Papers in Language Testing and Assessment, 2019, Volume 8, Issue 1, 31-51

Noun phrase complexity in integrated writing produced by advanced Chinese EFL learners

Lirong Xu[1]
Zhejiang University

This study aims to investigate the relationship between the noun phrase complexity of advanced Chinese EFL learners' integrated writing and the score assigned by expert raters. Their written performance was also compared with those of native English speakers (NS) at university level with particular reference to the use of noun phrases. One hundred and twenty integrated writing samples were collected from an English writing test administered in a southeastern province of China. Results showed that there was a moderately positive correlation between the use of complex nominals in test-takers' writing and the corresponding score. More specifically, non-native speakers of English (NNS) and NS groups differed significantly in the majority of noun phrase complexity measures. The implications are discussed concerning noun phrase complexity as a more reliable measure of syntactic complexity for an integrated writing test.

Key words: integrated writing; noun phrase complexity; academic writing

Introduction

Integrated writing tasks, commonly used in second language assessment, require examinees to integrate source texts (reading or listening material or both) into their writing which is thought to resemble more authentically the types of performances needed for academic studies (Cumming et al., 2006). The use of integrated writing tasks in language testing contexts is also believed to bring beneficial washback to language teaching and learning. In real academic settings, students are usually engaged in such activities as writing research papers based on multiple source texts or synthesizing ideas on a certain topic. This ability to write from sources through combining reading and writing skills has been considered crucial in the process of developing academic writing literacy. As a result, integrated writing tasks have been

Email address for correspondence: lynnxlr@163.com

employed in large-scale standardized tests such as the Test of English as a Foreign Language (TOEFL) in the US as well as in language classrooms that focus on academic English language development (Grabe & Zhang, 2013).

Students' writing performance is usually approached by analyzing particular linguistic features at lexical, syntactic or discourse level. Evidence has been mounting from corpus-based studies that academic writing is characterized by greater grammatical complexity at the phrasal level, especially noun phrase complexity, as writers develop their academic writing literacy (Biber, Gray, & Poonpon, 2011; Parkinson & Musgrave, 2014). Biber et al. (2011) hypothesized a series of developmental stages of complexity features for L2 learners of English based on the observed developmental patterns for L1 learners. Recently their hypothesized developmental index has been tested and confirmed by several studies on the academic writing of advanced English L2 learners (e.g., Parkinson & Musgrave, 2014; Staples, Egbert, Biber, & Gray, 2016; Ansarifar, Shahriari, & Pishghadam, 2018). In general, research findings of these studies indicated that the use of phrasal complexity features in writing increased with academic level. However, whether noun phrase complexity is related to writing proficiency as represented by writing scores remains relatively under-researched. This inquiry is meaningful as it would shed some light on refining the sub-construct definition of syntactic complexity for integrated writing tasks.

Literature review

Grammatical complexity at sentence level

Studies on grammatical complexity have been conducted by researchers from different angles in the sub-fields of linguistics. Researchers in second language acquisition and language assessment often refer to the cognitive complexity involved in comprehending a task, for example, the instructions given before a listening passage. They are mainly concerned about how the cognitive complexity of a particular task might influence students' language use (Skehan, 1998; Robinson, 2001). In psycholinguistics, the processing difficulty of a certain grammatical structure is often measured by the amount of time required to understand it (Biber et al., 2011). Meanwhile, in applied linguistics, grammatical complexity has been most commonly used to refer to the more advanced grammatical structures acquired by language learners as they progress along their learning path, whether it be lexical, syntactic or discoursal. Researchers' interest in L2 syntactic development arises from their belief that 'the ability to arrange words syntactically into phrases and phrases into clauses demonstrates the capacity to manipulate a language's combinatorial properties, which is argued to be a strong indicator of general language acquisition' (Crossley & McNamara, 2014, p.66).

Hunt (1965; 1966) coined the term T-unit and further evidenced that T-unit functioned as a reliable index of maturity in English writing among elementary students. According to Hunt (1966), T-unit refers to a minimal terminal unit, which is 'one main clause plus whatever subordinate clauses are attached to or embedded within that main clause' (p.737). Other T-unit motivated measures like Mean length of T-unit (MLTU) and clauses per T-unit (C/TU) have also been used widely (e.g., Ellis & Yuan, 2004; Larsen-Freeman, 2006). T-unit motivated measures to examine L2 writing development have been most thoroughly documented by Wolfe-Quintero, Inagaki, and Kim (1998) who, after reviewing an extensive body of 39 relevant studies concerning complexity measures, identified the best complexity measures as C/TU and dependent clauses per independent clauses.

