THE JOURNAL OF ASIA TEFL
Vol. 18, No. 3, Fall 2021, 745-763
http://dx.doi.org/10.18823/asiatefl.2021.18.3.1.745

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511 © 2004 AsiaTEFL.org. All rights reserved.

Syntactic Complexity Measures and Academic Writing Proficiency: A Corpus-based Study of Professional and Students' Prose

Rajab Esfandiari
Imam Khomeini International University, Qazvin, Iran

Mohammad Ahmadi
Imam Khomeini International University, Qazvin, Iran

This study examines large-grained, medium-grained, and fine-grained measures, as different dimensions of syntactic complexity, and their relationship with academic writing proficiency in the abstract sections of academic prose. We sampled 237 texts written by expert academic writers of research articles, PhD level L1 Persian writers, and MA level L1 Persian writers. Using one-way multiple analysis of variance (MANOVA) and ordinal logistic regression, we compared the abstracts in terms of their syntactic complexity and checked the predictability of large-grained, medium-grained, and fine-grained measures in academic writing proficiency. The findings revealed that the differences among the groups were not large enough for clausal and sentential measures of dependent clauses per T-unit, and clauses per sentence to show statistical significance. However, the differences among the groups on length-based measures of mean length of T-unit, and mean length of clause along with noun phrase measures of clauses per nominals, noun premodifiers, and prepositional phrases reached significance. Analysis of fine-grained measures of noun phrase complexity shows that four indices that measured the number and variation of dependents in the texts made unique statistically significant contribution to the prediction of academic writing proficiency. These findings suggest the importance of incorporating complementary measures of syntactic complexity (i.e., large-grained measures) as integral parts of L2 writing instruction practices. The implications of the findings in relation to L2 academic writing pedagogy are discussed.

Keywords: academic proficiency, academic writing, syntactic complexity measures

Introduction

From among complexity, accuracy, and fluency (CAF) triad proposed by Skehan (1996) to account for L2 learners' proficiency, complexity has received extensive treatment in L2 literature. As Crossley and McNamara (2014) put it, the reason why complexity has gained increasing attention from researchers is that the arrangement of words to clauses indicates the combinational property of the language, which can be reliably used to assess acquisition. Similarly, Ortega (2003) argues that syntactic complexity is of great significance because of the assumption that language development involves increasing the learners' syntactic complexity repertoire and their ability to use it in a variety of contexts.

Large-grained measures of syntactic complexity have mainly relied on dependent clause as an important measure of L2 writing. A dependent clause is defined as the clause which cannot stand on its own and "must be used with another clause to form a complete grammatical construction" (Richards & Schmidt, 2013, p. 162). Since dependent clauses "have greater influence on length" (Biber et al., p. 9)

than measures of other grammatical units (e.g., phrases), they have been favored as a significant measure of grammatical complexity in writing. Reliance on clausal metrics for gauging writing proficiency is also encouraged by the impression that increased grammatical complexity is the result of structural elaboration and clausal embedding (Biber et al., 2011). It is assumed that the proportion of dependent clauses, those which are structurally incomplete and semantically depend on the main clause of the sentence (Swan & Baker, 2008), to total number of clauses in written prose produced by learners is associated with the development in proficiency (Iwashita, 2006).

Although clausal measures (e.g., mean length of T-unit and dependent clauses per T-unit) have been used exhaustively in the previous studies, the inconsistency of the results obtained makes it difficult to interpret the findings (See Wolfe-Quintero et al., 1998). The inconsistency could be partly attributed to the lack of clear and operational definition of the term 'syntactic complexity' (Blute & Housen, 2012; Lu, 2010) which could possibly circumscribe the construct and identify its core components. The most significant problem with using large-grained measures of syntactic complexity (clausal measures in particular) is their inability to capture the dimensions of complexity other than clausal features (Biber et al., 2011).

While noun phrase measures of syntactic complexity better characterize advanced academic writing (Staples et al., 2016), not all writing produced by L2 English writers is academic, and there is no evidence that even most of it is (Casal & Lee, 2019). Hence, by evaluating academic writing solely through phrasal measures, we miss an important aspect of academic texts which is complexity at clausal levels. As Norris and Ortega (2009) put it "at a minimum SLA researcher should measure large-grained or general complexity, complexity by subordination, and complexity via phrasal elaboration, as well as possibly coordination if early proficiency data are also included" (p. 574). Norris and Ortega further argue that in order to capture large-scale or long-term changes in language complexity, we still need to have overall (large-grained) measures to the extent that fine-grained measures would miss these changes.

Following the call by previous research for employing syntactic complexity indices that measure large-grained, medium-grained, and fine-grained constructs (e.g., Norris & Ortega, 2009; Vyatkina ,2013), the present study focused on both fined-grained phrasal, and large-grained complexity measures. By employing a relatively large repertoire of the measures that tap into different dimensions of syntactic complexity, a more accurate understanding of L2 complexity could be achieved. As Norris and Ortega (2009) stated, large-grained and fine-grained complexity measures complement each other since they capture different aspects of syntactic complexity.

Although previous research has deepened our understanding of complexity features of academic texts produced by L2 writers, they typically employed one or two popular measures which result in the reduction of multidimensional construct of complexity into one or a few, at best (Bulté & Housen, 2014). Few recent studies, however, examined the link between complexity measures and writing proficiency or quality using a number of measures (e.g., Casal & Lee, 2019; Youn, 2014); they typically focused on the data obtained from rated essays in the context of classes (for example, first-year writing courses). Little research has been conducted to investigate and compare the complexity of the texts produced in the higher education context with those of expert academic writers. The gap becomes significant when we consider the importance of academic writing proficiency in higher education. The language that receives recognition throughout undergraduate studies may not be appropriate in higher education. Graduate students are likely to write research articles, theses, or dissertations soon. Accordingly, familiarity with the conventions of academic writing, especially syntactic complexity, is particularly important for the learners entering higher education.

As for fine-grained measures of phrasal complexity, while the measures proposed by Biber et al. (2011) provided adequate description of complexity features in the text, i.e., noun phrase modifiers, the attributive function of the nominals being modified (e.g., nominative, accusative, or dative noun phrases) is not clear to interpret. For example, post-modifying prepositional phrase '*of foreign languages*' could be used to modify the nominative '*the learners*' as in '*The learners of foreign languages took content courses*' or it can be used to modify the dative noun phrase as in '*The first research question concerned the effect of task complexity on the learners of foreign language'*. In addition, previous noun phrase

measures did not capture variation of noun phrase modifiers. As a result, there will be no difference between the complexity of the texts that employ the same number of noun modifiers with different modifier variation. Accordingly, throughout this study, we call them medium-grained measures of noun phrase complexity so that they are differentiated from fine-grained measures.

