CSCanada
Cross-Cultural Communication
Vol. 18, No. 4, 2022, pp. 78-87
DOI:10.3968/12889

ISSN 1712-8358[Print]
ISSN 1923-6700[Online]
www.cscanada.net
www.cscanada.org

A Comparison of Measurements of Syntactic Complexity in L2 Writing: Large-Grained Indices and Fine-Grained Clausal and Phrasal Indices

HUANG Qian[a],*

[a] Zhijiang College, Zhejiang University of Technology, Shaoxing, China.
*Corresponding author.

Received 21 August 2022; accepted 28 September 2022
Published online 26 December 2022

Abstract

As an important measure of second language (L2) writing, syntactic complexity has always been a significant concept in L2 writing research and pedagogy. Syntactic complexity has traditionally been manipulated through large-grained indices. Although large-grained indices are regarded as the most consistent measurement in L2 writing studies, its widespread application has been criticized recently mainly because of the difficulty in its interpretation and a potentially misaligned focus on clausal complexity like clausal subordination. In this article, the researcher aims to adopt the large-grained indices of syntactic complexity, fine-grained indices of clausal complexity, and fine-grained indices of phrasal complexity to predict the holistic quality of independent expository essays written by Chinese college students. Results revealed that fine-grained indices of phrasal complexity were better indices to predict the overall writing quality than either large-grained or fine-grained clausal indices. These results tally with Kyle and Crossley's (2018) claims with regard to the validity of syntactic complexity indices in predicting writing quality and also provide stronger support for the superiority of fine-grained indices of phrasal complexity.

Key words: Syntactic complexity; L2 writing; Large-grained indices; Fine-grained indices

Huang, Q. (2022). A Comparison of Measurements of Syntactic Complexity in L2 Writing: Large-Grained Indices and Fine-Grained Clausal and Phrasal Indices. *Cross-Cultural Communication, 18*(4), 78-87. Available from: http://www.cscanada.net/index.php/ccc/article/view/12889
DOI: http://dx.doi.org/10.3968/12889

1. INTRODUCTION

In the context of innovations in writing pedagogy and assessment, computer-based automated writing assessment techniques have received a lot of attention. These techniques employ natural language processing (NLP) tools to flag linguistic, syntactic, semantic, or rhetorical features associated with writing quality, and uses statistical tools or machine learning algorithms to generate scores and feedback based on patterns observed in these features (Wilson & Rod, 2019). The integration of modern information technology represented by computers and the Internet is gradually boosting the trend of language teaching modernization (Jiang Xueqing et al., 2011).

Among language complexity indices measured by the computer automatic scoring system, syntactic complexity indices have received far more attention than lexical complexity and morphological complexity (Kuiken et al., 2019a), but at are also more controversial (Liu Ligang & Miao Haitao, 2018). Syntactic complexity is often measured by large-grained indices such as the mean length of T-units and the ratio of embedded structures. However, in recent years, more and more research have focued on syntactic complexity at the clause and phrase levels (Kuiken et al., 2019a), because researchers found that large-grained indices lack interpretation power (Bardovi-Harlig, 1992; Norris & Ortega, 2009; Biber et al., 2011). This study aims to use the latest syntactic complexity software to analyze English academic writings by Chinese college students, compare the predictive effects of various indices, and find indices suitable for measuring the writing output of Chinese English learners.

2. RESEARCH BACKGROUND

Syntactic and linguistic complexity are widely studied in applied linguistics as indicators of language proficiency development and level (De Clercq & Housen, 2017).

Syntactic complexity refers to the variety and complexity of syntactic forms in language production (Ortega, 2003; Lu, 2011), which is an important concept in second language writing teaching and research, because syntactic structure growth is a necessary step in the process of learners' target language development. Measurements of syntactic complexity include quantified indices at multiple dimensions, including the length of production unit, the number of subordinate or parallel structures, the variety and complexity of syntactic structures, etc. (Norris & Ortega, 2009).

2.1 Large-Grained Indices of Syntactic Complexity

Large-grained indices of syntactic complexity, the traditional syntactic complexity indices, have been widely used to measure language production and development. The most commonly used large-grained indices are based on unit lengths (e.g., mean length of sentences, mean T-units, and clauses), calculated by dividing the total number of words by a chosen linguistic unit (Norris & Ortega, 2009), or by calculating the proportion of embedded structures (e.g., subordinate clauses, parallel structures or particular types of phrases) in T-units, clauses, and sentences (Lu Xiaofei & Xu Qi, 2016). Take the Mean Length of Sentence (MLS) as an example. The sentence *It seemed that the little girl failed in the exam because she ran out of the classroom with tears on her face* contains three clauses: a main clause, a *that*-clause and a *because*-clause. Among them, the main clause has two words, the *that*-clause has eight words, and the *because*-clause has twelve words. Therefore, the average clause length of this sentence is 7.33 (22/3).

