A Model of Textual Features Predicting Writing Quality of Chinese EFL Learners in the Continuation Task

Huaqing He[1], Lanyu Chen[2]

[1]China West Normal University, Nanchong, China

[2]Yuxi Normal University, Yuxi, China

Keywords: textual features, continuation task, writing quality, EFL learners

Abstract: Writing assessments, especially in China, have increasingly employed the continuation task, an integrated reading-writing task in which students read an incomplete story and then finish the story logically. This paper mainly employes the Coh-Metrix text processing Software to analyze the textual features that influence the writing quality of Chinese English as a foreign language (EFL) learner in the continuation task. Based on quantitative analysis, the following results are gained. Firstly, fluency, grammatical accuracy, lexical complexity, and cohesion are correlated with the writing score to such a degree that their indices account for 15% , 31.9% , 19.4% , and 23.6% of its variance, respectively. Secondly, the predictive model of writing quality is Writing Score (in a twenty-five-mark system) = -11.844+0.017×Number of Words+2.580×Grammatical Accuracy +0.041×The Ease of Constructing Mental Images for Content Words, which accounts for 52.4% of the variance in the evaluation of continuation writing quality. The findings are beneficial for both EFL teaching and English writing assessment.

1. Introduction

Integrated tasks—such as reading-writing or reading-listening-speaking tasks—have been increasingly employed on large-scale tests [1]. Reading and writing share some cognitive processes in second language (L2) integrated assessment tasks [2, 3]. The continuation task, a type of integrated reading-writing task in which students read and complete an incomplete story, has recently received considerable attention in China, arguably for two main reasons. First, it has been adopted in the National Matriculation English Test (the NMET, also called college entrance examination or gao kao)—one of the largest and most impactful standardized tests in the country [4, 5]—in an increasing number of provinces since its first use in the NMET in 2016 (specifically the version from Zhejiang Province, the NMET-ZJ). Second, the continuation task is beneficial to improving L2 students' learning and promoting instructional efficiency among teachers [6-13]. Several previous studies have focused on the alignment effect in the continuation task [7, 14-16] and the change in students' writing-related anxiety after training for it [17]. By contrast, the relationship between textual features and writing performance in the continuation task remains underexplored. The current study intends to fill this research gap by mainly taking advantage of the Coh-Metrix text processing Software to quantize textual features in Chinese EFL high school learners' continuation writing in order to establish a model for predicting continuation writing quality from textual features.

2. Literature review

2.1 The continuation task

Source text usage is becoming increasingly common on college-level writing tests [18]. The continuation task is a relatively new form of integrated reading-writing task that requires L2 learners to read an incomplete story and finish it logically, coherently, and creatively according to the material and prompts presented [16]. It is based on the interactive alignment model, which assumes that successful dialogue involves alignment between interlocutors [19]. The concept of alignment in the interaction between reading materials and L2 learners was extended by Wang who claimed that the

  DOI: 10.25236/iemetc.2024.016

continuation task helps language learners complete the writing task by providing a source text that offers textual and non-textual support, and that it enhances language learning via stimulating learner interaction and alignment with the source text [20, 21]. Empirical research shows that the continuation task helps learners alleviate writing-related anxiety [17, 22], reduces form-based mistakes [16], yields more accurate and complex language than independent writing tasks [6], and promotes vocabulary acquisition that is superior to the reading plus continued cloze task [13]. Researchers have also systematically verified the continuation task's reliability, validity, difficulty, and practicality, and have found it applicable for assessing learners' writing ability in large-scale and high-stakes language testing contexts [23, 24].

In view of the continuation task's theoretical basis and the empirical evidence for its ability to enhance learning as well as its reliability, validity, and feasibility regarding writing assessments, it has gained increasing popularity, including its adoption in the NMET in an increasing number of China's provinces [21, 24]. Accordingly, researchers have begun to study the continuation task on the NMET [10, 25, 26]. Chen [25] illustrated the continuation task's positive impact on the NMET through quantitative and qualitative analyses. Moreover, Chen and Zhang [26] carried out an empirical study and proposed a set of measures to ensure the high-quality evaluation of the continuation task on the NMET.