Literature Review

Multidimensional Nature of Syntactic Complexity

Recently, there has been a growing call for multidimensional investigation of syntactic complexity in L2 writing studies (Bulté & Housen, 2012; Lu & Ai, 2015; Norris & Ortega, 2009; Ortega, 2003). Norris and Ortega (2009) argue that multifaceted nature of syntactic complexity is theoretically and empirically motivated. The construct validity of syntactic complexity is, to a large extent, dependent upon its multidimensionality such that any improper characterization of the term 'complexity' may question the results of the studies.

Multidimensional nature of complexity is not often investigated adequately in L2 writing research. That is because the constructs of complexity are still poorly defined and operationalized in L2 writing literature. Bulté and Housen (2018) argue that in spite of abundance of empirical research studies exploring complexity in L1 and L2, the majority of them suffer from inappropriate operationalization of the construct in a way that its multidimensionality has not been captured. Given that complexity is "the most complex, ambiguous, and least understood" construct of the CAF triad (Housen & Kuiken, 2009, p. 463), it seems necessary to limit its scope of definitions so that its dimensions can be identified and operationalized.

The first step in developing measures of complexity is to consider what subconstructs might constitute the construct of complexity (Connor-Linton & Amoroso, 2014). As Norris and Ortega (2009) put it, complexity is a multidimensional construct that consists of overall complexity, subordination, phrasal complexity, and a variety of forms. In L2 research, overall complexity is usually defined as "the range of forms (i.e., items, structures, patterns, rules) available to a learner and as the degree of sophistication of these forms, reflecting the distinction between, respectively, absolute and relative complexity" (Housen et al., 2019, p. 9), with relative complexity referring to the difficulty of processing the L2 system and absolute complexity relating to the number of discrete linguistic features and the number of connections between linguistic elements (Bulté & Housen, 2012). Syntactic complexity is a multidimensional construct that entails distinct domains such as complexification at clausal, phrasal, and global levels. Clausal complexity is based on the premise that any addition to, or modification of, the simple clause pattern, which consists of only a subject, a verb and an object, will result in complex or elaborated grammar (Biber & Gray, 2016). Clausal complexity measures the amount of clausal subordination in the text. Subordination concerns the degree of embedding (Iwashita, p. 158) or the number of independent clauses which are subordinate to the main clause. Global (large-grained) syntactic complexity measures, on the other hand, do not clarify which type of modification contributes to the complexity of the texts (Kreyer & Schaub, 2018). Accordingly, these large-grained measures are sometimes called 'omnibus measures' (See Biber et al., 2020), as they combine complexity at phrasal and clausal levels. Phrasal complexity measures are finer-grained indices that capture complexation by complex noun phrases modified by pre- and post-modifiers. That is to say, the more constituents a noun phrase contains, the more complex it becomes. Accordingly, complexity measures that are used in the literature correspond to the different subcomponents that are believed to constitute the construct of complexity.

Large-grained Measures of Syntactic Complexity in L2

Many of the studies in writing used the construct of complexity to describe the learners' produced texts. However, the definition, operationalization, and measurement of the construct remain a major concern.

The studies adopt a production perspective on complexity on the ground that the language the learners produce becomes more complex as they move away from preliminary language skills to more advanced ones (Biber et al., 2011). Bulté and Housen (2014) argue that the vast majority of the studies have used complexity measures under the assumption that more language elements like morphemes, lengthier language units like texts, more embedded elements like subordinate clauses, more varied language structures, and more sophisticated and marked language features imply more complex language.

Given the fact that a simple clause has only a subject, a verb, and an object or a complement, any addition to, or modification of, these patterns results in complex or elaborated grammar (Biber & Gray, 2016). Based on this view, the researchers have used the measures like T-unit, C-unit, mean length of clause and so forth to assess syntactic complexity. In general, the measures which have been used to assess syntactic complexity are categorized into five major groups, which are presented in Table 1. They include length-based complexity, sentence-based complexity, subordination-based complexity, coordination-based complexity, and particular structures. Syntactic Complexity Analyzer (SCA), which is a computational tool for measuring the syntactic complexity of writing samples (See Lu, 2010), draws on the full set of 14 measures as categorized in Table 1.

Although the measures categorized in Table 1 are often used to predict group differences, the results are rarely open to interpretations. That is why they are sometimes called *omnibus measures* (Biber et al., 2020) or *global measures* (Norris & Ortega, 2009), as they can be increased in a wide variety of ways such as lengthier clauses or lengthier phrases (Casal & Lee, 2019). As a result, Biber et al. (2011) proposed trajectories for development of noun phrase complexity which are organized around two parameters of grammatical relations and syntactic functions.

TABLE 1
Lu's (2010) Classification of Syntactic Complexity Measures

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

As shown in Table 1, length-based complexity, sentence complexity, subordination-based complexity, coordination-based complexity, and particular structures are different sub-constructs of syntactic complexity. Length-based measures (Type 1) assume that the more words or constituents a unit of production (e.g., a clause) contains, the more complex it becomes. This view is based on the findings of some studies that as learners become more advanced and more proficient, they use more sophisticated syntactic complexity structures that are realized in longer units of production (Ortega, 2003). Sentence complexity (Type 2) assesses the amount of sophistication of the structures in the unit of the sentence. Subordination-based complexity (Type 3) takes into account the amount of embedding (dependent clauses) a unit of production contains. Coordination-based complexity (Type 4) measures complexification through coordinating conjunctions which "join linguistic units which are equivalent or of the same rank" (Richards & Schmidt,

2013, p. 117). Finally, particular structures (Type 5) measure complexity at the level of phrases (verb phrases, or noun phrases) in T-unit or clauses.

Phrasal (Medium-grained) Measures Syntactic Complexity

Building on the previous studies on multidimensional analysis of register variation and syntactic complexity, Biber et al. (2011) and Biber and Gray (2016) have documented that complexity in written register is characterized by dependent phrases, while in spoken register it is featured by dependent clauses. The communicative demand of any register informs the language features associated with that communicative need (Biber & Gray, 2011). Spoken register is influenced by the pressure of rapid online processing (Ravid & Berman, 2010) and immediate context, whereas written language is less affected by online processing of information.

Given the fact that spoken and written registers are structurally different (Halliday, 1985; Biber et al., 2011) and complexity metrics that we use to measure these registers need to be different, many scholars have cast doubt on the application of large-grained measures of syntactic complexity to writing (Biber & Gray, 2010; Norris & Ortega, 2009; Ravid & Berman, 2010). Norris and Ortega (2009), for example, state that complexity is multimodal and conventional ways of measuring complexity fail to capture its dynamic and evolving nature. Likewise, Bulte and Housen (2012) comment that large-grained measures of syntactic complexity fail to specify the relationship between complexity and second language development.

