGL 101B [low] versus ENGL 101C [intermediate] and PASS [upper-intermediate], ENGL 101B [low] and ENGL 101C [intermediate] versus PASS [upper-intermediate]). Thus, the implementation of the cumulative odds ordinal logistic regression with proportional odds was justified in the study. Additionally, variance inflation factors (VIFs) of both categorical and continuous predictors did not exceed 10.00, ranging from a low of 1.10 (prepositional phrases with a nonfinite complement clause) with a high of 1.62 (premodifying adjectives) (Cohen et al., 2013).

Although the deviance goodness-of-fit demonstrated that the predicted placement decisions were not significantly different from the observed placement decisions, $\chi^2(735) = 656.42, p = .983$, it should be noted that 748 cells (66.7%) were sparse with zero frequencies, which was expected when eight continuous predictors and one categorical predictor were plugged into the logistic regression model (Menard, 2002). Nonetheless, along with the length and topics of the essays, the incidence of seven nominal modifiers statistically significantly predicted the placement decisions for the 374 essays over and above the intercept-only model, $\chi^2(11) = 138.22, p < .001$.

TABLE 3
Correlations for Nominal Modifiers and Tokens

	1	2	3	4	5	6	7	8
1. Token	—							
2. PA	.44**	—						
3. PN	.30**	.20**	—					
4. PP <i>of</i>	.32**	.37**	.14**	—				
5. Con PP	.26**	.22**	.10	.09	—			
6. Abs PP	.21**	.25**	.06	.06	.17**	—		
7. PPC	.20**	.17**	.16**	.12*	.04	.04	—	
8. MPP	.13**	.21**	.11**	.38**	.12*	.15**	.09	—

Note. PA = Premodifying adjective, PN = Premodifying noun, PP *of* = Prepositional phrase *of*, Con PP = Prepositional phrase other than *of* (concrete/locative), Abs PP = Prepositional phrase other than *of* (abstract), PPC = Prepositional phrase with nonfinite complement clause, MPP = Multiple prepositional phrase.

* $p < .05$, two-tailed, ** $p < .01$, two-tailed.

TABLE 4
Descriptive Statistics of Nominal Modifiers Across Placement Decisions

Variable	ENGL 101B ($n = 83$)		ENGL 101C ($n = 127$)		PASS ($n = 164$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Token	245.11	75.64	271.04	62.73	296.51	75.19
Premodifying Adjective	7.45	4.15	8.91	4.34	11.16	5.02
Premodifying Noun	2.48	2.59	3.31	3.00	4.04	3.11
PP <i>of</i>	1.36	1.63	2.02	1.64	3.26	2.37
PP other than <i>of</i> (Concrete / Locative)	0.63	0.89	0.81	1.00	1.00	1.09
PP other than <i>of</i> (Abstract)	0.42	0.72	0.48	0.79	0.87	1.14
PPC	0.10	0.34	0.24	0.63	0.36	0.67
Multiple PP	0.13	0.38	0.28	0.54	0.56	0.89

Note. PP = prepositional phrase, PPC = Prepositional phrase with nonfinite complement clause, ENGL 101B = low, ENGL 101C, = intermediate, PASS = upper-intermediate.

Overall, an addition of one nominal premodifier and that of three nominal postmodifiers to an essay appeared to be positively associated with an increase in the predicted odds of passing the EPT across two cumulative splits of three placement decisions, after adjusting for the effects of other independent variables, as shown in the odds ratios in Table 5.