2.2 Past empirical research on the link between textual features and writing quality

Many studies have shown that textual (including cohesive, syntactic, and lexical) features are positively correlated with L2 writing quality. The study [27] indicated that lexical diversity and syntactic complexity comprise a big proportion of textual features in the assessment of writing quality. Lexical frequency, readability, and cohesion had a positive correlation with writing scores [28]. Crossley et al. [29] adopted nine Coh-Metrix syntactic indices to explore the syntactic complexity of the writings of 57 English learners within four months; they found that the syntactic complexity of the learners' writing improved noticeably over time, and that the reduced number of clauses was significantly related to the learners' improved writing scores. Crossley et al. [30] discovered that the use of global, local, and text cohesion collectively explained 42% of the variance in the overall judgment of writing proficiency. However, other scholars have drawn different conclusions. Bao [31] and Li [32] employed the Coh-Metrix computer program to examine the influence of writing proficiency on textual features and found no significant differences in lexical complexity among college English learners with different composition levels. Other studies demonstrated that the indicators of cohesive features analyzed by Coh-Metrix did not differ significantly between writings with high and low scores [27, 33].

Prior studies on the continuation task have primarily concentrated on college students [14, 16, 17, 34], while the trend of using this task in the NMET in an increasing number of Chinese provinces has drawn more attention to high school students. Furthermore, the connections between writing quality and textual features in the continuation task have not yet been systematically studied. The present study aims to fill the gap in the literature by exploring the link between writing quality and textual features in the continuation task.

3. Research design

3.1 Research participants

The participants in the study were 120 eleventh-grade EFL students (75 females, 45 males) from an ordinary senior high school in the Sichuan Province in China. They had earned an average of 16 out of 25 points on their most recent English writing exam, which was representative of most students.

3.2 The writing test

We adopted the continuation task of the NMET 2021 (see Appendix A), with the prompt providing the opening sentences of two continued paragraphs.

The reading material on the continuation writing task was a narrative that told the story of a pair of twins preparing gifts for their mother on Mother's Day. When choosing this reading material, we primarily considered that the overall content of the passage would be close to students' actual lives, and that all the words in the reading material are the basic words that students are supposed to have mastered at their level.

3.3 Measures

To study students' textual features, we used Coh-Metrix 3.0, which is an online tool for text analysis that combines a variety of technologies involving computer and corpus linguistics (see http://141.225.61.35/CohMetrix2017/). Coh-Metrix 3.0 automatically analyzes up to 106 grammatical, lexical, and semantic features of the text, covering 11 modules: referential cohesion, latent semantic analysis (LSA), lexical diversity, connectives, syntactic complexity, and syntactic pattern density, among others [35]. We used 20 measures (see Table 1), adapted from Shi et al. [36], to examine the textual features of the students' continuations from five dimensions (fluency, grammatical accuracy, lexical complexity, syntactic complexity, and cohesion). We chose these features because they are correlated with L2 writing proficiency [29, 30, 35, 37].