In an attempt to make an index of writing development through analyzing predominant lexico-grammatical features of spoken and written registers, Biber et al. (2011) proposed writing developmental trajectory which starts with the features that are more prevalent in conversation such as finite dependent clauses, continues with intermediate stage of nonfinite dependent clauses, and finally ends with dense uses of phrasal expressions (See Table 2). Unlike omnibus measures of syntactic complexity that collapse multiple structural units into a single measure, medium-grained noun phrase measures of syntactic complexity are based on both descriptive and predictive understanding of the construct in a way that the range of linguistic features is distinguished (Biber et al., 2020).

TABLE 2
Biber et al.'s (2011) Hypothesized Stages of Writing Development

Stage	Grammatical Structure	Examples from the Corpus of the Current Study
2	Attributive adjectives	Quantitative measures
	Relative clauses	Universities that provided the corpus
3	Nouns as pre-modifiers	Complexity measures
	Possessive noun as pre-modifiers	Learners' development
	Of phrase (concrete/locative meanings)	Group of learners
	Prepositions as noun post-modifiers other than of (concrete/locative meanings)	Clauses in the sentences
4	ed-participle as post-modifiers	Features identified through manual check
	ing-participle as post-modifiers	Measures targeting different components
	Attributive adjectives, nouns as pre-modifiers	Overall writing quality
	Of phrase (abstract meanings)	Perception of progress
	Prepositions as noun post-modifiers other than of (abstract meanings)	A wealth of research on writing development
5	Preposition + nonfinite complement clause	Difficulty in using these properties
	Complement clauses controlled by nouns	Tasks that the learners accomplish
	Appositive noun phrases	EFL (English as a foreign language)
	Multiple prepositional phrases as post-modifiers, with levels of embedding	Repetition of structures in sentences with discourse-level coherence within writing productions

Fine-grained Measures of Noun phrase Complexity

While phrasal complexity measures proposed by Biber et al. (2011) are assumed to be descriptively and predictively adequate and have attracted empirical investigations (Lan & Sun, 2019; Parkinson & Musgrave, 2014), they do not include functional attributes of the noun phrases being modified (e.g., nominative, accusative, or dative noun phrases), nor do they specify the variation of the modifiers (Kyle, 2016). As a result, in addition to large-grained measures, and phrasal measures of syntactic complexity, the present study also employed fine-grained measures of phrasal complexity proposed by Kyle (2016) which calculate the average number of dependents per each phrase type, the occurrence of particular dependent types, and the average occurrence of particular dependent types in particular types of noun phrases.

Previous Studies on Syntactic Complexity and Writing Proficiency

There is a growing body of literature on the relationship between writing proficiency and syntactic complexity, especially in academic settings. In a recent study, Jiang et al. (2019) explored large-grained and fine-grained measures of syntactic complexity to predict writing proficiency among intermediate and beginning L2 English learners. They employed syntactic complexity analyzer and dependency syntactically-annotated corpus to collect data for large-grained and fine-grained measures, respectively. The results indicated that large-grained measures of mean length of T-unit, and mean length of sentence, and clausal measure of dependent clauses per clause can better predict writing proficiency compared to other measures. In addition, more proficient writers placed heavier reliance on adverbial clauses, prepositional phrases, adjectival relative clauses, and complement clauses than less proficient writers did. In a similar study, Bi and Jiang (2020) used syntactic complexity for assessing EFL learners' writing quality. By analyzing 410 narrative essays of Chinese EFL learners, they found that mean length of sentence, complex nominal per clause, and clauses per T-unit were the strongest predictors of writing quality that explained nearly 37% of the variance in the dependent variable. The relationship between syntactic complexity measures and writing quality was also investigated by Casal and Lee (2019), who based their study on a corpus of 280 student papers across three proficiency levels. They revealed that there was a lack of relationship between clausal complexity and writing quality. On the other hand, length-based and medium-grained complexity measures could discriminate between proficiency levels.

In order to capture the multidimensionality of syntactic complexity, and in order to fill the gap of research on large-grained, medium-grained, and fine-grained complexity features of the academic texts produced by non-native academic writers in higher education, and compare them to those of expert academic writers, the first phase of the study investigates the extent to which MA, PhD, and expert academic writers differ in their syntactic complexity. Given the importance of phrasal complexity in advanced academic writing, and in order to overcome the limitations of large-grained complexity measures (See Biber et al., 2020), by employing the fine-grained indices of syntactic complexity at phrasal levels proposed by Kyle (2016), the second phase of the study explores the predictability of fine-grained measures of phrasal complexity in academic register. Therefore, the present study seeks to answer the following two research questions:

1. To what extent do MA, PhD, and expert academic prose writers differ in their syntactic complexity as assessed by large-grained syntactic and medium-grained phrasal complexity measures?
2. Which measure(s) of fine-grained phrasal complexity may predict academic writing proficiency?

Methodology

Construction of the Corpus

The study data is based on three corpora of texts randomly selected from peer-reviewed scholarly journals, PhD dissertations, and MA theses (Table 3 and Table 4). The inclusion of the journals for selecting RAs in this study was based on the criterion of *h*-index and publication history. We choose *h*-index because the validity of traditional scientific indices such as journal impact factor (JIF) has been questioned, as the method for computation is not transparent and reproducible (García-Pérez, 2010) and is biased by high number of citations which may question the fairness of scientific achievement (Teixeira et al., 2013). *H*-index, on the other hand, is robust to over-citations, as it is not based on mean scores (Harzing & Van der Wal, 2008). It is simply defined as follows:

A scientist has index *h* if *h* of his/her N_p papers have at least *h* citations each, and the other (N_p - *h*) papers have ≤ *h* citations each, where N_p = the number of papers published over n years (Hirsch, 2005, p. 16569).