TABLE 5
Ordinal Logistic Regression Coefficients of Nominal Modifiers on Placement Decisions

Variable ^a	B	β	SE	Wald	df	p	OR	95% CI	
								LL	UL
1st TH	1.70		0.47	13.09	1	< .000***	5.46	2.78	13.69
2nd TH	3.69		0.50	54.02	2	< .000***	40.07	14.98	107.22
Topic 4	-0.16	-0.02	0.35	0.20	1	.654	0.86	0.44	1.68
Topic 3	1.66	0.25	0.31	28.24	1	< .000***	5.24	2.84	9.64
Topic 2	0.91	0.14	0.30	9.87	1	.002**	2.48	1.41	4.37
Token	0.00	0.08	0.00	3.13	1	.077	1.00	1.00	1.01
PA	0.08	0.13	0.03	6.77	1	.009**	1.08	1.02	1.14
PN	0.05	0.05	0.04	1.58	1	.209	1.05	0.98	1.13
PP <i>of</i>	0.24	0.18	0.07	13.27	1	< .000***	1.27	1.12	1.44
Con PP	0.15	0.05	0.11	1.78	1	.182	1.16	0.93	1.44
Abs PP	0.26	0.09	0.13	4.14	1	.042*	1.29	1.01	1.66
PPC	0.13	0.03	0.21	0.40	1	.530	1.14	0.76	1.70
MPP	0.41	0.10	0.20	4.14	1	.042*	1.50	1.02	2.22

Note. Pseudo $R^2 = .31$ (Cox & Snell), $.35$ (Nagelkerke), $.17$ (McFadden). OR = odds ratio, CI = confidence interval, LL = lower limit, UL = upper limit, 1st TH = Threshold between ENGL 101B (low) and ENGL 101C (intermediate), 2nd TH = Threshold between ENGL 101C (intermediate) and PASS (upper-intermediate). Topic 2 = Rules and regulations, Topic 3 = Preferences for e-books, Topic 4 = Violence on TV. PA = Premodifying adjective, PN = Premodifying noun, PP *of* = Prepositional phrase *of*, Con PP = Prepositional phrase other than *of* (concrete/locative), Abs PP = Prepositional phrase other than *of* (abstract), PPC = Prepositional phrase with nonfinite complement clause, MPP = Multiple prepositional phrases.

^aThe categorical variable, Topic in this study was dummy coded. Topic 1 (Modern conveniences) was designated as the reference group.

* $p < .05$, ** $p < .01$, *** $p < .001$.

More specifically, given one addition of premodifying adjectives to an essay, the predicted odds of passing the EPT across two cumulative splits of three placement decisions was multiplied 1.08 times, 95% CI [1.02, 1.14], $\chi^2(1) = 6.77$, $p = .009$, holding constant all other variables. For each addition of prepositional phrases *of* to an essay, the predicted odds of passing the EPT across two cumulative splits of three placement decisions was multiplied 1.27 times, 95% CI [1.12, 1.44], $\chi^2(1) = 13.27$, $p < .001$, holding constant all other variables. Similarly, for every one addition of prepositional phrases other than *of* (abstract) to an essay, the predicted odds of passing the EPT across two cumulative splits of three placement decisions was multiplied 1.29 times, 95% CI [1.01, 1.66], $\chi^2(1) = 4.14$, $p = .042$, after accounting for all other predictors. For every one addition of multiple prepositional phrases to an essay, the predicted odds of passing the EPT across two cumulative splits of three placement decisions was multiplied 1.50 times, 95% CI [1.02, 2.22], $\chi^2(1) = 4.14$, $p = .042$, after holding constant all other predictors. However, an addition of three nominal modifiers to an essay was not statistically significantly associated

with placement decisions: premodifying nouns, prepositional phrases other than *of* (concrete/locative), and prepositional phrases with nonfinite clause complements.

Although the length of the essays was not a significant predictor, the topics of the essays appeared to be associated with the predicted odds of passing the EPT across two cumulative splits of three placement decisions. To illustrate, an essay on Topic 1 (Modern conveniences) was about 2.48 and 5.24 times likely to pass the EPT across two cumulative splits of three placement decisions than an essay on Topics 2 (Rules and regulations) and 3 (Preferences for e-books) respectively, when holding constant all other variables. However, the predicted odds of passing the EPT for an essay on Topic 1 (Modern conveniences) was not statistically different from those of passing the EPT for an essay on Topic 4 (Violence on TV), when holding constant all other variables. However, it is important to note that the effects of topics on placement decisions should be interpreted with caution as the number of original placement decisions was not balanced except for Topic 1 (Modern conveniences), as shown in Table 1. Thus, the effects of topics might be a simple reflection of unbalanced distribution across placement decisions in this study.