Despite its wide use in L2 writing research for its superiority over traditional sentence length, T-unit motivated measures are not received without criticism. Wolfe-Quintero et al. (1998) noted in their study that 61% of the studies reviewed failed to find a significant relationship between proficiency and complexity as measured by the use of subordinate clauses in T-units. In a similar vein, Ortega (2003) synthesized 25 empirical studies to evaluate cumulative evidence on the use of T-unit related measures as indices of college-level L2 writers' overall proficiency. She called for multiple complexity measures tapping different sources of syntactic complexification at both phrasal and clausal level. Furthermore, both Norris and Ortega (2009) and Lu (2010) advocated the inclusion of measures that would capture phrasal elaboration at advanced levels of English proficiency. Biber, Gray, and Staples (2016) challenged the T-unit approach because a few holistic measures were not adequate for capturing the entire system of grammatical complexity. They maintained that phrasal complexity features should be considered in future research because they were more indicative of formal written register.

Grammatical complexity at phrasal level

After conducting an extensive body of investigations into register variation between spoken and written discourse (e.g., Biber, 1988; Biber, 1995; Biber, Johansson, Leech, Conrad, & Finegan, 1999), Biber et al. (2011) claimed that subordinate clauses are more characteristic of spoken discourse while phrasal elaboration more specific to written discourse (formal writing). Two sentences extracted from their study (2011, p. 14) clearly illustrate this.

1. *Well, since he got so upset, I just didn't think we would want to wait for Tina to come back*. T-unit length: 20

2. *This may be part of the reason for the statistical link between schizophrenia and membership in the lower socioeconomic classes*. T-unit length: 20

These two sentences can be considered equally difficult by the same T-unit length, but they are more commonly found in two different registers. Clausal subordination (as exemplified in Sentence 1) is a feature of conversation rather than of academic writing. Instead, elaboration at the phrasal level (as exemplified in Sentence 2) is a salient feature of formal writing. Halliday (1989) once expressed a similar idea that linguistic complexity of writing relied more on complex nominal groups than dependent clauses.

Research on this linguistic phenomenon has been burgeoning. For example, Lu (2010) incorporated specific complex nominals in their large automated analysis of syntactic complexity in second language writing besides other T-unit measures. He deconstructed complex nominals into the following parts (i) nouns plus adjective, possessive, prepositional phrase, relative clause, participle, or appositive, (ii) nominal clauses, (iii) gerunds and infinitives in subject position. Lu (2011) later carried out a corpus-based study of college-level writing among Chinese EFL learners with 14 syntactic complexity measures, suggesting that phrasal elaboration increased with proficiency or school level. A similar study by Yoon (2017) also examined college-level Chinese EFL learners' argumentative essays, which suggested that students across proficiency levels exhibited differences in phrase-level syntactic, lexical and morphological measures but not in clause-level measures. Yoon, therefore, argued for the validity of a multidimensional linguistic complexity for different constructs.

Hypothesized developmental stages for complexity measures

Building on the extensive studies on register difference between written and spoken corpora and also the developmental patterns observed for L1 learners, Biber et al. (2011) hypothesized a similar developmental stage of complexity for L2 English learners. According to this hypothesis, L2 English learners generally progress from finite dependent clauses, then intermediate stages of nonfinite dependent clauses, and to the last stage requiring dense use of phrasal dependent structures (like constitutes in noun phrases). Several studies have provided empirical evidence in support of this hypothesis (e.g., Parkinson & Musgrave, 2014; Biber et al., 2016; Ansarifar et al., 2018). Parkinson and Musgrave (2014) tested the developmental progression by utilizing a subset of the developmental stage to compare L2 writing by English for Academic Purposes (EAP) students preparing for graduate study and already matriculated MA students. They found that EAP students' writing showed a lower proportion of nouns modifying other nouns and post-noun modifying prepositional phrases when compared with MA-level writing. A similar conclusion can be found in Ansarifar et al. (2018), who compared abstracts written by graduate students and expert writers in the field of applied linguistics. Their results confirmed that expert writers exhibited higher phrasal complexity than graduate students.

A review of the literature, however, reveals that relatively little research has been conducted to investigate test-takers' integrated writing performance through the lens of grammatical complexity at phrasal level. Even less is known concerning whether noun phrase complexity features have an impact on writing scores assigned by language testing experts. Taguchi, Crawford, and Wetzel (2013) investigated L2 English argumentative essays produced by students from various language backgrounds and found that lower-rated essays used more finite/non-finite dependent clauses while higher-rated essays are characterized by more attributive adjectives and post-noun-modifying prepositional phrases. Biber et al. (2016) undertook a lexico-grammatical analysis of the discourse features of integrated writing produced by TOEFL iBT test-takers across mode of production (speech or writing), task type (independent or integrated) and score levels. They found that apart from lexical variation, TOEFL iBT test-takers have been aware of the less salient but equally important distinction between clausal versus phrasal grammatical complexity. The features associated with their integrated writing are mostly noun phrase features- nouns, nominalizations, and noun phrase modifiers- even though no significant difference is found across score levels. Meanwhile, Grant and Ginther (2000) indicated that higher rated L2 essays contained more subordination when they used computer-tagged linguistic features to characterize L2 writing differences. On the other hand, although L2 writers produced texts more aligned with academic writing (for example, longer noun phrases) over a semester, this syntactic growth failed to predict human ratings reliably (Crossley & McNamara, 2014).