Large-grained syntactic complexity indices are widely used. In Ortega's (2003) synthesis study on writing, a total of twenty-seven studies were counted, sixteen of which only used the mean length of T-units (MLT) to measure syntactic complexity. Even in recent international studies (e.g. Hwang et al., 2020; Marcus et al., 2020) and domestic studies (e.g., Bao, 2009; Chen, 2010; Lu, 2011; Xu et al., 2013; Li & Liu, 2016; Lei, 2017), MLT was also used as an important indicator of language capacity and development. The basic assumption of the large-grained indices is that the longer the structure, the more complex it is, but they tend to ignore the internal characteristics of these structures, therefore lacks deep interpretation power (Liu Ligang & Miao Haitao, 2018). Consider the following sentences:

(a) *The beautiful girl in red won the competition for her class*.

(b) *The girl was very happy, because she has won the competition*.

Both sentences contain eleven words and a T-unit, but the internal linguistic structure is very different. In (a), the subject is modified by an adjective (*beautiful*) and a prepositional phrase (*in red*); the predicate verb is modified by a prepositional phrase (*for her class*). Sentence (b) contains a *because*-dependent clause. These two example sentences show that MLT can only provide general information on syntactic complexity, but cannot explain the specific composition of language structures within T-units.

2.2 Fine-Grained Indices of Clausal Complexity

Fine-grained indices of clausal complexity measure the average clausal dependent types or structures in clauses (Kyle & Crossley, 2018). Clausal complexity metrics contains those that measure the amount of subordination. They are computed by counting all clauses and dividing them over a given production unit of choice (i.e., mean number of clauses per T-unit or C-unit). Take (a)(b) as an example, sentence (a) contains one clause, and sentence (b) contains two clauses. From the perspective of internal dependent structures, each clause in sentence (a) contains a prepositional modifier (1/1) that modifies the verb, whereas each clause in sentence (b) contains 0.5 adjective predicative (1/2) and 0.5 adverbial clauses (1/2).

Although clausal complexity indices may solve the opacity of large-grained indices such as MLT and MLC to a certain extent, the excessive use of the clausal complexity indices in the study of second language writing has also been questioned. Biber et al. (2011) compared twenty-eight fine-grained lexical-grammatical features of two different genres in a corpus of conversational texts and academic writings. It was found that the clausal complexity features are more prominent in dialogues than in academic writings, because academic papers contain more complex phrasal components (such as noun phrases). For example, sentences (c) and (d) share the same T-unit length, but there are four dependent clauses in sentence (c) (excerpted from a dialogue) and no dependent clauses in sentence (d) (excepted from an academic article). This suggests that clause complexity indices may bias towards formal styles.

(c) *Well, since he got so upset, I just didn't think we would want to wait for Tina to come back.*

(d) *This may be part of the reason for the statistical link between schizophrenia and membership in the lower socioeconomic classes*. (Cited from Biber et al., 2011)

2.3 Fine-Grained Indices of Phrasal Complexity

Fine-grained indices of phrase complexity focus on the complexity at the phrasal level, and are calculated at three dimensions: frequency of dependent items, number of dependent items per noun phrase, and frequency of dependent items per noun phrase. Take the following sentence for example: *The man in the blue shoes bought the little girl food*. There are four noun phrases in this sentence: noun subject, prepositional object, indirect object, and direct object. In addition, there are six phrasal dependents: three determiners, two adjective modifiers and a prepositional phrase modifier. Therefore, each noun

phrase contains 0.75 dependent items with determiners, 0.5 dependent items with adjective modifiers, and 0.25 with prepositional phrases.

Studies have found inconsistent degree of effectiveness for different syntactic complexity metrics when judging texts of different genres (Biber et al., 2011; Polio & Yoon, 2018). Norris & Ortega (2009) noted that beginners typically use more parallel structures, intermediate learners use more dependent clauses, and advanced learners use more complex phrases—a process that mirrors the development from fundamental communication skills to advanced academic writing skills. Taguchi et al. (2013) analyzed the complexity of clauses and phrases of L2 compositions of different scores, and found that compared with articles with lower scores, articles with higher scores contained more features of phrase complexity (e.g., more adjective modifiers, prepositional phrase modifiers, etc.) and fewer clause complexity features (e.g., subordinating conjunctions, relative clauses). Kyle & Crossley (2018) analyzed a corpus of TOEFL composition and proposed that phrasal complexity indices may be superior to clausal complexity indices and the traditional large-grained syntactic complexity indices in predicting writing quality.

In summary, phrasal complexity indices highlight the dynamics and effectiveness at the phrase level, and are more sensitive than clausal complexity in judging the complexity of syntactic structures at different levels and of different styles. Therefore, it may have a greater advantage at predicting the quality of L2 writing output.

At present, there are only a handful of studies on the syntactic complexity of Chinese learners' writing (Lei, 2017; Gao, 2021), and the existing research mainly adopts large-grained syntactic complexity indices. For instance, Gao Xia (2021) used the phrasal complexity indices to analyze the writing output of Chinese second language learners for the first time. The corpus included texts of as many as 20 styles and 304 topics written by middle school and college students. However, its findings failed to confirm the importance of phrasal complexity indices in academic writing and L2 writing. To address this gap, the present study will use large-grained syntactic complexity indices, and fine-grained clausal and phrasal indices to analyze the writing quality of Chinese college students' essays on a single topic, and compare the role of various complexity indices in predicting writing outcomes. It mainly addresses the following research questions:

RQ1: What is the relationship between traditional indices of syntactic complexity and holistic writing scores?