A final model with the four significant nominal modifiers was fitted to compute the overall success of prediction for placement decisions. As illustrated in the confusion table (see Table 6), the overall accuracy of prediction for placement decisions amounted to 52.41% (196/374); the correct placement decisions (i.e., the main diagonal of the confusion table [i.e., 196]) was divided by the total number of placement decisions (i.e., 374). The confusion table shows the positive association between original and predicted placement decisions for the 374 essays. Goodman and Kruskal's γ indicated a robust, positive association between original and predicted placement decisions, which was statistically significant ($G = .59, p < .001$). Similarly, symmetric Somers' d showed that there was a strong, positive association between original and predicted placement decisions ($d = .39, p < .001$).

The overall findings of this study bolster the claim that nominal modifiers are useful for differentiating between adjacent levels of L2 writers despite the holistic nature of placement decisions; therefore, the results of this study appear to contradict the previous finding that individual grammatical complexity measures were generally not good predictors of holistic writing scores (Biber et al., 2016). As illustrated in Tables 5 and 6, the four nominal modifiers in the developmental stages of grammatical complexity (i.e., premodifying adjectives, prepositional phrases *of*, prepositional phrases other than *of* [abstract], multiple prepositional phrases) was found to be positively associated with placement decisions for test takers. Additionally, the positive and robust association between observed and predicted placement decisions existed for test takers. In this regard, the results of this study appear to support the substantial effects of grammar competence of

L2 writers on raters' judgments of the overall quality of essays (Alderson & Huhta, 2005; McNamara, 1990).

TABLE 6
Confusion Table of Original and Predicted Placement Decisions

Original	Predicted			
	ENGL 101B	ENGL 101C	PASS	Total
ENGL 101B (low)	23 (27.7%)	44 (53.0%)	16 (19.3%)	83(100%)
ENGL 101C (intermediate)	19 (15.0%)	57 (44.9%)	51 (40.2%)	127 (100%)
PASS (upper-intermediate)	7 (4.3%)	41 (25.0%)	116 (70.7%)	164 (100%)
Total	49 (16.8%)	142 (38.0%)	183 (45.2%)	374 (100%)

The following students' essays on Topic 1 (Modern conveniences) seem to further support the finding that the use of nominal modifiers was positively associated with raters' placement decisions; in the following specific examples, nominal modifiers are underlined.

We can go anywhere we won't without worrying about the time or the way. Today, we can travel by car, by train, or even by plane. However, 30 years, people who want to go another country need to spend many money. When my father was a student, he walked to school, it would take him 3 hours to go to school. But now, we can arrive at school in 10 minutes. [Modern conveniences_Essay_76_ENGL 101C_ Intermediate]

These resources provided to the modern community are essential for the development and smooth running of the society. Making life easier for its people, Amenities like automatic teller machine, fast food, etc. has become an important part of the society. [Modern conveniences_Essay_122_PASS_Upper Intermediate]

Rather than implementing nominal modifiers, Essay 76 heavily relied on personal pronouns such as *we* and *he*. As a characteristic of conversational discourse (Biber et al., 2002), the use of personal pronouns might have affected the placement decision awarded to Essay 76. On the other hand, Essay 122 exhibited more nominal modifiers, supporting the previous finding that advanced L2 writers tended to yield more complex noun phrases (Parkinson & Musgrave, 2014; Taguchi et al., 2013).

Regarding the usefulness of two nominal premodifiers including premodifying adjectives and premodifying nouns, the results of this study seem inconsistent with previous findings. The use of two premodifiers (premodifying adjectives and premodifying nouns) was found to be negatively correlated with human ratings awarded to argumentative

essays (Staples & Reppen, 2016), and no significant differences in the use of premodifying adjectives were detected among abstracts in research articles written by advanced and less advanced Persian graduate-level L2 writers (Ansarifar et al., 2018).