TABLE 3
Overview of Journals Included in the Corpus

Journal	Years of Publication	H factor
Language Learning	1948-1953, 1955-1956, 1958-ongoing	38
Applied Linguistics	1980-ongoing	38
TESOL Quarterly	1981-ongoing	36
Modern Language Journal	1916-1996, 1998-2001, 2005-ongoing	36
English for Specific Purposes	1980-1981, 1986-ongoing	25

The first sub-corpus consists of the abstract sections of 109 research articles (RAs). The second and third subcorpora included the abstract sections of 48 PhD dissertations and 80 MA theses compiled randomly from databases of Iranian state universities (See Table 4). All the texts followed IMRD (Introduction-Methods-Results-Discussion) format and were published between 2018 and 2020. The collection of recently published texts characterizes 'the present day' trends in academic writing (Biber & Gray, 2016). Only research studies representing empirical studies were included so that rhetorical and linguistic variations could be controlled for. The selection of the abstracts was based on stratified sampling technique. That is, texts were divided based on the strata of the journal/university types, years and issues of publication. Then, simple random sampling was used for each stratum. Students' texts were diverse in their focus, but we included only those texts that covered the same areas of study as the academic journals did (e.g., English education as in *TESOL Quarterly*, or studies targeted at specific discourse communities as in *English for Specific Purposes*). Accordingly, areas of studies like language testing, pragmatics, etc. were excluded from our study, as they do not conform to those of selected journals. We chose abstracts because of the following reasons: (a) they have manageable length, which makes cross comparisons easier than other sections, and (b) they are the first part that editors and reviewers read, and they determine whether readers keep reading the rest of the RA (Abrahamson, 2008). Accordingly, RA abstracts, as important part-genres, are getting increasing attention since they "foreground important claims, minimize methodology and background statements, and pack information into visuals" (Hyland, 2000, p. 86).

In this study, we follow Ansarifar et al. (2018), who assumed that the students in PhD group can have better academic expertise than their MA peers because of two reasons: Firstly, in Iran, PhD candidates are admitted based on a rigorous nationwide entrance examination and evaluation of their scientific backgrounds. Secondly, PhD dissertations have been written after four years of intensive study (p. 62).

TABLE 4
Descriptive Details of the Three Corpora

Corpora	Number of texts	Mean length of abstracts	Total number of words
MA	80	242.58	19406
PhD	48	398.67	19136
EW	109	177.69	19368

Note. MA = Master of Arts; PhD = Doctor of Philosophy; EW =Expert Writer

Syntactic Complexity Measures

In order to check the difference in the syntactic complexity of MA, PhD, and expert academic texts, the present study employed 5 large-grained measures of syntactic complexity, and 6 medium-grained measures of complexity. Large-grained syntactic complexity measures were selected from Lu's (2010) classification which includes 14 measures classified into five major categories of length of production, sentence complexity, subordination, coordination, and particular structures, as presented in Table 1 and explained in the literature review section. Lu's classification of the measures was "drawn from WolfeQuintero et al. (1998) and Ortega (2003) inter alia" (Kyle, 2016, p. 52). As Lu (2011) put it, they are comprehensive complexity indices commonly used in writing development (e.g., Bae & Min, 2020; Bailey & Judd, 2018; Kyle & Crossley, 2018; Lee, 2021; Liu & Li, 2016). The measures included in this study are mean length of T-unit (MLT), mean length of clause (MLC), clauses per sentences (C/S), dependent clauses per T-unit (DC/T), and complex nominals per T-unit (CN/T).

In deciding the large-grained measures of syntactic complexity, we considered the inclusion of those that could capture multidimensionality of the construct but at the same time address the issue of redundancy and distinctness (Norris & Ortega, 2009). As a result, based on Lu's (2010) classification of syntactic complexity measures which consists of five types of indices, we chose one measure from type 2, 3, and 5, and two measures from type 1 (See Table 1). The inclusion of two measures from type 1 was based on Norris and Ortega's (2009) distinctness argument that while it is superficially similar to the two other clausal measures, MLC must be considered a measure that taps into complexification at the phrasal level, as it can only increase through the addition of pre- or postmodification within a phrase or through nominalization (p. 561). Conversely, we did not include type 4, following Bardovi-Harlig's (1992) suggestion that coordination is a suitable measure for early stages of L2 development.

The medium-grained noun phrase measures of syntactic complexity consist of attributive adjectives, attributive adjectives and noun pre-modifiers, noun premodifiers, possessives, and post-modifying prepositional phrases. Since other features were infrequent in the corpus of the present study (with normalized frequency of less than 5 times per 1000 words), we did not examine them.

Automated Text Analysis Tools

Large-grained syntactic complexity analyses were performed with the aid of Syntactic Complexity Analyzer (SCA), which is a freely available online tool that computes 14 syntactic complexity indices (Lu, 2010). We choose SCA because of its free availability, inclusion of large number of measures (5 types of measures, each of which is assumed to capture one dimension of syntactic complexity), batch processing, and high reliability indices which are reported to range from 0.83 to 1.00 for structural unit identification, and 0.83 to 1.00 for correlation with human annotators (Lu & Ai, 2015). Initially, SCA uses Stanford parser to segment the text into individual sentences, which are then tokenized, and part-of-speech (POS) tagged (Liu & Li, 2016). Then using Tregex, the nod matches within a Tree are identified, and finally 14 measures of syntactic complexity are calculated.

Noun phrase modifiers, which were used to assess noun phrase complexity, were extracted in two phases. First, the texts were grammatically annotated through an automated text analysis tool called Stanford CoreNLP parser. It is a free text-analysis tool that requires Java to parse the text. The accuracy of POS tags is reported to be around 97% with trained data (Chen & Manning, 2014), which is state of art.

Then, a computer program written by Python was employed to count the frequency of 6 modifiers in the corpora already tagged by Stanford CoreNLP. Then, 10 percent of text was tagged and noun phrase modifiers were extracted manually by the two researchers, and the results were compared with those of automatically extracted ones. Overall, all of the modifiers had the accuracy rate of more than 90 percent with attributive adjective being the highest (97.35), and prepositional phrases being the lowest (90.78).

Fine-grained noun phrase measures were calculated through a computer program called Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC) which has been designed to measure a broad range of fine-grained indices of syntactic complexity sophistication and complexity in student writing (Kyle, 2016). Overall, TAASSC calculates 132 fine-grained measures of phrasal complexity which are categorized in four types. TAASSC uses Stanford dependency parser for annotating the text. Thus, the accuracy of TAASSC depends, in part, on the accuracy of Stanford parser. Unlike holistic measures, the measures included in TAASSC are function-based in a way that functional attributes of the noun phrases being modified are taken into account (e.g., number of dependents in nominal subjects).

Statistical Analyses

This study intended to investigate the extent to which large-grained and fine-grained syntactic complexity measures as predictor (independent) variables could predict academic writing proficiency as outcome (dependent) variable. We acknowledge that the three corpora of MA, PhD, and RA abstracts are, by no means, representative of the wide range of proficiency levels in academic writing in different contexts. The main objective of this study was to explore fine-grained, medium-grained and large-grained syntactic complexity measures for assessing abstract writing in higher education. The results may not be applicable to other assessment settings such as first-year undergraduate English-as-a-foreign-language (EFL) contexts.