Generally speaking, 'the conceptualization of syntactic complexity in L2 writing research incorporates a richer and more fine-grained set of dimensions than those considered in theoretical frameworks for writing assessment or L2 writing rating scales' (Lu, 2017, p.498). L2 writing research attaches more importance to the sophistication aspect of syntactic complexity, which is somewhat missing in writing rating scales. This inquiry is meaningful in that it could provide ESL teachers with a more effective complexity measure for students' writing development. On the other hand, raters can be more informed when rating advanced EFL writers even though noun phrase complexity is not stated explicitly as descriptors in the rating scale. Automated essay evaluation can also benefit from implementing more valid complexity indices into their evaluation system. In summary, the present study aims to approach this under-researched issue by investigating college-level students' integrated writing performance with particular reference to noun phrase complexity. Two research questions can be formulated as follows:

RQ1: Is there a significant correlation between noun phrase complexity measures and writing scores assigned by language raters? To what extent can they predict writing scores?

RQ2: Are there any significant differences in the use of noun phrases according to the hypothesized developmental stages of complexity measures between NNS and NS?

Methods

Participants

Our writing samples were collected from a nation-wide EFL writing contest administered in China. All test-takers were college-level undergraduates of similar age and across various disciplines. They were required to write a timed argumentative essay of appropriately 500 words based on a prompt about traditional Chinese characters (See Appendix A). It is worth noting that these students have gone through an initial screening test before signing up for the second-round writing test. In other words, the participants of our current study were those enrolling for the second-round test at the provincial level. A total of 120 integrated argumentative writing samples from a southeastern province were collected (word count: Mean=532; SD=56).

Human ratings

An analytical 10-point rating scale was used in our study (See Appendix B). The grading was done by two experienced native or EFL raters who were familiar with the writing test and the scoring guidelines. After rater training, they would mark each essay independently and assign both a holistic score and three category scores according to the rating scale provided. When divergence (2 points or above) occurred between two raters, the essay would be sent to a third rater for a third grading to ensure the reliability of human scoring (He & Sun, 2015). The mean score between these two raters was recorded as the final score for each test-taker, which ranged from 4.2 to 9.1 (score: Mean=7.06; SD=0.99). Inter-rater reliability between raters for the integrated writing is strong, with Cronbach's alpha reaching 0.95. It was worth noting that each piece of writing was also assigned three category scores according to the assessment criteria on the content/idea, organization/development, and language respectively. As we were more concerned with students' language use in an integrated writing task, we only referred to the language score given by raters.

Instruments

In order to have a general picture of how the use of complex nominals in students' integrated writing relates to the score assigned by raters, the L2 Syntactic Complexity Analyzer (L2 SCA) developed by Lu (2010) was utilized. The web-based L2 SCA produces nine linguistic units of each text and 14 indices of syntactic complexity measures. Syntactic complexity indices are classified into four categories (1) length of

production units, (2) amount of coordination, (3) amount of subordination, (4) degree of phrasal sophistication and overall sentence complexity (Lu, 2011; Lu & Ai, 2015). The L2 SCA produced these quantitative indices automatically after each text was processed. Pearson's product-moment correlations between linguistic complexity indices and writing score were calculated. Simple regression analyses were then run to assess the predictive power of syntactic complexity on the writing score.

Table 1. Hypothesized developmental stages for noun phrase complexity measures

Stage	Grammatical structures	Code	Example
3	Simple phrasal embedding in the noun phrase: nouns as pre-modifiers	3.1	Cable channel
	Simple PPs as post-modifiers, especially prepositions other than *of* when they have concrete/locative meanings	3.2	House in the city
	Relative clauses	3.3	..the guy that made the call
4	Nonfinite relative clauses	4.1	The method used here... Studies employing…
	More phrasal embedding in the NP= attributive adjectives, nouns as pre-modifiers	4.2	Self-reported disease status
	Simple PPs as post-modifiers especially with prepositions other than *of* when they have abstract meanings	4.3	The specific growth rate at small population sizes
5	Preposition + nonfinite complement clauses	5.1	The idea of using…
	Complement clauses controlled by nouns	5.2	The hypothesis that female…
	Extensive phrasal embedding in the NP: multiple prepositional phrases as post-modifiers, with levels of embedding	5.3	The presence of layered structure at the borderline of cell territories

Manual coding of noun phrases

A more fine-grained analysis of students' writing followed after we identified the general trend of complex nominal use along the score range. We adjusted the hypothesized developmental stages for complexity measures proposed by Biber et al. (2011) with particular reference to noun phrase complexity measures (See Table 1).

We excluded the earlier stages of hypothesized complexity measure as our samples were written by advanced Chinese EFL learners. All the writing samples went through a strict manual coding of noun phrase complexity measures. They were documented with each instance of noun phrases which was signaled in the developmental sequences. To identify the possible developmental index of noun phrases, we also included native speakers' writings in the sub-section of the Louvain Corpus of Native English Essays (LOCNESS, Granger, 1996). We randomly selected 98 pieces of argumentative essays produced by British or American University students. These essays collected from LOCNESS are of similar length (approximately 600 words) with our exam scripts and are suitable for comparative purposes. Detailed information concerning two datasets is shown in Table 2.