The reason for this difference seems to have resulted from the inclusion of research articles and the nature of different statistical approaches (Staples & Reppen, 2016). More specifically, the population of this study (i.e., incoming international undergraduate students) could have resulted in the difference, as Ansarifar et al. (2018) analyzed master's thesis and PhD dissertation abstracts written by advanced L1 Persian writers. Additionally, the separate analysis of attributive and participial adjectives as nominal premodifiers in Ansarifar et al.'s study might have led to the difference with respect to the significance of adjectives as nominal premodifiers; in this study, attributive and participial adjectives as nominal premodifiers were collectively coded as premodifying adjectives due to the low frequency of participial adjectives in the current data set. Lastly, as Staples and Reppen (2016) investigated the correlations between nominal premodifiers and human ratings, the implementation of a cumulative odds logistic regression might have resulted in the difference with respect to the significance of adjectives as nominal premodifiers.

Prepositional phrases other than *of* (concrete/locative) and prepositional phrases with nonfinite complement clauses were not found to be significant predictors for course placement decisions in this study. This finding appeared to be consistent with previous findings on the developmental stages of grammatical complexity (Ansarifar et al., 2018; Parkinson & Musgrave, 2014) that prepositional phrases other than *of* (concrete/locative) did not exhibit statistically significant differences between advanced and less advanced graduate-level L2 writers. However, this comparison also needs to be made with caution, as the two comparison studies primarily analyzed research articles, while argumentative essays written by L2 writers were analyzed in this study.

Especially noteworthy were the invariant effects of the four nominal modifiers on placement decisions across two cumulative splits of three placement decisions (i.e., ENGL 101B [low] versus ENGL 101C [intermediate] and PASS [upper-intermediate], ENGL 101B [low] and ENGL 101C [intermediate] versus PASS [upper-intermediate]) in this study, as evidenced by the satisfaction of the assumption of proportional odds in the cumulative ordinal logistic regression model. The satisfaction of this assumption supported not only the robustness of the four nominal modifiers as useful discriminators between adjacent levels of L2 writers but also their constant effects on human raters' judgments of the overall quality of argumentative essays across three placement decisions. Accordingly, the findings of this study further confirm the previous finding on complex nominals in argumentative essays; it was found that two complex nominal measures (complex nominals per clause and T-unit), which were the closest proxies for the four nominal

modifiers in this study, exhibited linear developmental patterns for adjacent levels of Chinese ESL university-level L2 writers (Lu, 2011).

However, when it comes to the overall accuracy of prediction for placement decisions, the final model with the four nominal modifiers made more accurate predictions for PASS essays than the other two placement decisions. More specifically, as shown in Table 6, 16 out of 83 essays (19.3%) written by low-level test takers were incorrectly classified as PASS essays, while 51 out of 127 essays (40.2%) written by intermediate-level test takers were graded as PASS essays. As the underlying cause for these deviations is difficult to pinpoint due to the research design in the current study, the imprecise prediction for essays written by low- and intermediate-level test takers seems to merit future investigations.

Somewhat surprisingly, in this study, the length of an essay was not found to be a significant predictor for placement decisions, contradicting findings from previous studies (Ferris, 1994; Grant & Ginther, 2000; Larsen-Freeman, 1978, 2006). Although these studies supported the commonly held intuition by ESL instructors that the length of essays is a significant discriminator for L2 writers, the results of the current study demonstrated that the length of an essay alone could not be used to predict the quality of written works when the holistic rating scale was implemented. Consistent with the current study, Connor (1990) also pointed out that the length of essays only accounted for 6% of the variance of the holistic scores of argumentative and persuasive essays written by first language English high school students.