Table 1. Measures of analyzing the textual features in the writing

Category	**Measures**	**Description**	**Tool**
Fluency	Text length	the total number of words per essay	Coh-Metrix 3.0
Grammatical accuracy	Three-point grammatical accuracy scale	the ability to be free from grammatical errors while using language to communicate	Human rating
Lexical complexity	Measure of textual lexical diversity	a measure of lexical diversity that is not influenced by text length	Coh-Metrix 3.0
	Incidence of content words	the number of nouns, adverbs, adjectives, and main verbs per 1,000 words	
	Concreteness of content words	a measure of the extent to which the content words are concrete or abstract	
	Imageability of content words	a measure of the ease of constructing mental images for content words	
Syntactic complexity	Mean sentence length	the total number of words in each sentence	Coh-Metrix 3.0
	Number of words before the main verb	the average number of words before the main verb	
	Number of modifiers per noun phrase	the mean number of modifiers per noun phrase	
	Passive voice density	the number of agentless passive forms per 1,000 words	
	Syntactic similarity of adjacent sentences	the extent to which adjacent sentences in a sample have similar structures	
Cohesion	Incidence of all connectives	occurrence of all connectives per 1,000 words	Coh-Metrix 3.0
	Incidence of causal connectives	occurrence of causal connectives per 1,000 words	
	Incidence of logical connectives	occurrence of logical connectives per 1,000 words	
	Incidence of adversative and contrastive connectives	occurrence of adversative and contrastive connectives per 1,000 words	
	Incidence of temporal connectives	occurrence of temporal connectives per 1,000 words	
	Incidence of additive connectives	occurrence of additive connectives per 1,000 words	
	LSA overlap between adjacent sentences	similarity between the two adjacent sentences	
	LSA overlap between adjacent paragraphs	similarity between the two adjacent paragraphs	
	LSA given-new	the proportion of new information in each sentence	

3.4 Research procedure

First, all participants finished the continuation task within 40 minutes. The entire testing process in the classroom was supervised by the school's English teachers. Second, three teachers with extensive experience in scoring the NMET rated the students' writings on the continuation task according to the five-level holistic scoring rubric for the continuation task on the NMET, with a total possible score of 25 (see Appendix B). The rubric included four key criteria: (1) the connection between the main ideas of the reading material and completion of the writing task; (2) the completion and richness of the continued writing; (3) the choice of vocabulary and grammatical structure; and (4) the overall structure and coherence of the article. We adopted Pearson's correlation analysis to test the consistency of the scores given by the three teachers; the consistency was high ($r=.83$, $.82$, $.80$, respectively, $p<.01$), indicating the reliability of the scoring results. After checking the consistency of the scores calculated by the three teachers, we took the mean score as the final score of each student. Third, We roughly adopted three-point grammatical accuracy scale to holistically judge the grammatical accuracy of the students' writing: one point means there are many grammatical errors in the writing sample that affect understanding of the text; two points indicate that there are some errors in the writing sample that generally do not affect understanding of the content; three points mean that there are few or no grammatical errors in the writing sample, and understanding of the content is not affected. Grammatical accuracy was scored using human rating. Finally, we analyzed students' 19 textual features on the continuation task using Coh-Metrix 3.0.

4. Results and discussion

Since the purpose of this study is not only to build a predictive model of text quality, but also to test the predictive power of the model, we randomly divided 120 writing samples into two sets at a ratio of 3:1, namely, a training set with 90 writing samples and a test set with 30 writing samples. The results of independent-samples T test showed that there was no significant difference between the scores of the two sets ($t =.136$, $p =.892$) (see Table 2).

Table 2. Differences in writing scores between two sets

Writing Samples	N	Mean	Standard Deviation	F	p	t	p
a training set	90	14.75	2.69	2.026	.157	.136	.892
a test set	30	14.68	2.18				

After the training set and the test set were divided, we perform Pearson correlation analysis to find out which textual features are significantly related to the writing quality (score) on the training set and the more relevant ones are further conducted by the unitary linear regression analysis. Then, a predictive model is generated through the multivariate linear regression analysis. Finally, the predictive model is verified using the data on the test set.

4.1 Pearson correlation analysis and unitary linear regression analysis

we conducted Pearson's correlation analysis with a two-tailed test and found out that there were relationships of varying degrees between writing quality in the continuations and all textual features except for syntactic complexity. Table 3 only reports positive correlations with statistical significance, namely p-values less than 0.05.

Table 3. Correlations between textual features and writing scores

Category	Index	r	p
Fluency	Text length	$.399^{**}$	.000
Grammatical accuracy	Three-point grammatical accuracy scale	$.571^{**}$	.000
Lexical complexity	Concreteness of content words	$.388^{**}$	.000
	Imageability of content words	$.451^{**}$	.000
Cohesion	LSA overlap between adjacent sentences	$.304^{**}$	.004
	LSA overlap between adjacent paragraphs	$.494^{**}$	.000
	LSA Given-New	$.431^{**}$	.000

Note: **. Correlation is significant at the 0.01 level (2-tailed).