Table 2. Profile of two datasets

	Number of writings	Average length of texts	Total number of words	Total number of noun phrases
Exam scripts	120	532	63869	1636
LOCNESS	98	667	65423	1458

Results

The following section begins with presenting results relevant to the first research question concerning whether the use of complex nominals (CN) in test-takers' writing is related to the score assigned by language raters. Then the frequency of the use of noun phrase complexity measures between exam scripts and LOCNESS extracted texts is compared.

RQ1: Is there a significant correlation between noun phrase complexity measures and writing scores assigned by language raters? To what extent can they predict writing scores?

With regard to the first research question put forward in the current study, we analyzed syntactic complexity focusing on the use of CN in all samples based on the web-based L2 SCA. A correlation analysis was then run. Results showed that both CN and CN/C (complex nominal per clause) are moderately correlated to the language score assigned by raters ($r=.342$, $p<.001$ and $r=.218$, $p<.05$ respectively). Table 3 below summarizes all the linguistic indices that are either positively or negatively related to the writing score ($p<0.05$).

Table 3. Pearson correlations: linguistic indices to writing score.

Index	*r*	*p*
W (Words)	0.318	0.000**
S (Sentence)	0.233	0.010*
CP (Coordinate phrases)	0.189	0.039*
CN (Complex nominals)	0.342	0.000**
C/S (Clauses per sentence)	-0.216	0.018*
C/T (Clauses per T-unit)	-0.212	0.020*
DC/C (Dependent clauses per clause)	-0.225	0.013**
DC/T (Dependent clauses per T-unit)	-0.234	0.010*
CN/C (Complex nominals per clause)	0.218	0.017*
VP/T (Verb phrases per T-unit)	-0.194	0.034*

**indicates a significance level with p<.001.
*indicates a significance level with p<.05.

In order to know the predictive power of noun phrase complexity on writing scores, we conducted a simple linear regression analysis with CN alone as a predictor and writing score as the dependent variable. The Pearson correlation (r= 0.342, p<0.0001, N=120) suggested that more CNs in test-takers' writing tend to result in a higher score. The use of CNs alone can account for approximately 11% of the variance in the overall writing score. As so many other factors might contribute to raters' judgment of test-takers' writing proficiency (e.g., content, organization, structure, vocabulary use, etc.), this predictability of CN in writing score is relatively high (Yu, 2010).

We proceeded to perform a stepwise regression analysis to find the best fitting model to account for the writing score. As can be seen in Table 4, the analysis yielded a significant model which is composed of CN, DC/C, and T/S. Together, they could account for approximately 20% of variances of writing score. We also ensured that these three predictor variables did not have problems with multi-collinearity as tolerance values were well above 0.1 and VIF around 1.

Table 4. Stepwise linear regression to predict writing score

Model		Unstandardized coefficients		Standardized coefficients	*t*	Sig.	Collinearity Statistics	
		B	Std. error	Beta			Tolerance	VIF
1.	Constant	5.123	0.489		10.481	0.000		
R^2=	CN	0.24	0.006	0.342	3.956	0.000	1	1
0.117								
2.	Constant	6.064	0.576			0.000		
R^2=	CN	0.025	0.006	0.354	10.527	0.000	0.998	1.002
0.175	DC/C	-2.412	0.838	-0.242	4.208	0.005	0.998	1.002
3.	Constant	7.319	0.843		8.679	0.000		
R^2=	CN	0.027	0.006	0.374	4.474	0.000	0.983	1.017
0.203	DC/C	-2.297	0.829	-0.230	-2.771	0.007	0.993	1.007
	T/S	-1.302	0.646	-0.169	-2.015	0.046	0.980	1.020

CN=complex nominal; DC/C= dependent clauses per clause; T/S= T-unit per sentence

Table 4 further demonstrates that CN is the most important predictor of writing score in the regression model. On the other hand, both DC/C (dependent clauses per clause) and T/S (T-unit per sentence) have a slightly negative effect on writing score. According to Lu (2010), such syntactic complexity features as DC/C and T/S are categorized into Subordination and Coordination respectively. In this aspect, our research finding was contrary to Grant and Ginther (2000) who suggested that more proficient English L2 writers tend to incorporate more subordination in their timed essays.

RQ2: Are there any significant differences in the use of noun phrases according to the hypothesized developmental stages of complexity measures between NNS and NS?