4.2. Relative Strength of Nominal Modifiers

The second research question regarded the extent to which the relative strength of the effects of the nominal modifiers aligned with their corresponding developmental stages of grammatical complexity (Biber et al., 2011). To investigate the relative strength of the effects of the nominal modifiers on placement decisions in this study, fully standardized coefficients of predictors in the model were computed, following steps² outlined by Menard (2011). As illustrated in the fully standardized coefficients in Table 5, prepositional phrases *of* appeared to have the most potent impact on placement decisions, followed by premodifying adjectives, multiple prepositional phrases, and prepositional phrases other than *of* (abstract). Regarding prepositional phrases *of*, the most substantial predictor in the model, a 1.00 standard deviation increase in this grammar structure, was

² The equation proposed by Menard (2011) was used to compute the fully standardized coefficients of the ordinal logistic regression model in this study. It should be noted that when the assumption of proportional odds was met in the polytomous ordinal logistic model as it was in this study, the computation procedures for the dichotomous dependent variable are the same as those for the polytomous dependent variable.

associated with a 0.18 standard deviation increase in the logit of placement decisions when adjusting for other effects of the predictors. Among the statistically significant nominal modifiers, the relative strength of prepositional phrases other than *of* (abstract) on placement decisions appeared to be the weakest, as evidenced by its fully standardized coefficient, 0.09.

Overall, the computation of standardized coefficients of the nominal modifiers in the cumulative ordinal logistic regression model rendered it possible to establish the relative strength of the effects of the nominal modifiers on placement decisions. It was found that prepositional phrases *of* seemed to have the most substantial effect on raters' judgments of the overall quality of essays after adjusting for the effects of other variables; on the other hand, prepositional phrases other than *of* (abstract) exhibited the weakest effect on raters' judgments when holding constant the effects of other variables.

Accordingly, the order of the relative strength of effects of the four nominal modifiers on placement decisions did not appear to align with their corresponding developmental stages of grammatical complexity. More specifically, the most influential grammar structure on placement decisions, prepositional phrases *of*, was located in Stage 3 in the developmental hierarchy in grammatical complexity. By contrast, the least influential grammar structure on placement decisions, prepositional phrases other than *of* (abstract), was located in the more advanced Stage 4.

As indicated in the description of the study population, this misalignment between the relative strength of significant nominal modifiers and their corresponding developmental stage could be attributed to the range restriction that the most advanced international undergraduates were already excluded from the study population due to the exemption policy at this institution (Bachman, 2004). Accordingly, it could be that this policy already exempted the most advanced international undergraduates with a high score on the standardized English proficiency tests (e.g., TOEFL and IELTS) from the EPT prior to analysis in this study. Moreover, timed argumentative essays might not elicit complex nominal modifiers as Kyle and Crossley (2018) pointed out the scarcity of phrasal nominal modifiers in timed argumentative essays despite their usefulness for discriminating between advanced and less advanced L2 writers.

In summary, the fully standardized coefficients of significant nominal modifiers in this study indicated that prepositional phrases *of* as nominal postmodifiers had the most potent impact on placement decisions. Furthermore, the misalignment between the relative strength of significant nominal modifiers and their corresponding developmental stages seemed to partially contradict the proposed developmental stages of grammatical complexity (Biber et al., 2011). Accordingly, it is tentatively suggested that the use of prepositional phrases *of* could be most useful for differentiating between timed

argumentative essays written by L2 writers who are not adequately advanced enough to be exempted from an EPT.

5. CONCLUSION

The overall results of this study indicate that the four nominal modifiers are salient grammar features characteristic of L2 writers, partially validating Biber et al.'s (2011) developmental stages of grammatical complexity for L2 writers. In particular, the results of logistic regression analysis show that the effects of the use of prepositional phrases *of* on placement decisions by human raters are the strongest.

Educational implications of this study are that the use of nominal modifiers in developmental stages of grammatical complexity proposed by Biber et al. (2011) is found to be positively related to human raters' judgments of the overall quality of essays in the context of an EPT. Thus, ESL instructors of remedial academic writing courses could exploit this finding to predict what grammar structures are learnable and teachable for those L2 writers who are not advanced enough to be exempted from an EPT.