RQ2: What is the relationship between fine-grained indices of clausal complexity and holistic writing scores?

RQ3: What is the relationship between fine-grained indices of phrasal complexity and holistic writing scores?

RQ4: Do fine-grained indices of phrasal complexity predict writing scores more accurately?

3. RESEARCH METHODS

3.1 Learner Corpus

This study collects 120 argumentative explanatory essays written by sophomores in a university in Zhejiang Province for analysis. The title of the article is: *A Ten-Year Development Comparison of KFC and McDonald's in the Chinese Market*. It is an expository writing with charts. The required number of words for each essay is 150-180, and the writing time is 30 minutes. Each article is scored by three professional raters on a 100-point scale, and the final score is obtained by averaging the three scores. Articles with large differences in individual scoring results were discussed among the raters and the initial scores were then adjusted. The scoring criteria consist of three components: language quality (40%), content (30%) and organization (30%), and are weighted in a ratio of 4:3:3. The total number of words in the corpus is 19,306 words, and the average score of all articles is 70.7 (N=120, SD=5.83)

3.2 Research Tools

Measuring large-grained indices. To work out large-grained indices of syntactic complexity, Lu's (2010) L2 syntactic complexity analyzer (L2SCA) indices was adopted. The L2SCA system generates indices of syntactic complexity on fourteen measures. The structure types measured by L2SCA statistics are words, verb phrases, complex nominals, coordinate phrases, clauses, dependent clauses, T-units, complex T-units and sentences. See Table 1 for a list of the fourteen L2SCA indices and a description of each index.

Table 1
Description of SCA Variables

Index abbreviation	Index name	Index description
MLS	mean length of sentence	number of words per sentence
MLT	mean length of T-unit	number of words per T-unit
MLC	mean length of clause	number of words per clause
C/S	clauses per sentence	number of clauses per sentence
VP/T	verb phrases per T-unit	number of verb phrases per sentence
C/T	clauses per T-unit	number of clauses per T-unit
DC/C	dependent clauses per clause	number of dependent clauses per clause
DC/T	dependent clauses per T-unit	number of dependent clauses per T-unit
T/S	T-units per sentence	number of T-units per sentence
CT/T	complex T-unit ratio	number of complex T-units divided by T-units
CP/T	coordinate phrases per T-unit	number of coordinate phrases per T-unit
CP/C	coordinate phrases per clause	number of coordinate phrases per clause
CN/T	complex nominals per T-unit	number of complex nominals per T-unit
CN/C	complex nominals per clause	number of complex nominals per clause

Measuring fine-grained clausal indices. Fine-grained indices of clausal complexity were measured by the free text analysis tool TAASSC (Kyle, 2016; Kyle & Crossley, 2018). There are 31 clausal complexity indices in TAASSC. Twenty-nine of these indices measure the average number of specific constructs in each clause. Kyle (2016) pointed out that there are three main differences between the clausal complexity indices included in TAASSC and those included in L2SCA (i.e., MLC, DC/C, CP/C, and CN/C). First, TAASSC calculates the length of clauses as the number of direct dependents each clause contains, rather than the number of words, which prevents structures containing more words (e.g., prepositional phrases) from getting more weight. Second, TAASSC counts each clause type separately, rather than combining structures such as subordinate clauses or complex nominal phrases. Finally, TAASSC considers not only finite clauses but also non-finite clauses. The other two metrics included in TAASSC take into account the total number of dependencies per clause. One index represents the average number of dependent items per clause, and the other represents the standard deviation of the number of dependent items per clause. Table 2 lists the clause dependency types in TAASSC (clausal complexity index = particular dependent type/ clause number).