In this section, we continue to explore whether NNS of English (Chinese advanced EFL learners) have written in a similar way as NS of English regarding the use of noun phrases. Here we only adopted a subset of the developmental stages in Biber et al. (2011, pp.30) because we were mainly concerned with those stages relating to noun phrase complexity features. After we completed coding all the noun phrases according to their hypothesized stages as exemplified in Table 1, we firstly presented the distributional pattern of noun phrases across NNS and NS groups based on their raw frequency counts in texts. Figure 1 graphically displays the raw frequency of each type of noun phrases for both NNS and NS groups. As the two datasets were designed to a comparable size, we then performed the Log-likelihood (LL) test to see whether NNS and NS writers differed significantly in the use of each type of noun phrases. We used LL test because there were rare occurrences of a certain type of noun phrases (less than 5) in both datasets, for example, extensive phrasal embedding in the NPs

hypothesized to be in Stage 5. The LL test uses 'the asymptotic distribution of the generalized likelihood ratio' (Dunning, 1993). It does not assume data to be normally distributed and allows comparisons between both rare and common occurrences. Here we followed Oakes (1998) in that for a difference to be significant at the significance level of 0.01, LL score with one degree of freedom must be greater than 6.64. The LL significance test results are presented in Table 5.

As shown in Figure 1, for both NNS and NS groups, the most common types of noun modifiers are relative clauses (coded as 3.3). The second most common types of noun modifiers are nonfinite relative clauses (coded as 4.1). From Table 5, we can observe that NNS and NS groups did not show any significant difference when using relative clauses and nonfinite relative clauses to modify nouns in their essays. It can be inferred in this regard that advanced Chinese L2 learners have approximated English L1 university students.

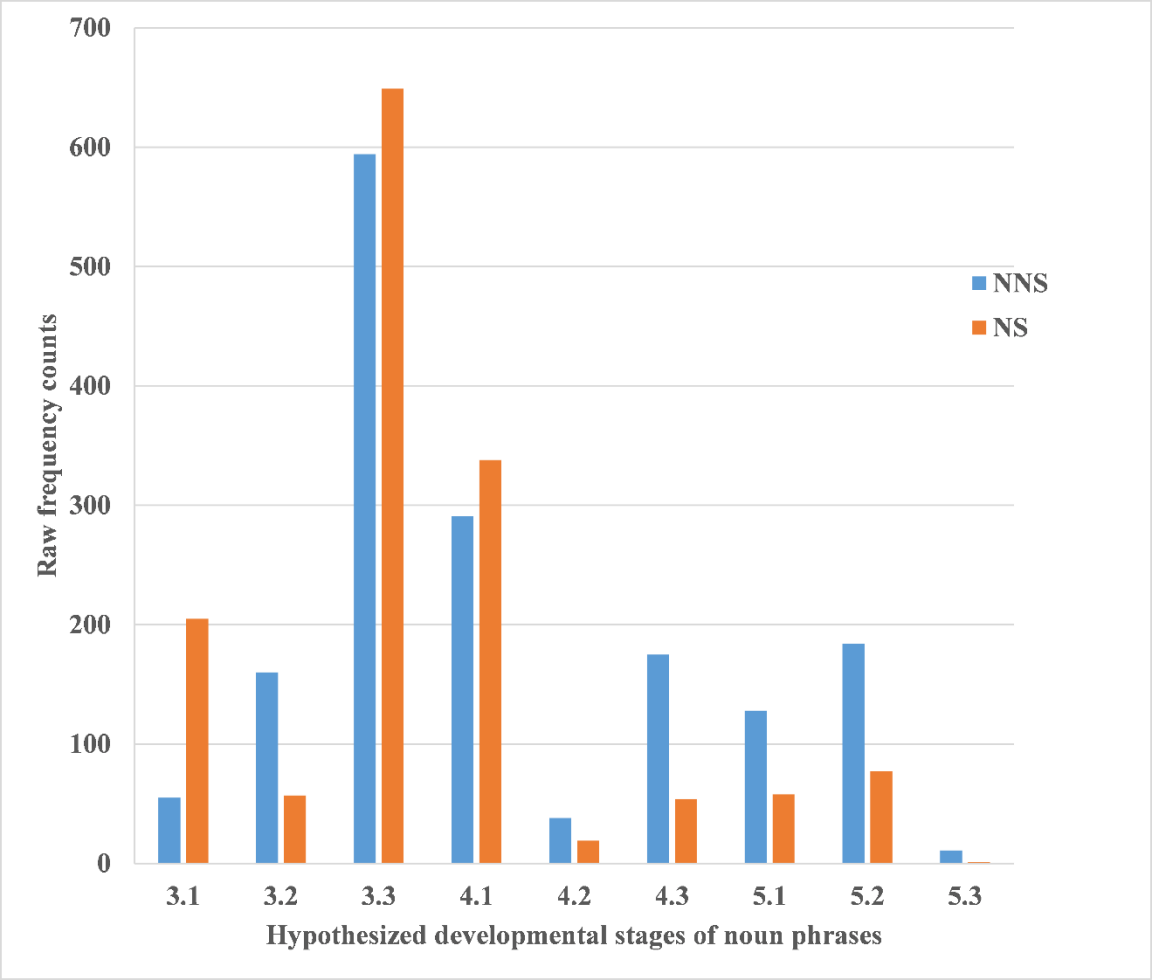

Figure 1. Noun phrase complexity measures in NNS and NS datasets.