In order to answer research question 1, we used SCA to calculate large-grained syntactic complexity measures. 5 syntactic complexity measures were compared across three groups of academic writers. Thus, one-way Multiple Analysis of Variance (MANOVA) was run which was followed by post-hoc pairwise comparisons for significant results. One-way ANOVA with post-hoc multiple comparisons were performed for comparing significant features identified in MANOVA. Bonferroni post-hoc correction was employed to adjust the *P*-values.

The second research question addressed the predictive power of fine-grained measures of phrasal complexity for three groups of academic writers. We performed ordinal regression by using 132 measures of phrasal complexity as independent variables to predict the academic writing proficiency (dependent variable), which was operationalized as the academic writing MA students, PhD students, and published expert RA writers produce. Such an operational definition of academic writing proficiency excludes high-school student writing, creative writing, and the types of writing practiced at BA programs in universities such as first-year student writing. These indices, which were taken from Kyle (2016), include indices for seven noun phrase types and ten phrasal dependent types. As Kyle further put it, these 132 measures can be squeezed into three main types: The first type of indices measures the number of dependents (modifiers of the head in the dependency grammar) per each phrase type (e.g., direct object); the second type calculates the frequency of occurrence of dependents regardless of noun phrase types in which they occur; and, finally, the third type calculates the average of dependents within noun phrase types.

Generalized linear model was used to obtain a more detailed and more accurate picture of the relationships between the predictor variables and the outcome (academic writers' proficiency levels). Statistical packages like SPSS use a model called proportional odds (PO) which is probably the most commonly used model for estimating ordinal outcome variables (Long & Freese, 2006). The PO model, in which the effect of each predictor variable on the categories of outcome variables is assumed the same, is used to predict the probability of being below, at, or above a particular level of an outcome variable (Liu & Koirala, 2012).

Results and Discussion

Results

Research question 1

Our first research question concerned the extent to which MA, PhD, and expert academic writers differed with regard to large-grained indices of syntactic complexity. The indices were chosen in a way that they could capture different dimensions of syntactic complexity as proposed by Lu (2010). The descriptive statistics are shown in Table 5, and Table 6. Expert writers (EWs) outperformed PhD writers (PhDWs), and PhDWs outperformed MA writers (MAWs) in all of the measures except for Adjective + noun premodifiers, and possessives.

TABLE 5
Descriptive Statistics of Large-Grained Syntactic Complexity Measures across Proficiency Levels

Measure	Group	Number of texts	Mean	Std. Deviation
Mean length of T-unit (MLT)	EW	109	29.00	5.42
	PhD	48	26.68	4.87
	MA	80	23.65	4.42
Mean length of clause (MLC)	EW	109	19.27	5.62
	PhD	48	15.97	4.44
	MA	80	13.54	3.18
Clauses per sentences (C/S)	EW	109	1.79	0.38
	PhD	48	1.63	0.29
	MA	80	1.66	0.29
Dependent clauses per T-unit (DC/T)	EW	109	0.57	0.33
	PhD	48	0.51	0.25
	MA	80	0.47	0.21
Complex nominals per T-unit (CN/T)	EW	109	4.92	1.24
	PhD	48	3.92	0.95
	MA	80	2.97	1.07

TABLE 6
Descriptive Statistics of Medium-grained Noun Phrase Complexity Measures across Proficiency Levels

Measure	Group	Number of texts	Mean	Std. Deviation
Adjectives	EW	109	72.57	23.70
	PhD	48	71.80	20.66
	MA	80	69.48	22.91
Noun premodifiers	EW	109	45.99	27.14
	PhD	48	43.32	19.16
	MA	80	25.10	18.50
Adjectives + noun	EW	109	15.54	10.25
	PhD	48	16.07	12.94
	MA	80	15.22	10.70
Possessives	EW	109	7.03	9.18
	PhD	48	10.14	9.28
	MA	80	9.92	10.10
Prepositions	EW	109	85.22	25.70
	PhD	48	72.15	15.42
	MA	80	69.49	21.30

A one-way between-groups multivariate analysis of variance (MANOVA) was performed to investigate academic proficiency differences in large-grained, medium-grained and fine-grained noun phrase syntactic complexity measures. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted. There was a statistically significant difference

among the academic writers on the combined dependent variables, $F(20, 422) = 9.559$, $p = 0.000$; Pillai's Trace = 0.624; partial eta squared = 0.312. As shown in Table 7, using a Bonferroni adjusted alpha level of 0.005, MLT ($p = 0.000$), MLC ($p = 0.000$), CN/T ($p = 0.000$), noun premodifiers ($p = 0.000$), and prepositional phrases ($p = 0.000$) reached statistical significance.

TABLE 7
Tests of Between-Subjects Effects

Dependent Variables	df	Mean Square	*F*	Sig.	Partial Eta Squared
MLT	2	623.33	22.25	0.00*	0.17
MLC	2	727.68	33.48	0.00*	0.23
C/S	2	0.50	4.52	0.01	0.04
DC/T	2	0.20	2.64	0.07	0.02
CN/T	2	83.12	65.48	0.00*	0.37
Adjective premodifier	2	212.54	0.41	0.67	0.00
Noun premodifier	2	9230.70	21.33	0.00*	0.16
Adjective + Noun	2	10.44	0.09	0.92	0.00
Possessives	2	243.69	2.68	0.07	0.02
Prepositions	2	6075.28	12.06	0.00*	0.10

* The results are significant at the $p < 0.005$ level.

The results for the significant dependent variables were considered separately by employing ANOVA and post-hoc Tukey HSD test (Table 8). According to the results obtained, MAWs used statistically shorter T-units than PhDWs ($p = 0.007$), and EWs ($p = 0.000$). In comparison, PhDWs employed shorter T-units than EWs ($p = 0.041$). Concerning MLC, low proficiency writers produced lower length of clauses compared to high proficiency writers. The difference between EWs and other groups reached significance at ($p = 0.000$), and the difference between PhDWs and MAWs was significant at ($p = 0.016$). As for CN/T, the differences between lower and higher proficiency groups were significant at ($p = 0.000$). Concerning medium-grained noun-phrases measures, MA abstracts were characterized by significantly less normalized frequency of noun premodifiers than those of PhD ($p = 0.000$), and EWs ($p = 0.000$). EWs used significantly more prepositional phrases for constructing abstracts than both MAWs ($p = 0.000$) and PhDWs ($p = 0.004$). However, the difference between MAWs and PhDWs was not statistically significant ($p = 0.803$).