A positive relationship was found between text length and writing scores, between the three-point grammatical accuracy scale and writing scores, between concreteness of content words and writing scores, between imageability of content words and writing scores, between LSA overlap between adjacent sentences and writing scores, between LSA overlap between adjacent paragraphs and writing scores, and between LSA Given-New and writing scores at the significance of 0.000, .000, .000, .000, .004, .000, .000, respectively. The correlation coefficients were .399, .571, .388, .451, .304, .494, .431, respectively.

We selected the indicators that have the highest correlation with the writing performance in each category and carried out unitary linear regression analysis (see Table 4).

Table 4. Results of the unitary linear regression analysis

Category	**Predictable Variables**	**R**	**R^2**	**Adjusted R^2**	**F**	**Unstandardized Coefficients**	**Standardized Coefficients**	**T**
Fluency	(constant)	.399	.159	.150	16.679***	9.392	.399	6.998***
	Text length					.033		4.084***
Grammatical accuracy	(constant)	.571	.326	.319	42.635***	7.952	.571	7.450***
	Three-point grammatical accuracy scale					3.015		6.530***
Lexical complexity	(constant)	.451	.203	.194	22.430***	-5.789	.451	-1.332
	Imageability of content words					.046		4.736***
Cohesion	(constant)	.494	.244	.236	28.434***	11.116	.494	15.322***
	LSA overlap between adjacent paragraphs					7.668		5.332***

Note: Dependent Variable: writing score; *** p ≤ . 001

Table 4 presents that the adjusted effect sizes were .150, .319, .194, .236, respectively; namely, text length, three-point grammatical accuracy scale, imageability of content words, and LSA overlap between adjacent paragraphs accounted for 15%, 31.9%, 19.4%, and 23.6% of the variance in the writing scores, respectively. According to Cohen [38], the effect size of grammatical accuracy was large (R^2>.25). This means that grammatical accuracy had strong predictive power for the continuation writing scores.

4.2 Multivariate linear regression analysis

Pearson correlation analysis shows that there is a collinearity between the seven indicators and the writing scores (see Table 3), so we conduct multivariate linear regression analysis for these indicators according to the stepwise entry method, and the results are shown in Table 5.

Table 5. Model summary

Model	**Predictable Variables**	**R**	**R^2**	**Adjusted R^2**	**F**	**Unstandardized Coefficients**	**Standardized Coefficients**	**T**	**Collinearity Statistics**	
									Tolerance	**VIF**
1	(constant)	.571	.326	.319	42.635***	7.952		7.450***		
	Variable 1					3.015	.571	6.530***	1.000	1.000
2	(constant)					-10.816		-3.065**		
	Variable 1					2.884	.546	7.201***	.996	1.004
	Variable 2	.708	.501	.489	43.623***	.043	.418	5.512***	.996	1.004
3	(constant)					-11.844		-3.457***		
	Variable 1					2.580	.489	6.412***	.920	1.087
	Variable 2					.041	.394	5.343***	.982	1.018
	Variable 3	.735	.540	.524	33.703***	.017	.209	2.724**	.907	1.103
4	(constant)					-3.352		-3.852***		
	Variable 1					6.780	.512	6.922***	.908	1.102
	Variable 2					4.517	.977	4.370***	.099	10.060
	Variable 3					3.212	.211	2.851**	.907	1.103
	Variable 4	.760	.578	.558	29.095***	-3.136	-.616	-2.750**	.099	10.106
5	(constant)					-11.128		-3.352***		
	Variable 1					2.600	.493	6.780***	.896	1.117
	Variable 2					.101	.985	4.517***	.099	10.063
	Variable 3					.019	.234	3.212**	.889	1.124
	Variable 4					-.072	-.694	-3.136**	.097	10.349
	Variable 5	.776	.603	.579	25.498***	6.294	.174	2.295*	.824	1.214