Parkinson and Musgrave (2014) only found a marginally significant difference between EAP and MA groups (English as L2 learners) regarding the use of relative clauses as post-modifying clauses. Furthermore, Ansarifar et al. (2018) did not find

any significant difference in the use of relative clauses across MA, Ph.D. (English as L2 learners) and expert writers in their academic writing. These studies demonstrate that relative clauses as noun post-modifying device are relatively easy to acquire for NNS writers. The same is true for nonfinite relative clauses in our research finding. Specifically, both Parkinson and Musgrave (2014) and Ansarifar et al. (2018) analyzed nonfinite relative clause in terms of two sub-categories, namely, *-ed* participle as post-modifiers and *-ing* post-modifiers. However, they only noticed *-ed* participle as post-modifiers displayed a statistical significance between EAP and MA groups (Parkinson & Musgrave, 2014), and between MA and expert writers (Ansarifar et al., 2018).

Table 5. Noun phrases in NNS and NS datasets.

Stage	Code	Exam scripts	LOCNESS	Log-likelihood
		raw freq.	raw freq.	
3	3.1	55	205	88.55 (p = 0.000)*
	3.2	160	57	53.42 (p = 0.000)*
	3.3	594	649	1.29 (p = 0.256)-
4	4.1	291	338	2.48 (p = 0.116)-
	4.2	38	19	6.92 (p = 0.009)*
	4.3	175	54	70.24 (p = 0.000)*
5	5.1	128	58	28.71 (p = 0.000)*
	5.2	184	77	47.80 (p = 0.000)*
	5.3	11	1	9.99 (p = 0.002)*

*indicates a statistically significant difference with p<0.01.
-indicates a non-significant difference with p>0.01.

As can be observed in Table 5, Chinese L2 learners differed significantly from L1 writers in the use of nouns as pre-modifiers (coded as 3.1). NS tend to use more nouns as pre-modifiers in their essays. Parkinson and Musgrave (2014) also observed a greater portion of nouns as pre-modifiers in MA students' writing than in EAP students' writing. A similar pattern was observed in Ansarifar et al. (2018), who indicated that more experienced writers tend to use more regular nouns as noun pre-modifiers. Even though this grammatical feature was hypothesized to be acquired at a relatively early stage according to Biber et al. (2011), it might pose a greater challenge to English L2 learners than to native writers. For grammatical features at Stage 3, the second observed significant difference was in the use of simple PPs as post-modifiers. NNS group showed a tendency to use more simple PPs as noun post-modifiers, which were dominantly simple PPs to define locative or abstract meanings. For grammatical features at Stage 4 and Stage 5, we found a significant difference in all categories except nonfinite relative clauses (coded as 4.1). In this respect, it can be inferred that

NNS also approximated NS. However, for the rest of noun phrase complexity measures, we found advanced Chinese L2 learners were predisposed to outperforming NS university students.

Discussion

Noun phrase complexity and writing score

Concerning the first research question, our finding is congruent with that of Lu (2011), who found a significant positive relationship between CN and college-level ESL writing development besides other measures. Ai and Lu (2013) observed a statistically significant increase of CN per T-unit (CN/T) for Chinese L2 learners at higher proficiency levels than those at lower proficiency levels. The mean value of CN/C dropped from Chinese L2 learners at higher proficiency levels to those at lower proficiency levels, but this drop was not statistically significant. As far as CN was concerned, Gevara (2015) also observed significant pairwise differences among several adjacent proficiency levels and the effect size value for CN was considered moderate.

Our findings echoed previous studies in the sense that higher-rated essays tend to display more frequent use of nominal groups in their academic writing. Taguchi et al. (2013) once reported higher rates of occurrences of post-noun modifying prepositional phrases, but substantially lower rates of occurrence of subordinating conjunctions and *that*-relative clauses in high-rated essays. Yang, Lu, and Weigle (2015) reported that for two different topics, Pearson correlation between CNP/C (complex NPs per clause) and writing score was 0.12 and 0.20 respectively. Bulté and Housen (2014) showed syntactic phrasal complexity as measured by MLNP (mean length of noun phrase) correlated with holistic ratings at 0.373.

On the other hand, Bulté and Housen (2014) also discovered mean length of sentence (MLS) and mean length of T-unit (MLTU) correlated positively with subjective human ratings of writing quality. Yang et al. (2015) also found a similar pattern with T-unit complexity as a consistently strong predictor of writing score across two topics. However, this was not the case in our study. We did not find any correlation between T-unit related measures and writing score.

Although L2 learners' writing became indeed more syntactically complex, they were not predictive of human judgments of writing quality when assigned language use scores (Crossley & McNamara, 2014). In Crossley and McNamara (2014)'s study, the overall production of clauses became the only and strongest predictor of human ratings in that fewer embedded clauses were an indicator of higher ratings. Rather, nominal groups or complex phrasal elements failed to predict human judgments of

writing quality. Meanwhile, our research findings indicated a minor negative effect of embedded clauses and coordinating clauses on writing score: the use of nominal groups contributed greatly to a higher writing score. As the writing samples used in Crossley and McNamara (2014) were timed descriptive essays, genre differences between argumentative and descriptive writing, rater characteristics, or writing conditions all might lead to different findings.