Table 2
Clausal Dependent Types Analyzed by TAASSC

Structure	Abbreviation	Example of Structure
adjective complement	acomp	He [looks]$_{gov}$ [unhappy]$_{acomp}$
adverbial clause	advcl	*The boy [grows up]$_{gov}$ [as time flies]$_{advcl}$*
adverbial modifier	advmod	*[Suddenly]$_{advmod}$, The rabbit [appears]$_{gov}$*
Agent	agent	*My money has been [stolen]$_{gov}$ by the [thief]$_{agent}$*
Auxiliary verb	aux	She[is]$_{aux}$ [dancing]$_{gov}$
Bare noun phrase temporal modifier	tmod	*Last [week]$_{tmod}$, I [bought]$_{gov}$ a new bag*
Clausal complement	ccomp	*I am [sure]$_{gov}$ [that she is right]$_{ccomp}$*
Clausal coordination	cc	*[Alice sings]$_{gov}$ and [Anna dances]$_{cc}$*
Clausal prepositional complement	pcomp	*They [heard]$_{gov}$ [about you missing classes]$_{pcomp}$*
Clausal subject	csubj	*[What I said]$_{csubj}$ [is]$_{gov}$ very important*
Conjunction	conj	She [sings]$_{gov}$ and [dances]$_{conj}$
controlling subject	xsubj	*[Ken]$_{xsubj}$ likes to [watch]$_{gov}$ TV*
Direct object	dobj	*My mother [bought]$_{gov}$ me a new [watch]$_{dobj}$*
Discourse marker	discourse	*[Well]$_{discourse}$, I [like]$_{gov}$ flowers*
Existential "there"	expl	*[There]$_{expl}$ may [be]$_{gov}$ a bookshop*
Indirect object	iobj	*My mother [bought]$_{gov}$ [me]$_{iobj}$ a new watch*
Modal auxiliary	modal	*She [may]$_{modal}$ [be]$_{gov}$ nice.*
Negation	neg	*He did [not]$_{neg}$ [steal]$_{gov}$* the *money.*
Nominal complement	ncomp	*He [is]$_{gov}$ a [doctor]$_{ncomp}$*
Nominal subject	nsubj	The [girl]$_{nsubj}$ [is]$_{gov}$ beautiful The [athlete]$_{nsubj}$ [ran]$_{gov}$ quickly
Open clausal complement	xcomp	*I am [ready]$_{gov}$ [to leave]$_{xcomp}$* *[Going]$_{gov}$ [fishing]$_{xcomp}$ is fun*
Parataxis	parataxis	*"That song," he [said]$_{parataxis}$, "[is]$_{gov}$ beautiful"*
Passive auxiliary verb	auxpass	*His car has [been]$_{auxpass}$ [stolen]$_{gov}$*
Passive clausal subject	csubjpass	*[That she helped others]$_{csubjpass}$ was [praised]$_{gov}$ by everyone*
Passive nominal subject	nsubjpass	*[Jeff]$_{nsubjpass}$ was [laughed]$_{gov}$ by John*
Phrasal verb particle	prt	*They [gave]$_{gov}$ [up]$_{prt}$ their friendship*
Prepositional modifier	prep_	*They [went]$_{gov}$ [into the supermarket]$_{prep\_into}$*
Subordinating conjunction	mark	*He watched TV [after]$_{mark}$ he [finished]$_{gov}$ homework*
Undefined dependent	dep	N/A

Note. "*gov*" represents the governor of the dependent

Measuring fine-grained phrasal indices. In order to compute fine-grained indices of phrasal complexity, the phrasal complexity metric in Kyle's (2016) Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC) is adopted. TAASSC analyzes seven noun phrases types and ten phrasal dependent types (Table 3). There are three types of fine-grained indices of phrasal complexity included in TAASSC. The first type measures the average number of dependents contained in each phrase type (e.g., nominal subjects). The second type counts the occurrences of particular dependent types (e.g., adjective modifiers) in various types of noun phrases. The

third one calculates the average frequency of a certain type of dependent term in a particular type of noun phrase (e.g., adjective modifiers in nominal subjects). Each index of TAASSC consists of two versions, one for noun phrases containing pronouns and the other for noun phrases that do not contain pronouns. This is because noun phrases can be composed of pronouns. But except in rare cases such as relative clauses, pronouns do not collocate with direct subordinates. Therefore, pronoun-containing phrases may affect the proportion of dependent terms if not counted separately. In addition, TAASSC also shows how well the mean represents the data by calculating the standard deviation (Kyle, 2016). A total of 132 fine-grained indices of phrasal complexity (and variations) are included in TAASSC. Table 4 provides an overview of the different types of phrasal indices.

Table 3
Phrase Types and Dependent Types Analyzed by TAASSC

Structure	Abbreviation	Example of structure
Phrase Types		
Nominal subject	nsubj	[The boy in the blue shoes] $_{nsubj}$ bought that little girl food.
Passive nominal subject	nsubj_pass	*[That little girl] $_{nsubj\_pass}$ was bought food by the boy in the blue shoes.*
Agent	agent	*That little girl was bought food by [the boy in the blue shoes] $_{agent}$.*
Nominal complement	ncomp	*She is [a little girl] $_{ncomp}$.*
Direct object	dobj	*The boy in the blue shoes bought that little girl [food] $_{dobj}$.*
Indirect object	iobj	*The boy in the blue shoes bought [that little girl] $_{iobj}$ food.*
Prepositional object	pobj	The boy in [the blue shoes] $_{pobj}$ bought that little girl food.
Dependent Types		
Determiners	det	*[The] $_{det}$ boy in [the] $_{det}$ blue shoes bought [that] $_{det}$ little girl food.*
Adjective modifiers	amod	*The boy in the [blue] $_{amod}$ shoes bought that [little] $_{amod}$ girl food.*
Prepositional phrases	prep	*The boy [in the blue shoes] $_{prep}$ bought that little girl food.*
Possessives	poss	*That is [his] $_{poss}$ blue shoes.* *[Jeff's] $_{poss}$ shoes are blue.*
Verbal modifiers	vmod	*I don't have anything [to do] $_{vmod}$.*
Nouns as modifiers	nn	This is an [apple] $_{nn}$ tree.
Relative clause modifiers	rcmod	*I bought the shoes [you like] $_{rcmod}$.* *The boy [who helped others] $_{rcmod}$ is admirable.*
Adverbial modifiers	advmod	*He is a [really] $_{advmod}$ good boy.*
Conjunction "and"	conj_and	*Jeff [and] $_{conj\_and}$ John.*
Conjunction "or"	conj_or	Jeff [or] $_{conj\_or}$ John.