Note: Dependent Variable: writing score; *** p ≤ . 001; ** p ≤ . 01; *p ≤ . 05

Variable 1: Three-point grammatical accuracy scale; Variable 2: Imageability of content words;

Variable 3: Text length; Variable 4: Concreteness of content words; Variable 5: LSA overlap between adjacent sentences

Among the five models, model 5 has the highest goodness of fit and can explain 57.9% of the variance in the writing scores. However, the tolerances of variables 2 and 4 in both model 5 and model 4 are low, and the variance inflation factor is high, suggesting the existence of collinearity between two variables. Pearson correlation analysis shows that the correlation coefficient of variable 2 and variable 4 is high ($r=.948$, $p\leq.01$), suggesting that there is a strong collinearity between two variables. Therefore, model 3 is the optimal model, and the three predictors combined account for 52.4% of the variance in the continuation writing scores. To be specific, the predictive model of continuation writing quality is: Writing score (in a twenty-five-mark system) =-11.844+2.580×Three-point grammatical accuracy scale+0.041×Imageability of content words+0.017×Text length.

4.3 Model verification

The predictive model of continuation writing quality was generated, we need to verify it. Based on the predictive model, we calculated the scores of 30 writing samples in the test set（group 2, pair 1）and compared them with the scores obtained by manual grading（group 1, pair1）, and found that the correlation between the two groups reached a significant level ($r=.696$, $p=.000$). However, the paired sample T-test showed a significant difference between the two groups ($t=-3.002$, $p=.005$).

Table 6. Paired samples statistics

Writing Samples		N	Mean	Standard Deviation	r	sig.	t	sig.
Pair 1	Group 1	30	14.68	2.18	.696	.000	-3.002	.005
	Group 2	30	15.56	1.89				
Pair 2	Group 1	26	14.86	2.06	.759	.000	-1.905	.068
	Group 2	26	15.37	1.80				

The four writing samples with the largest differences in scores between two groups were excluded (see pair 2 in table 6), namely, for 86.67% of the writing samples in the test set, the paired sample T-test showed no significant differences in scores between the two groups($t=-1.905$, $p=.068$), correlation coefficient between the two groups is .759 ($p=.000$).

4.4 Discussion

We systematically examined the link between writing scores and textual features and found relationships to varying degrees between writing quality (the holistic score) and all textual features (including fluency, grammatical accuracy, lexical complexity, and cohesion) except for syntactic complexity.

In terms of fluency, our finding is like Du and Cai's research results [28] that the longer EFL learners write compositions within the time limit, the higher their English proficiency level is.

Importantly, grammatical accuracy has strong predictive power for writing quality in our research, which is consistent with previous findings [39-42].

From the perspective of lexical complexity, our findings support previous studies [27, 28] that advanced English learners use more varied and complex vocabulary in writing than poor learners.

The results on cohesion approximately support prior studies [28, 30] that high-proficiency English learners tend to use cohesive devices to produce more coherent texts than low-proficiency English learners.

Surprisingly, syntactic complexity does not correlate with students' writing quality in our study which is inconsistent with the studies [27-29]. The discrepancy might have resulted from the difference in participants' proficiency levels in the studies. The participants in their research who were English native speakers or college-level ESL learners could produce complex syntactic structures, whereas our participants who were eleventh-grade EFL students from an ordinary high school in the Sichuan Province could only write relatively simple syntactic structures.

5. Conclusion

We aimed to explore the relationship between textual features and students' writing quality on the continuation task to offer suggestions for EFL learning and instruction, especially in China, where more provinces are planning to implement a new integrated reading-writing task in the NMET following the exam reform policy. Our findings revealed that fluency, grammatical accuracy, lexical complexity, and cohesion have significant correlations with continuation writing scores. To be specific, three indices, including text length, grammatical accuracy, and imageability of content words, combined account for 52.4% of the variance in the continuation writing scores. The findings in our study also imply that grammatical accuracy has the strongest predictive power for writing quality in continuations; this conveys the importance of grammar in EFL learning and instruction for both teachers and students.