TABLE 8
Pairwise Tukey HSD Comparisons of Syntactic Complexity Measures Academic Writers' Groups

Language Measure	Writer Groups		Mean Difference	Sig.
MLT	EW	PhD	2.32*	0.041
		MA	5.35*	0.000
	PhD	EW	-2.32*	0.041
		MA	3.03*	0.007
	MA	EW	-5.35*	0.000
		PhD	-3.03*	0.007
MLC	EW	PhD	3.30*	.000
		MA	5.74*	.000
	PhD	EW	-3.3*	.000
		MA	2.44*	.016
	MA	EW	-5.74*	.000
		PhD	-2.44*	.016
CN/T	EW	PhD	0.99*	0.000
		MA	1.95*	0.000
	PhD	EW	-0.99*	0.000
		MA	0.96*	0.000
	MA	EW	-1.95*	0.000
		PhD	-0.96*	0.000
Noun premodifier	EW	PhD	2.67	0.755
		MA	20.88*	0.000
	PhD	EW	-2.67	0.755
		MA	18.22*	0.000
	MA	EW	-20.88*	0.000
		PhD	-18.22*	0.000
Prepositional phrases	EW	PhD	13.07*	.004
		MA	15.73*	.000
	PhD	EW	-13.07*	.004
		MA	2.66*	.803
	MA	EW	-15.73*	.000
		PhD	-2.66*	.803

* The results are significant at the $p < 0.05$ level.

Research question 2

Given that one of the purposes of the current study was to explore syntactic complexity measures in terms of their predictive power and descriptive adequacy, the second phase examines fine-grained measures of syntactic complexity. Complexity in the abstract sections of MA theses, PhD dissertations, and research articles was investigated using 132 measures of phrasal complexity proposed by Kyle (2016). In order to accomplish this purpose, fine-grained measures were entered into ordinal logistic regression model to determine which of the independent variables (if any) could predict the dependent variable (proficiency level of academic writers).

In order to obtain a more detailed and more accurate picture of the relationships between the predictor variables and the outcome, the researchers employed generalized linear model (GLM) in SPSS. GLM assumes that any observed response is a linear sum of multiple individual underlying responses (Fagerland & Hosmer, 2016). Independent variables could be both categorical and continuous. The dependent variables do not need to be normally distributed, but they need to belong to the exponential family (binominal, multinominal, Poisson, ….). One of the prominent features of GLM is that all of the inference tools that we use for model checking in loglinear and logistic regression models apply for GLM too, e.g., Wald and Likelihood ratio tests, Deviance, Residuals, Confidence intervals, Overdispersion (Agresti, 2018). In sum, compared to other methods of assessing the predicative ability of a set of predictor variables with regard to the outcome which is ordered dependent variable in our study (e.g., Multinominal logistic regression, and Ordinal logistic regression), running ordinal regression by GLM poses the advantage of generating both Wald test and Likelihood ratio chi-square test, and also producing the odd ratios.

Having checked the assumptions of ordered dependent variables, multicollinearity, the presence of at least one continuous, categorical, or ordinal variable, and proportional odds, we finally came up with 14 variables. The variables were entered into SPSS. As shown in Table 9, the value obtained by the Likelihood Ratio Chi-Square test indicates that our full model was a significant improvement in fit over the null (no predictors) model ($\chi^2(14) = 66.166$, $p < 0.001$). This simply means that prediction of the outcome variable based on the predictors of the current study is significantly more accurate than frequency-based prediction without taking predictor variables into account.

TABLE 9
Omnibus Test of Phrasal Complexity Measures

Likelihood Ratio Chi-Square	df	Sig.
66.17	14.00	0.00

According to Table 10, out the of 14 phrasal complexity measures that were entered into the model, four measures made unique statistically significant contribution to the model (dependents per nominal, dependents per nominal subjects, dependents per direct object, and dependents per direct object (Std.)). The EXP(B) values indicated that the strongest predictor of academic writing proficiency levels was dependents per nominal, which records an odds ratio of 627.27. This means that every one unit increase on this predictor variable increases the odds of being in a higher category of proficiency writing by a factor of 627.27. The second strongest predictor variable is dependents per direct object (Std.) with the EXP (B) value of 2.94 followed by dependents per nominal subject, and dependents per direct object, respectively.

TABLE 10
Tests of Model Effects and Parameter Estimates of Phrasal Complexity Measures

Predictor variables			Type III		
	B	Exp(B)	Wald Chi-Square	df	Sig.
Dependents per nominal	6.44	627.27	13.98	1	0.00
Dependents per nominal subjects	0.94	2.56	6.04	1	0.01
Dependents per direct objects (Std.)	1.08	2.94	4.58	1	0.03
Dependents per direct object	1.29	0.27	0.08	1	0.00

Discussion

In response to the first research question regarding the difference among expert, PhD, and MA academic writers in terms of syntactic complexity as measured by large-grained and medium grained indices, the present study employed one-way MANOVA to examine if significant differences existed among the texts produced by three writer groups. The results showed that sentence-based, and subordination-based measures of syntactic complexity could not capture complexity across three levels of academic writing proficiency. On the other hand, length-based (MLTU), phrasal-based (MLC and CN/T), and medium-grained measures (noun premodifiers) of syntactic complexity could discriminate between adjacent and non-adjacent levels (with the exception of prepositional phrases that could not discriminate between adjacent levels of PhD and MA academic writers). Lack of predictability of sentence, and subordination measures of syntactic complexity does not suggest that academic texts are not complex with regard to these structures; rather, it implies that the writers of academic texts "do not produce such structures more or less frequently" (Casal & Lee, 2019, p. 59).

The results obtained from length-based syntactic complexity measures conforms to those of Youn (2014), where MLTU distinguished between "three different levels of proficiency and pragmatic performance well" (p. 278), although the participants in Youn's study were all undergraduate students enrolled in a four-year university course. While T-unit measures have provided mixed results in previous

research, with some studies showing no association between large-grained T-unit measures such as MLTU and L2 syntactic growth and some others exhibiting strong correlation (Crossley & McNamara, 2014), this study lends support to the claim that high proficiency writers produce longer T-units. In the same vein, the results of this study corroborate previous findings in that academic writers progressively use more complex nominals per T-units, and greater mean length of clauses (Casal & Lee, 2019), and specific phrasal modification features of noun pre-modifiers and prepositional phrases (Ansarifar et al., 2018; Parkinson & Musgrave, 2014).

The fact the current study found no link between academic writing proficiency and clausal measures of C/S and DC/T is not surprising given that less proficient groups in our study were graduate academic writers who had already mastered the principles of undergraduate academic writing, inter alia, basic clause structures.