Our findings also provided some empirical evidence in support of the hypothesized developmental progression by Biber et al. (2011) from the perspective of language assessment. In the present study, more proficient Chinese L2 writers tend to produce more CNs, which were found more common in the academic written register. On the contrary, subordination and coordination became less common in highly rated integrated writing. It can be inferred that advanced Chinese L2 learners had developed some awareness of phrasal embedding in their writing. They tended to embed more attributive or noun elements into NPs and were aware of the need to extend NPs by adding more post-modifiers like PPs when necessary. They also displayed capability in producing extensive phrasal embedding in NPs, which was believed to be acquired later compared with other grammatical devices. In general, they showed a tendency to write more academically by condensing meaning into noun phrases instead of adding one dependent clause after the other. Our study also indicated that language raters were aware of the syntactic features more aligned with academic writing when rating integrated writing performance. This scenario offered some evidence for employing integrated writing tasks to predict students' potential academic writing literacy.

NS and NNS writing at college level

Results from our study demonstrated that Chinese university English L2 learners approximated or even outperformed NS university students in almost all the noun phrase complexity features, except the use of regular nouns to modify nouns. Our findings ran contrary to Ai and Lu (2013) who found a statistically significant difference in the mean values of CN/C (complex nominals per clause) and CN/T (complex nominals per T-unit) across NNS and NS of English. They showed on average a smaller proportion of CNs in NNS university students' writing than NS university students' writing. In spite of the difference, they also suggested that NNS at higher proficiency levels better approximated NS in the degree of phrasal sophistication. Their NNS writing samples were collected under non-testing conditions and from a relatively wider range of Chinese university students across different proficiency levels. Our participants can be seen as a relatively homogenous group, whose English proficiency is believed to be at an advanced level. Therefore, we can infer that advanced Chinese L2 learners are at the same level with NS writers as regards the use of noun phrase complexity in their academic writing. This contrastive

evidence further justified the use of noun phrase complexity measures in integrated writing tasks, which was more suitable for assessing English L2 learners at a relatively advanced level. However, due to limited sample size, the generalizability of our research findings should be interpreted with caution.

Conclusions and implications

Investigations into syntactic complexity are abundant in English L2 writing research. Meanwhile, little is known concerning the relationship between noun phrase complexity measure and writing score assigned by language raters. With two types of syntactic complexity proposed by Biber et al. (2011) in mind, this study examined phrasal complexity in integrated writing produced by advanced Chinese L2 learners of English to see whether and to what extent it correlated to writing score. In order to gain a complete picture of the use of noun phrase complexity measures for Chinese university students, we also compared our writing samples with those written by NS university students.

Results revealed that firstly, there was a positive correlation between the use of complex nominals and the writing score; secondly a slight negative effect of subordination and coordination on writing score. These findings offered some evidence partly in support of Norris and Ortega (2009) in that more advanced L2 learners established syntactic complexity through phrasal elaboration instead of clausal coordination or subordination in later stages of second language acquisition. Thirdly, through comparing our writing samples with NS writers, we suggested that generally advanced Chinese L2 learners of English approximated NS university students concerning the use of noun phrase complexity features in their writing. They displayed a certain facility for extending NPs through embedding pre-modifiers and post-modifiers, which is more aligned with formal academic writing.

Our research findings might have implications for a clearer construct definition of syntactic complexity in English L2 writing assessment, especially for the integrated writing test. From the perspective of language assessment, for integrated writing task to have sound predictive validity in making inferences about test-takers' academic ability in a future study, the linguistic features specific to academic writing should be given due attention. Bachman and Palmer (1996) identified the need for a framework that 'enables us to use the same characteristics to describe what we believe are the critical features of both language test performance and non-test language use' (p. 10).

Secondly, corpus-based studies on academic writing can provide more insight into the critical grammatical features characterizing academic writings at different proficiency levels and thus could potentially figure into the descriptors writing for L2 writing rating scale. Language raters would become more informed about proper complexity

measures when assigning scores to integrated writing samples. Using corpus analysis tools to inform the development or revision of writing rating scale has been on the rise in response to the call for validating rating scale empirically (Gevara, 2015; Banerjee, Yan, Chapman, & Elliott, 2015). However, we acknowledge the fact that whether descriptors for noun phrase complexity measures should be incorporated or made explicit in rating scale warrants more empirical work in the future.

One limitation of this study is that we only address one aspect of language use in test-takers' integrated writing; that is, noun phrase complexity. However, no single predictor variable can fully account for L2 English writing performance. Researchers are recommended to examine multiple profiles of English L2 writing in the future, across proficiency levels or NNS and NS groups (e.g., Jarvis, Grant, Bikowski, & Ferris, 2003; Friginal, Li, & Weigle, 2014). More recently, Yan and Staples (2016) and Biber et al. (2016) advocated a multi-dimensional analysis of L2 writing because co-occurring lexico-grammatical features were more reliable for describing and capturing L2 writing proficiency. Finally, the interaction between syntactic complexity, accuracy and other sub-constructs of writing and how they combined to influence writing quality is a promising strand for future research. A fuller account of these variables would certainly contribute to more fine-grained construct definition of syntactic complexity and L2 writing proficiency.