Table 4
Phrasal Indices Included in TAASSC

Index type	Average		Standard deviation		Total
	pronouns	no pronouns	pronouns	no pronouns	
Number of dependents per nominal	8	8	8	8	32
Occurrence of particular dependents	10	10			20
Occurrence of particular dependents per particular nominal	40	40			80
Total	116		16		132

3.3 Procedure

In order to probe into the relationship between syntactic complexity indices and holistic scores of writing quality in independent essays from the *Spoken and Written English Corpus of Chinese Learners*, a stepwise multiple linear regression analysis was adopted in this study to analyze the data derived from each index type (large-grained indices of syntactic complexity, fine-grained indices of clausal complexity, and fine-grained indices of phrasal complexity), and to compare relationship between the three types of syntactic complexity indices and the holistic writing scores.

First of all, normality was checked. It is necessary to check the normality of data because data obeying normal distribution or approximate normal distribution is the prerequisite to ensure the validity of the data in the statistical methods (i.e., Pearson Correlation) to be used subsequently. Any variables that violated a normal distribution were discarded. To check normality, SPSS was adopted in this study. In this study, the number of samples is 120, so the result of K-S test was employed.

Next, Pearson Correlation was conducted on the remaining variables to examine whether they were meaningfully correlated with holistic essay score or not.

Pearson Correlation analysis is used to measure the linear correlation between two variables and the degree was represented by Pearson Correlation Coefficient |r|. Any variables that did not reach an absolute correlation value of r> 0.1 with the holistic essay scores and a significance of p< 0.05 were removed from further analysis.

Then, the remaining variables were checked for multi-collinearity. Multi-collinearity refers to the phenomenon that independent variables are highly correlated and information overlaps (Xu Hongchen, 2013). This phenomenon is very disadvantageous to multiple linear regression analysis because it confuses the boundaries between the predictive variables, which renders researchers unable to determine the contribution of the predictive variables to the regression equation.

Finally, the remaining variables were entered into a stepwise multiple regression using SPSS. Then, comparisons among three models for three types of indices were conducted.

4. RESULTS ANALYSIS

4.1 Relationship Between Syntactic Complicity and Holistic Scores

Among the syntactic complexity indices of L2SCA (Lu, 2010, 2011), only three were normally distributed: C/T, DC/C, and CN/C (Table 5), of which two variables did not reach the minimum correlation threshold with the overall score of articles in the corpus r>0.1 and p<0.05 and were therefore removed from further analysis. The remaining variable (CN/C) did not need to be tested for multicollinearity, so it was directly fed into stepwise regression. The final model showed that CN/C (number of compound noun phrases per clause) explained 5.4% of the variance in the overall score of the article (R=0.233, R^2=0.054).

Table 5
Correlations for SCA Variables with Normal Distribution

	Person Correlation [a] (N=120)	Sig. (2-tailed)
Clauses per T-unit (C/T)	-0.072	0.435
Dependent clauses per clause (DC/C)	-0.010	0.916
Complex nominals per clause (CN/C)	0.233	0.011*

Note. The dependent variables are holistic writing scores

4.2 Relationship Between Clausal Complicity and Holistic Scores

Among the data for fine-grained indices of clausal complexity, only two of Kyle's (2016) clausal complexity indices demonstrated normal distributions (nominal subject per clause and direct object per clause). The other clausal complexity indices violated the normality assumption and were therefore excluded from further analysis. Of the remaining two variables (Table 6), one variable did not meet the minimum correlation threshold r>0.1 and p<0.05 with writing scores. Since there was only one variable left in the clause complexity index, further analysis of multicollinearity was unnecessary, so the variable was directly put into stepwise regression. The resulting model, including the only remaining variable (number of noun subjects per clause), explained 5.7% of the overall score variance (R = 0.239, R^2 = 0.057).

Table 6
Correlations for Clausal Complexity Variables with Normal Distribution

	Person correlation [a] (N=120)	Sig. (2-tailed)
Nominal subject per clause	-0.239	0.008**
Direct object per clause	0.025	0.782

Note. The dependent variables are holistic writing scores

4.3 Relationship Between Phrasal Complicity and Holistic Scores

Among the data for fine-grained indices of phrasal complexity, there were eighteen Kyle's (2016) phrasal complexity indices demonstrating normal distributions (see Table 4.7). Other phrasal complexity indices violated the assumption of normality and were removed from further consideration. Of the remaining eighteen variables (Table 6), six variables: dependents per prepositional object, dependents per prepositional object (no pronouns), adjective modifiers per nominal, adjective modifiers per nominal (no pronouns), determiners per prepositional object (no pronouns) and determiners per prepositional object reached the minimum correlation thresholds of r>0.1 and p<0.05 with holistic writing scores of the examined essays. Other variables were removed from further analysis. Of the remaining six variables, four was excluded due to their inconsistencies of multicollinearity. In the meanwhile, the remaining two variables (adjective modifiers per nominal and determiners per prepositional object) were entered into the stepwise regression. The resulting model explained 10.1% (R = 0.317, R^2 = 0.101) of the variance of the holistic essay scores.