While the primary purpose of the current study was not to trace Biber et al.'s (2011) proposed trajectories of writing development, our findings are consistent with the argument by Biber et al. that finite dependent clauses are expected to occur at early stages of writing development and move towards final stages of dependent phrases. Our study revealed that noun premodifiers distinguished low proficiency (MAWs) and higher proficiency groups (PhDWs and EWs). In line with the findings of Casal and Lee (2019), we also found reliance on prepositional phrases made distinctions between less proficient and more proficient academic group. This clearly supports Biber et al.'s (2011) argument that as academic writers become more advanced, their writing is characterized by more occurrences of noun phrase post-modifiers. However, increase in noun phrase modifiers in our corpus did not occur at the cost of decrease in clausal complexity as suggested by some studies (e.g., Byrnes et al., 2010, as cited in Casal & Lee, 2019, p. 59).

The second research question concerned the predictability of fine-grained measures of syntactic complexity across EWs, PhDWs, and MAWs. By drawing on generalized linear model, our study has also revealed that fine-grained indices of 'dependents per nominal', 'dependents per nominal subjects', 'dependents per direct object', and 'dependents per direct object (Std.)' could significantly predict academic writers' proficiency levels. Generally, these indices suggest that higher proficiency academic writers tend to use more dependents in their writings which strongly supports Biber et al's. (2011) argument that advanced academic writing is characterized by more noun-modifying features. Our study also found that expert academic writers use a wider range of dependents per direct object, an interesting finding about functional attributes of noun phrase modifiers as well as their diversity that Biber et al.'s (2011) proposed measures of writing proficiency did not take into account.

The other finding of this study is that expert academic research article writers used far larger dependents per nominal in a way that for every one unit of increase in this measure, there is a predicted increase of 6.44 in the log odds of a writer being at a higher level of writing proficiency level. This is an important finding because it suggests that the average number of dependents (modifiers) that each nominal contains is in a strong and direct correlation with the complexity of a text. This finding supports Biber et al.'s (2011) hypothesis that as learners become more proficient, their writing is characterized by greater amount of meaning embedded in phrases.

The number of dependents per nominal subject also contributes to the complexity of a text. That simply means that in order for student writers to be more academic, they need to include less bare nominal subjects and more structures with multiple dependents in the subjective position. This is in line with Biber et al.'s (2011) claim that as writers become more advanced, their writings represent more features of academic writing.

An important finding of our study is that as writers become more proficient, they tend to use more and a wider range of dependents per direct objects. The functional property of noun phrase modifiers and their range of distribution were not included in Biber et al's. (2011) hypothesis. The measure provides us with more detailed information regarding the distribution of noun phrase modifiers in academic writing, because the functional attributes of noun phrases (e.g., accusative noun phrases in this measure) are

reported. This is in line with Kyle's (2016) study, although, unlike the current study, Kyle did not deal with the abstract sections.

Conclusion

This study has examined large-grained, medium-grained, and fine-grained measures of syntactic complexity for assessing writing complexity of research article abstracts across three proficiency tiers. The corpora in our study fell into three categories of EWs, PhDWs, and MAWs. Drawing on the suggestion by Biber et al. (2011) that progress in writing development involves complexification at clausal levels at early stages, and then continues into intermediate stages of nonfinite dependent clauses and moves towards increased noun phrase complexity, the present study intended to explore the extent to which the three proficiency tiers differ in each of the measures, and to investigate the predictive ability of fine-grained noun phrase complexity measures.

In summary, our findings provide further evidence to support the argument that academic writing complexity moves from clausal embedding to phrasal embedding, to the extent that clausal measures of C/S, and DC/T did not show the distinctions in proficiency levels suggesting that student writers in our study (PhDWs and MAWs) may have already mastered clause structures. By contrast, large-grained measures of syntactic complexity (MLC, CN/T, noun premodifiers, and prepositional phrases) detected proficiency differences for academic writers. In addition, the results of ordinal logistic regression indicated that fine-grained measures of phrasal complexity had a predictive ability of 67.17 with the strongest predictor being dependents per nominal which reported an odd ratio of 627.27.

These findings would suggest medium-grained and large-grained length-based measures of syntactic complexity offer more progress-sensitive account of academic writing development than large-grained clausal measures. Accordingly, owing to the fact that complexity is a multi-dimensional construct, large-grained length-based measures of syntactic complexity particularly MLTU, MLC and CN/T, as this study and other studies showed, could complement specific measures such as noun premodifiers, and prepositional phrases.

Fine-grained measures, as our study showed, were highly predictive of academic writing proficiency. Given the predictive strengths of fine-grained measures of syntactic, we can safely argue that academic writing assessment could benefit from the measures that provide "structural/ syntactic/functional distinctions" (Biber et al., 2020, p. 13). That is to say, fine-grained measures provide the researchers with detailed and interpretable account of academic wiring complexity.

The arguments above highlight the importance of length-based measures of syntactic complexity in academic writing instruction and assessment, as this study supported the assumption that longer structures are more complex (longer T-units, longer clauses, and phrasal structures). Although the use of long noun phrases is the hallmark of advanced academic writing (Ortega, 2012), more advanced academic writers in our study have complexified the texts through the large-grained structures too. That being said, both specific language structures and large-grained language structures contributed to the complexification of academic texts in the present study. One particular type of specific language structures is noun phrase modification features such as noun premodifiers, prepositional phrases, and dependents in a particular unit of analysis (e.g., number of dependents per nominals, or number of nominals per direct object). As for large-grained features like MLTU, and CN/T, there is a variety of grammatical choices that can potentially make a longer T-unit, including embedded phrases, embedded clauses, extra adjectives, or adverbs (Biber et al., 2011). Thus, familiarity with these devices may enhance the learners' academic literacy such as better appreciation of academic texts, and more professional research reports (Lan & Sun, 2019).

This study has found that advanced academic writing benefits from "increased repertoire of complex structures and functional capacities of particular complex structures" (Casal & Lee, 2019, p. 60). The predictability of complex academic language features in general and length-based measures of complexity

in particular suggest that the explicit instruction of these features is most likely to help graduate academic writers develop a more complex writing style. For example, consciousness raising tasks can be designed to draw students' attention to the saliency of noun phrase modifiers in academic writing and to their syntactic roles (e.g., nominal subjects vs. direct objects). Furthermore, students' writing can be compared to those of advanced writers in terms of prominent syntactic features (e.g., mean length of T-unit, dependents per nominals, etc.), and any loosely-organized syntactic structures could be collaboratively adjusted in order to meet the requirements of expert academic writing.

Limitations and Future Research

While this study has examined syntactic complexity in academic writing by utilizing both fine-grained and large-grained measures, there were some limitations that should be taken into consideration. First, due to labor-intensive, and time-consuming process of qualitative check, which was part of the analysis, a relatively small corpus of academic RA abstracts was examined. Thus, future studies might compile larger corpora that could offer a deeper insight into the way the differences among academic writers with varying levels of proficiency are realized. Also, this study compared noun phrase modifiers only across three groups of MA, PhD and RA academic writers. It might be an interesting area of research for future studies to investigate large-grained and fine-grained syntactic complexity measures across a wide range of academic tiers.