Acknowledgments

I would like to thank Professor Sylviane Granger and the Centre for English Corpus Linguistics (CECL), Université catholique de Louvain for allowing my access to LOCNESS. I would also like to express my sincere gratitude to the anonymous reviewers of PLTA for their feedback and insightful comments. All errors that remain in the paper are mine.

References

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

Ansarifar, A., Shahriari, H., & Pishghadam, R. (2018). Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *Journal of English for Academic Purposes*, 31, 58–71.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Banerjee, J., Yan, X., Chapman, M., & Elliott, H. (2015). Keeping up with the times:

Revising and refreshing a rating scale. *Assessing Writing*, 26, 5–19.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.

Biber, D. (1988). *Variation across speech and writing.* Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. Edinburgh: Pearson Education Limited.

Biber, D., Gray, B., & Poonpon, K. (2011). Characteristics of conversation to measure complexity in L2 writing development. *TESOL Quarterly*, 45(1), 5–35.

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

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

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

Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL* (Report No. MS-30). Princeton, NJ: Educational Testing Service.

Dunning, T. (1993). Accurate Methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), 61–74.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition,* 26(1), 59-84.

Friginal, E., Li, M., & Weigle, S. C. (2014). Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing*, 23, 1–16.

Gevara, J. R. (2015). Using corpus complexity analyses to refine a holistic ESL writing placement rubric. *Papers in Language Testing and Assessment*, 4(2), 34–59.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9(2), 123–145.

Granger, S. (1996). From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast: text-based cross-linguistic studies* (pp. 37–51). Lund: Lund University Press.

Grabe, W., & Zhang, C. (2013). Reading and writing together: A critical component of English for academic purposes teaching and learning. *TESOL Journal*, 4(1), 9–24.

Halliday, M. A. K. (1989). *Spoken and written languag*e. Oxford: Oxford University Press.

He, L., & Y, Sun. (2015). Investigation of the effects of prompt characteristics on Chinese test-takers' integrated writing performance (in Chinese). *Foreign Language Teaching and Research*, 47(2), 237-250.

Hunt, K. W. (1965). A Synopsis of clause-to-sentence length factors. *The English Journal*, 54(4), 300-309.

Hunt, K. W. (1966). Recent measures of syntactic development. *Elementary English,* 43(7), 732–739.

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing*, 12(4), 377–403.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 30(4), 579-589.

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. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, 34(4), 493–511.

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16–27.

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.

Oakes, M. (1998). *Statistics for corpus linguistics.* Edinburgh: Edinburgh University Press.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics* 24(4), 492-518.

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

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22 (1), 27–57.

Skehan, P. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across levels of study, discipline and genre. *Written Communication,* 33(2), 149-183.

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.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity* (Technical Report No. 17). Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.

Yoon, H. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency and construct multidimensionality. *System*, 66, 130-141.

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing,* 28, 53–67.

Yan, X., & Staples, S. (2016). *Investigating lexico-grammatical complexity as construct Validity evidence for the ECPE writing tasks: A multidimensional analysis* (Report No. 2016-01). Michigan: Michigan Language assessment.

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 31(2), 236–259.

Appendix A. Integrated writing task

Read the following paragraphs which present contradictory views. Write a passage about the issue, clearly stating your opinion and explaining your reasons. You should write about 500 words.

At this year's Chinese People's Political Consultative Conference, the renowned film director Feng Xiaogang proposed the revival of 50 to 200 traditional Chinese characters (繁体字) in primary and secondary school textbooks to help increase understanding of our traditional culture. The proposal sparked an intense debate on whether traditional characters should be revived.

Supporters claim that restoring traditional characters will help pass on Chinese cultural heritage to the next generations since the essence of traditional values is conveyed in the pattern of the characters. For example, "見" in the character "親" indicates that having an intimate relationship necessitates seeing each other; "心" in the character "愛" demonstrates that love is from the heart.

Opponents, however, argue that traditional characters, which usually have more strokes (笔划) than simplified ones, are more difficult to learn and use. In addition, classical texts are all printed in their simplified version nowadays, indicating that traditional culture can also be promoted and preserved through simplified characters.

Appendix B. Analytic Rating scale

Category	Descriptions
Content/Ideas (40%)	1. Writing effectively addresses the topic and the task; 2. Writing presents an insightful position on the issue; 3. The position is strongly and substantially supported or argued.
Organization/Development (30%)	1. Writing is well-organized and well-developed, using appropriate rhetorical devices (e.g. exemplifications, analysis, comparison/contrast, etc.) to support the thesis or to illustrate ideas; 2. Writing displays coherence, progression, consistency, and unity; 3. Textual elements are well-connected through explicit logical and/or linguistic transitions.
Language (30%)	1. Spelling is accurate; 2. Writing displays consistent facility in the use of language; 3. Writing demonstrates appropriate register, syntactic variety, and effective use of vocabulary.