Table 7
Correlations for Phrasal Complexity Variables with Normal Distribution

	皮尔逊相关 [a] (N=120)	显著性 (双尾)
Dependents per nominal	0.080	0.387
Dependents per nominal subject	-0.044	0.634
Dependents per prepositional object	0.186	0.042*
Dependents per nominal subject (no pronouns)	-0.018	0.846
Dependents per prepositional object (no pronouns)	0.202	0.027*
Dependents per nominal subject (SD)	-0.119	0.195
Dependents per prepositional object (SD)	0.033	0.722
Dependents per nominal subject (no pronouns, SD)	-0.140	0.126

To be continued

Continued

	皮尔逊相关[a] (N=120)	显著性 (双尾)
Dependents per prepositional object (no pronouns, SD)	0.052	0.572
Determiners per nominal	0.127	0.167
Adjective modifiers per nominal	-0.207	0.023*
Determiners per nominal (no pronouns)	0.142	0.122
Adjective modifiers per nominal (no pronouns)	-0.215	0.018*
Prepositional phrases per nominal subject	0.004	0.963
Determiners per nominal subject (no pronouns)	-0.001	0.987
Prepositional phrases per nominal subject (no pronouns)	0.003	0.971
Determiners per prepositional object	0.225	0.014*
Determiners per prepositional object (no pronouns)	0.215	0.018*

Note. The dependent variables are holistic writing scores

Table 8
Phrasal Complexity Indices Entered into the Regression Model

Model	Predictors included	R	R2
1	Determiners per prepositional object	0.225	0.051
2	Determiners per prepositional object, adjective modifiers per nominal	0.317	0.101

5. DISCUSSION

5.1 Predictability of Large-Grained Syntactic Indices on Writing Scores

Correlation was found between the large-grained indices of syntactic complexity and essay scores, but the effect was small. In this study, only three of Lu's (2010, 2011) fourteen L2SCA indices passed the normality test, and only one index (complex nominals per clause) reached the minimum threshold of Pearson correlation and was finally included in the predictor model. The resulting model explained 5.4% of the variance in article scores. The positive correlation between complex nominals per clause and holistic essay scores indicates that essays which include more complex nominals (nouns with modifiers, nominal clauses, gerunds and infinitives that function as subjects) in clauses tend to be given higher scores. The results, however, didn't show great effects and explained only a small portion of the variance in holistic scores of writing quality. Traditional unit-length indicators, such as MLT and MLC, were inadequate in predicting writing scores. Notably, metrics in L2SCA related to sentence complexity (C/S) and dependent clause usage (such as CT/T, DC/C, etc.) also showed no correlation with scores, which lend support to Biber et al.'s (2011) judgment that clausal relationships may not be suitable for measuring the level of L2 academic writing.

Take two consecutive sentences selected from low-scoring and high-scoring articles as examples:

(e) ***It seems that*** *KFC has the stronger competition with MacDonald's.* ***We can easily see that*** *there are always more KFC than MacDonald's in China.* (60.29 marks)

(f) *KFC surpassed MacDonald's in the total number of stores and the sum of profit, although they are original on different levels.* ***We can also easily draw a conclusion that*** *the speed of the development of MacDonald's is conspicuously faster than that of KFC.* (81.80 marks)

In sentence (e), there are five clauses (two main clauses, two *that*-clauses and a comparative adverbial clause), three nouns with modifiers (*stronger competition*, *more KFC*, and *MacDonald's in China*) and two nominal clauses (a predicative clause and an object clause). Therefore, the number of complex nominals per clause in this sentence is 1 (5/5). There are also five clauses in sentence (f) (two main clauses, an *although*-clause, a *that*-apposition clause and a comparative adverbial clause), and five nouns with modifiers (*the total number of stores*, *the sum of profit*, *different levels*, *the speed of the development of MacDonald* and *that of KFC*) and one nominal clause (an appositive clause). Therefore, the number of complex nominals per clause is 1.4 (6/5).

Despite the seemingly huge gap between the two passages in terms of their writing quality, they contain the same number of clauses. This means that for articles written by college students in China, at least a part of the syntactic complexity indices such as C/S and C/T may be invalid, given their generally high level of English proficiency. Another reason for the possible invalidity of large-grained syntactic complexity indices is that great emphasis has been placed on teaching formulaic chunks in writing classes in China. The part of the example sentences in bold types (e.g., *We can easily draw a conclusion that…*) are typical sentence patterns included in the teacher's 'test-taking skill kits'. An experienced rater, however, will be able to tell at a glance which sentences contain overly-used modular language and therefore do not genuinely reflect the writers' language proficiency. Still, some of the SCA metrics such as the CN/C index, which gauges the number of complex nominal components, can reflect different writing levels to a certain degree.