More importantly, the partial validation of the developmental stages of grammatical complexity in this study is believed to assist the development of rating scales, as the empirical findings of this study could facilitate the establishment of an explicit relationship between rating scale descriptors and salient linguistic features exhibited by L2 writers (Brindley, 1998). Similarly, the group of salient linguistic features exhibited by L2 writers in this study could be incorporated into rater training materials to further clarify existing rating scale descriptors for grammatical complexity (e.g., effective complex constructions). As ambiguous rating scale descriptors have been identified as the source of discrepant ratings among raters, the utilization of corpus-driven rating scale descriptors could help resolve the differences among raters in the context of EFL in South Korea (Kim, 2011). Similarly, it may be desirable for rater training sessions to draw raters' attention to the findings of this study.

Study limitations are that the reported results may not be generalizable to highly advanced L2 writers and other genres of writing, such as persuasive essays and research articles written by L2 writers. In this regard, future research should be conducted to gauge the extent to which writing genres moderate the relationship between the use of grammar structures and human ratings. Additionally, the lack of inter-rater reliability is a potential weakness of this study. Thus, future research should check inter-rater reliability among raters and, more importantly, adjust for rater severity using many-facet Rasch measurement (Eckes, 2015). Most importantly, due to the holistic nature of placement decisions in this study, future research on analytic rating scales is necessary to further

elucidate the extent to which the use of nominal modifiers is related to human raters' judgments of grammatical complexity alone.

Applicable levels: Tertiary

REFERENCES

- Alderson, J. C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, 22(3), 301-320.
- Ansarifar, A., Shahriari, H., & Pishghadam, R. (2018). Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *Journal of English for Academic Purposes*, 31, 58-71.
- Bachman, L. F. (2004). *Statistical analyses for language assessment*. New York: Cambridge University Press.
- Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken and written English*. London: Longman.
- Biber, D., & Gray, B. (2015). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge, England: Cambridge University Press.
- 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., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37(5), 639-668.
- Brindley, G. (1998). Describing language development? Rating scales and SLA. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 112-140). New York: Cambridge University Press.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42-65.
- Carr, N. T. (2011). *Designing and analyzing language tests*. New York: Oxford University Press.

- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: Routledge.
- 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.
- Eckes, T. (2015). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments* (2nd ed.). Frankfurt, Germany: Peter Lang.
- Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing*, 13(2), 208-238.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9(2), 123-145.
- Jacobs, H., Zingraf, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.
- Kim, B. (2011). Resolving discrepant ratings in writing assessments: The choice of resolution method and its application. *English Teaching*, 66(2), 211-231.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275-304.
- Kyle, K., & 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.
- Lan, G., Lucas, K., & Sun, Y. (2019). Does L2 writing proficiency influence noun phrase complexity? A case analysis of argumentative essays written by Chinese students in a first-year composition course. *System*, 85, 102-116.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 12(4), 439-448.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590-619.
- 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.
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, 19(3), 246-276.
- McNamara, T. F. (1990). Item response theory and the validation of an ESP test for health professionals. *Language Testing*, 7(1), 52-76.
- Menard, S. W. (2002). *Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage.
- Menard, S. W. (2011). Standards for standardized logistic regression coefficients. *Social Forces*, 89(4), 1409-1428.

- Musgrave, J., & Parkinson, J. (2014). Getting to grips with noun groups. *ELT Journal*, 68(2), 145-154.
- North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15(2), 217-263.
- 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.
- Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexicogrammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing*, 32, 17-35.
- 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.
- Turner, C. E., & Upshur, J. A. (2002). Rating scales derived from student samples: Effects of the scale maker and the student sample on scale content and student scores. *TESOL Quarterly*, 36(1), 49-70.
- Weigle, S. C., & Friginal, E. (2015). Linguistic dimensions of impromptu test essays compared with successful student disciplinary writing: Effects of language background, topic, and L2 proficiency. *Journal of English for Academic Purposes*, 18, 25-39.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity* (Technical Report No. 17). Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawaii.
- Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28, 53-67.