The current study has some limitations and implications for future research. First, the participants were accustomed to independent task writing based on a provided outline and recently exposed to an integrated writing task. Whether the lack of new writing skills affects our findings requires deeper investigation. Second, we collected writing samples from 120 students, the sample size should be expanded in future research. Last, grammatical accuracy was scored using human rating in the study which was possibly subjective to a certain extent, thus, measuring grammatical accuracy in future studies will be as objective as possible.

Acknowledgements

This work was supported by the Social Science Research "14th Five-Year Plan" 2024 project in Nanchong, under Grant No.NC24B286.

References

[1] Plakans, L. (2015). Integrated second language writing assessment: Why? What? How? *Language and Linguistics Compass*, 9(4), 159–167. https://doi:10.1111/lnc3.12124

[2] Plakans, L., Liao, J., & Wang, F. (2019). "I should summarize this whole paragraph": Shared processes of reading and writing in iterative integrated assessment tasks. *Assessing Writing*, 40, 14–26. https://doi:10.1016/j.asw.2019.03.003

[3] Pan, R., & Lu, X. (2023). The design and cognitive validity verification of reading-to-write tasks in L2 Chinese writing assessment. *Assessing Writing*, 56. https://doi:10.1016/j.asw.2023.100699

[4] Cheng, L., & Qi, L. (2006). Description and examination of the National Matriculation English Test. *Language Assessment Quarterly*, 3(1), 53–70. https://doi:10.1207/s15434311laq0301_4

[5] Zhao, J. (2016). The reform of the National Matriculation English Test and its impact on the future of English in China: Will English lose its predominance in the Chinese foreign language landscape? *English Today*, 32(2), 38–44. https://doi:10.1017/S0266078415000681

[6] Jiang, L., & Chen, J. (2015). The continuation task: Effects on written accuracy, complexity, and fluency. *Modern Foreign Languages*, 38(3), 366–375.

[7] Ren, W., & Lyu, X. (2021). A meta-analysis of the alignment of the continuation writing task. *Foreign Language World*, 4, 44–52.

[8] Wang, C. (2012). The continuation task—An effective method in enhancing foreign language acquisition. *Foreign Language World*, 5, 2–7.

[9] Wang, C. (2015). Why does the continuation task facilitate L2 learning? *Foreign Language Teaching and Research*, 47(5), 753–762.

[10] Wang, C. (2022). Teaching independent of testing for more learning enhancement. *Foreign Language Teaching and Research*, 54(3), 425–432.

[11] Wang, C. (2023). Xu as a natural linkage for language interfaces. *Modern Foreign Languages*, 46(2), 237–244.

[12] Zhang, S., & Zhang, L. J. (2021). Effects of a xu-argument based iterative continuation task on an EFL learner's textual and affective development: Evidence from errors, self-initiated error corrections, and foreign language learning attitude. *System*, 98, 1–12. https://doi:10.1016/j.system.2021.102481

[13] Zhang, X., Du, L., & Zhu, X. (2022). Effects of the continuation task and the RCC task on Japanese as a foreign language vocabulary learning. *Applied Linguistics*, 43(4), 725–745. https://doi:10.1093/applin/amac006

[14] Qiu, H., & Wang, M. (2022). Effects of input enhancement on alignment effect in the continuation task. *Journal of PLA University of Foreign Languages*, 45(2), 70–78.

[15] Wang, Q. (2021). The facilitating role of aligned output in the continuation task in L2 learning. *Foreign Language World*, 6, 15–22

[16] Wang, C., & Wang, M. (2015). Effect of alignment on L2 written production. *Applied Linguistics*, 36(5), 503–526. https://doi:10.1093/applin/amt051

[17] Zhan, J., Jiang, L., & Huang, L. (2022). An investigation of L2 writing anxiety in the comparative continuation task. *Journal of PLA University of Foreign Languages*, 45(2), 61–69.