The above results are generally consistent with the results from Kyle (2016) and Kyle & Crossley (2018). That is, coarse-grained complexity indices have limited predictive power for academic writings. However, in Kyle & Crossley (2018), the last index that entered the regression equation in the L2SCA index was the Mean Length of Clause (MLC, R^2=0.058), but no correlation between MLC and score was found in this study. The reason is probably that the corpus sources of this study are relatively homogeneous (all written by second-

year college students), while the sources of the TOEFL texts used by Kyle & Crossley (2018) might be more heterogeneous. The results of this study are also contrary to the findings of Gao Xia (2021). In Gao Xia (2021), three indices of SCA: CT/T (complex T-unit ratio), CP/T (coordinate phrases per clause), CN/C (complex nominals per clause) explained 18.5% of the score variance, and the predictive power of coarse-grained indices far exceeds that of the other two categories of indicators. This discrepancy can be explained by the different types of corpora used by Gao Xia (2021) which contain 20 different styles and different levels of writings by middle school and university students. This further shows that syntactic complexity indices are sensitive to different text types (Biber et al., 2011; Polio & Yoon, 2018) and language ability differences (Norris & Ortega, 2009; Kuiken & Vedder, 2019b).

5.2 Predictability of Fine-Grained Clausal Indices on Writing Scores

The relationship between fine-grained indices of clausal complexity and writing scores was also significant, and stronger than large-grained indices of syntactic complexity. In this study, only two clausal complexity indices reached the first threshold of normal distribution, and only one index (nominal subject per clause) was included in the final predictor model after the Pearson correlation test, explaining 5.7% of the variance in essay scores. The nominal subject per clause is negatively correlated with overall essay scores. This suggests that too many nominal subjects used in clauses can lead to lower scores. This finding is consistent with Kyle & Crossley (2018), which also used academic writings as their testing corpus. This result provides some evidences for Biber et al.'s (2011, pp.29-30) hypothesis that writers will progress from writing characterized by finite dependent clauses to writing characterized by non-finite dependent clauses (e.g., infinitive clauses). Chinese college students tend to have a relatively weak grasp of nominal components including infinitives, gerunds, and nominal clauses (Lu, 2010). The absence of these structures has resulted in an increase in the proportion of nominal subjects in the clauses, which in turn, negatively affects the overall scores of the essays.

In addition, clausal complexity indices are calculated by dividing dependent items by the number of clauses. Although they specifically identify the dependent items contained in clauses, they do not further analyze the phrasal structures inside the dependent items. For example, clausal complexity indices can distinguish whether the subject in a clause is composed of a noun phrase or a noun clause, but it cannot further explain what kind of modifiers the noun subject takes (Is it modified by an adjective or a prepositional phrase?). In addition, due to the above-mentioned influence of teaching and study habits, Chinese college students tend to over-use subordinate clauses (especially adverbial subordinate clauses). Thus, clause complexity metrics may not effectively judge their writing outcomes.

5.3 Predictability of Fine-Grained Phrasal Indices on Writing Scores

The relationship between fine-grained indices of phrasal complexity and essay scores was significant, and the correlation effect far exceeded the previous two indices. Among them, eighteen phrase complexity indices were normally distributed, and six indices were significantly correlated with essay scores. After removing the four variables that violated multicollinearity, the regression model generated by the remaining two variables (adjective modifiers per nominal and determiners per prepositional object) explained 10.1% of the score variance. The findings further support the superiority of methods analyzing specific linguistic structures over the use of opaque linguistic indices such as the MLC in predicting the scores of L2 writings (Norris & Ortega, 2009; Kyle & Crossley, 2018; Kuiken & Vedder, 2019b). However, this result also reveals a potential pitfall of fine-grained indices, i.e., variables associated with fine-grained indices tend to be rare in written texts (Kyle & Crossley, 2018). Some constructs in TAASSC (e.g., passive nominal subject and indirect object) rarely occur in the corpus of this study, leading to the exclusion of such variables from the analysis. The two variables that were highly correlated with essays scores in this study were:

(1) **Determiners per prepositional object**. The findings demonstrated that the index of determiners per prepositional object explained 5.06% of the variance in essay scores. The positive correlation between determiners per prepositional object and holistic essay scores indicates that high-scoring essays tend to include more prepositional object with more determiners. The determiners included many parts of speech, such as: articles, possessive pronouns, noun genitives, demonstratives, numerals, etc., but its role in measuring writing is often ignored. From the perspective of daily teaching experience, the use of determiners (*the*, *their*, *every*) in high-scoring texts such as (g) can often explain the writer's language skills.

(g) *With the advent of globalization, international companies begin to set up their branches in every corner of the world.*

(2) **Adjective modifiers per nominal**. The result indicated that the index of adjective modifiers per nominal explained 4.28% of the variance in essay scores. This shows that adjective modifiers are not only very common in college students' English writings, but also reflect their writing ability. The use of adjective modifiers is not only a way to construct complex nouns (Lu, 2010), but also reflects the author's ability to nominalize verbs or reduce clauses in order to compress information. For example, by rephrasing *as the number of stores increases rapidly* into *the rapid increase in the number of shops*, more information can be packaged into the original sentence.

Another possible high-scoring indicator is adverb modifiers, but since Chinese college students generally underuse adverbial modifiers and the size of corpus analyzed in this study is too small, this index did not pass the normal distribution test.