The Authors

Rajab Esfandiari is an associate professor in the Department of English Language, Faculty of Humanities at Imam Khomeini International University, in Qazvin, Iran. His areas of interest and specialty include teaching and assessing L2 writing, multifaceted Rasch measurement, L2 classroom assessment, and EAP teaching and testing.

Department of English Language, Faculty of Humanities
Imam Khomeini International University, Qazvin, Iran
Tel: +98-912-887-7274
Email: esfandiari@hum.ikiu.ac.ir

Mohammad Ahmadi is a visiting professor in the Department of English Language, Faculty of Humanities at Imam Khomeini International University, in Qazvin, Iran. His research interest areas include corpus linguistics, second language writing, computational analysis of natural language, and formulaic language.

Department of English Language, Faculty of Humanities
Imam Khomeini International University, Qazvin, Iran
Tel: +98-912-394-9371
Email: ahmadi.m8362@gmail.com

References

Abrahamson, E. (2008). 22 things I hate: Mini rants on management research. *Journal of Management Inquiry, 17*(4), 422-425.

Agresti, A. (2018). *An introduction to categorical data analysis*. John Wiley & Sons.

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*(1), 58-71.

Bae, J., & Min, S. (2020). Genre-based analysis of syntactic complexity in l2 college students' writing: pedagogic scope and directions. *The Journal of Asia TEFL, 17*(3), 937-953.

Bailey, D. R., & Judd, C. (2018). The effects of online collaborative writing and TOEIC writing test-preparation on L2 writing performance. *The Journal of Asia TEFL, 15*(2), 383-397.

Bardovi-Harlig, K. (1992). The relationship of form and meaning: A cross-sectional study of tense and aspect in the interlanguage of learners of English as a second language. *Applied Psycholinguistics, 13*(3), 253-278.

Bi, P., & Jiang, J. (2020). Syntactic complexity in assessing young adolescent EFL learners' writings: Syntactic elaboration and diversity. *System, 91*, 1-10.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9*(1), 2-20.

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. 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., & Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. *Journal of English for Academic* Purposes, *46*, 1-15. https://doi.org/10.1016/j.jeap.2020.100869

Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). John Benjamins Publishing Company.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*(1), 42-65.

Bulté, B., & Housen, A. (2018). Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International Journal of Applied Linguistics, 28*(1), 147-164.

Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment. *Modern Language Journal, 94*(supplement 1), 1-235.

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

Chen, D., & Manning, C. D. (2014, October 26-28). A fast and accurate dependency parser using neural networks [Paper presentation]. The 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Qatar. https://www.emnlp2014.org/papers/pdf/EMNLP2014082.pdf

Connor-Linton, J., & Amoroso, L. (Eds.). (2014). *Measured language: Quantitative studies of acquisition, assessment, and variation*. Georgetown University Press.

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*(1), 66-79.

Fagerland, M. W., & Hosmer, D. W. (2016). Tests for goodness of fit in ordinal logistic regression models. *Journal of Statistical Computation and Simulation, 86*(17), 3398-3418.

García-Pérez, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in

Psychology. *Journal of the American Society for Information Science and Technology, 61*(10), 2070-2085.

Halliday, M. A. (1985). Systemic background. *Systemic Perspectives on Discourse, 1*, 1-15.

Harzing, A. W. K., & Van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics, 8*(1), 61-73.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences, 102*(46), 16569-16572.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461-473.

Housen, A., De Clercq, B., Kuiken, F., Vedder, I. (2019). Multiple approaches to complexity in second language research. *Second Language Research, 35*(10), 3-21.

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing.

Hyland, K. (2000). Hedges, boosters and lexical invisibility: Noticing modifiers in academic texts. *Language Awareness, 9*(4), 179-197.

Iwashita, N. (2006). Syntactic complexity measures and their relation to oral proficiency in Japanese as a foreign language. *Language Assessment Quarterly: An International Journal, 3*(2), 151-169.

Jiang, F. K., & Hyland, K. (2017). Metadiscursive nouns: Interaction and cohesion in abstract moves. *English for Specific Purposes, 46*, 1-14.

Jiang, J., Bi, P., & Liu, H. (2019). Syntactic complexity development in the writings of EFL learners: Insights from a dependency syntactically-annotated corpus. *Journal of Second Language Writing, 46*, 1-13.

Kreyer, R., & Schaub, S. (2018). The development of phrasal complexity in German intermediate learners of English. *International Journal of Learner Corpus Research, 4*(1), 82-111.

Kyle, K (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic complexity sophistication* [Unpublished doctoral dissertation]. Georgia State University.

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., & Sun, Y. (2019). A corpus-based investigation of noun phrase complexity in the L2 writings of a first-year composition course. *Journal of English for Academic Purposes, 38*, 14-24.

Lee, J. (2021). Syntactic complexity, clausal complexity, and phrasal complexity in L2 writing: The effects of task complexity and task closure. *The Journal of Asia TEFL, 18*(1), 108-124.

Lei, L., & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics, 23*(2), 216-243.

Liu, L., & Li, L. (2016). Noun phrase complexity in EFL academic writing: A corpus-based study of postgraduate academic writing. *The Journal of Asia TEFL, 13*(1), 48-65.

Liu, X., & Koirala, H. (2012). Ordinal regression analysis: Using generalized ordinal logistic regression models to estimate educational data. *Journal of Modern Applied Statistical Methods, 11*(1), 242-254.

Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata*. Stata Press.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474-496.

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.

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*(1), 16-27.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578.

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.

Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann & B. Szmrecsanyi (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127-155). de Gruyter.

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*(1), 48-59.

Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First Language, 30*(1), 3-26.

Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics*. Routledge.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17*(1), 38-62.

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.

Swan, M., & Baker, D. (2008). *Practical English usage: International students' edition*. Oxford University Press.

Teixeira, M. C., Thomaz, S. M., Michelan, T. S., Mormul, R. P., Meurer, T., Fasolli, J. V. B., & Silveira, M. J. (2013). Incorrect citations give unfair credit to review authors in ecology journals. *PLoS One, 8*(2), 1-4.

Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal, 96*(4), 576-598.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. University of Hawaii at Manoa.

Youn, S. J. (2014). Measuring syntactic complexity in L2 pragmatic production: Investigating relationships among pragmatics, grammar, and proficiency. *System, 42*, 270-287.

(Received May 20, 2021; Revised August 20, 2021; Accepted September 18, 2021)