6. CONCLUSION

This paper revolves around two main criticisms of syntactic complexity metrics: the lack of interpretability for large-grained indices (Bardovi-Harlig, 1992; Norris & Ortega, 2009; Biber et al., 2011) and the excessive focus on clausal complexity metrics (Biber et al., 2011). A corpus of 120 essays written by Chinese college students were analyzed to compare the effectiveness of large-grained syntactic complexity indices and fine-grained indices of clausal and phrasal complexity in predicting overall writing scores. The results showed that phrasal complexity indices are the best predictors of writing scores. This result suggested that not only in academic writing, but also in the advanced stage of L2 writing, phrasal complexity is a more valuable tool than clausal complexity. In addition, the relationship between various dimensions of syntactic complexity indices and learners' language level may be complicated (Lu & Xu, 2016), so using hybrid complexity indices to predict the quality of second language writing can complement each other and achieve better effects (Bulté & Housen, 2018). Corpus research on syntactic complexity plays a guiding role in foreign language teaching as well, especially second language writing teaching. Language teachers can adopt syntactic complexity indices as teaching tools via a process of 'reverse engineering', guiding students to gradually increase the complexity of their written output. For lower-proficiency students, emphasis can be placed on eliciting more subordinate clauses, while for higher-proficiency students, they can be trained to use more complex phrases and phrasal modifiers.

This study has certain limitations. First, it is based on a specific type of L2 writing (expository essays on a single topic). In these texts, there might be an absence of some particular language structures and a plethora of some other language features, which have hindered some indices from playing a role. On the other hand, due to the lack of strictly manually-graded essays, the number of texts used in this study is relatively small, which may compromise the generalizability of the research. Future studies should employ larger corpora and examine the application of various complexity metrics to texts of different styles and levels. Second, this study examines whether there is a linear relationship between complexity indices and writing scores. Future research can adopt longitudinal research designs (e.g., Gray et al., 2019; Schaub, 2019) to explore the nonlinear and dynamic trends between complexity indices and second language ability.

REFERENCES

Bao, G. (2009). Syntactic complexity in EFL learners' essays: A multidimensional perspective. *Foreign Language Teaching and Research, 41*(4), 291-297.

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL quarterly*, *26*(2), 390-395.

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.

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.

Chen Guiyuan. (2010). A study of the writing performance measures and their attributes. *Modern Foreign Languages, 33*(1), 72-81.

De Clercq, B., & Housen, A. (2017). A cross-linguistic perspective on syntactic complexity in L2 development: Syntactic elaboration and diversity. *The Modern Language Journal*, *101*(2), 315-334.

Gao Xia. (2021). Syntactic complexity in English writing by Chinese EFL learners at different levels. *Foreign Language Teaching and Research, 53*(2)*,* 224-237.

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

Hwang, H., Jung, H., & Kim, H. (2020). Effects of written versus spoken production modalities on syntactic complexity measures in beginning-level child EFL learners. *The Modern Language Journal*, *104*(1), 267-283.

Jiang, X. Q., Cai, J., & Tang, J. L. (2011). Impacts of an automated essay scoring tool on the development of writing proficiency of Chinese college EFL learners. *Shandong Foreign Language Teaching Journal, 32*(6), 36-43.

Kuiken, F., & Vedder, I. (2019b). Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish. *International journal of applied linguistics*, *29*(2), 192-210.

Kuiken, F., Vedder, I., Housen, A., & De Clercq, B. (2019a). Variation in syntactic complexity: Introduction. *International Journal of Applied Linguistics*, *29*(2), 161-170.

Kyle, K. (2016). Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication. Georgia State University, Atlanta, GA, 2016.

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.

Lei, L. (2017). Syntactic complexity in the academic writings by Chinese EFL learners. *Journal of PLA University of Foreign Languages, 40*(5),1-10.

Li, M. X., & Liu, Y. B. (2016). A corpus-base study on the syntactic complexity of English writing by Chinese middle school English learners. *Journal of Northeast Normal University (Philosophy and Social Sciences), 279*, 140-145.

Liu, L. G., & Liao, H. T. (2018). Language complexity theories and measurement. *Foreign Language Research, 34*(1), 52-55.

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. F., & Xu, Q. (2016). L2 syntactic complexity analyzer and its applications in L2 writing research. *Foreign Language Teaching and Research, 48*(3), 409-420.

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.

Polio, C., & Yoon, H. J. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics*, *28*(1), 165-188.

Schaub, S. The longitudinal development of clausal and noun-phrasal complexity in German intermediate learners of English. *Learner Corpus Research 2019 Warsaw, 12–14 September*, 76.

Ströbel, M., Kerz, E., & Wiechmann, D. (2020). The relationship between first and second language writing: Investigating the effects of first language complexity on second language complexity in advanced stages of learning. *Language Learning*, *70*(3), 732-767.

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.

Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, *58*(1), 87-125.

Xu, H. C. (2013). *Learning statistics from examples of second language research*. Beijing: Foreign Language Teaching and Research Press.

Xu, X. Y., Wang, W. N., & Xiong, Y. Y., et al. (2013). A study on the syntactic complexity of English essays written by Chinese students of English. *Foreign Language Teaching and Research, 45*(2), 264-275.