[18] Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9(1), 27–55. https://doi:10.1016/j.asw.2004.01.002

[19] Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences*, 27(2), 169–190. https://doi:10.1017/s0140525x04000056

[20] Wang, C. (2011). An analysis of three teaching preferences in light of an efficient route to L2 learning. *Foreign Language Teaching and Research*, 43(4), 540–549.

[21] Wang, C. (2016). Learning by extension. *Modern Foreign Languages*, 6, 784–793.

[22] Zhang, X. (2016). Multidimensional analysis of the effects of the continuation task on the ESL writing process. *Foreign Language World*, 6, 86–94.

[23] Wang, C., & Qi, L. (2013). A study of the continuation task as a proficiency test component. *Foreign Language Teaching and Research*, 45(5), 707–718.

[24] Liu, Q., & Chen, K. (2016). A study on the design of the continuation task used in the biannual National Matriculation English Test. *Foreign Language Teaching in Schools*, 1, 1–5 (Middle School, Edition).

[25] Chen, K. (2019). A study on the effects of the reading-writing integrated continuation task used in the National Matriculation English Test. *Foreign Language Teaching in Schools*, 11, 1–6 (Middle School, Edition).

[26] Chen, K., & Zhang, J. (2020). A study on ensuring the rating quality of the continuation task in the National Matriculation English Test. *China Examinations*, 12, 38–43.

[27] McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86. https://doi:10.1177/0741088309351547

[28] Du, H., & Cai, J. (2013). A Coh-Metrix-based model of linguistic features predicting argumentative writing quality of EFL learners. *Modern Foreign Languages*, 36(3), 293–300.

[29] Crossley, S. A., Roscoe, R., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication*, 31(2), 184–214. https://doi:10.1177/0741088314526354

[30] Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive

devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1–16. https:// doi:10.1016/j.jslw.2016.01.003

[31] Bao, G. (2011). The effects of school type and writing proficiency on EFL learners' lexical complexity across course program levels. *Journal of PLA University of Foreign Languages*, 34(4), 55–60.

[32] Li, X. (2014). Linguistic features of compositions by English majors from a multidimensional perspective: A Coh-Metrix analysis. *Journal of Nanjing Institute of Technology* (Social Science Edition), 14(1), 41–46.

[33] Plakans, L., & Gebril, A. (2017). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing*, 31, 98–112. https://doi:10.1016/j.asw.2016.08.005

[34] Peng, J., Wang, C., & Lu, X. (2020). Effect of the linguistic complexity of the input text on alignment, writing fluency, and writing accuracy in the continuation task. *Language Teaching Research*, 24(3), 364–381. https://doi:10.1177/1362168818783341

[35] McNamara, D., Graesser, A., McCarthy, P. M., & Cai, Z. (2014), *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi:10.1017/CBO9780511894664

[36] Shi, B., Huang, L., & Lu, X. (2020). Effect of prompt type on test-takers' writing performance and writing strategy use in the continuation task. *Language Testing*, 37(3), 361–388. https://doi:10.1177/0265532220911626

[37] Gebril, A., & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly*, 10(1), 9–27. https://doi:10.1080/15434303.2011.642040

[38] Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*. Lawrence Erlbaum Associates.

[39] He, L., & Shi, L. (2012). Topical knowledge and ESL writing. *Language Testing*, 29(3), 443–464. https://doi:10.1177/0265532212436659

[40] Wu, J. (2018). The effect of language distinguishing features on English native speakers' written Chinese quality. *Language Teaching and Linguistic Studies*, 2, 11–20.

[41] Wu, J., Zhou, W., & Lu, D. (2019). Assessing Chinese L2 writing quality on basis of language features and content quality. *Chinese Teaching in the World*, 33(1), 130–144.

[42] Cheng, Y. (2022). Analysis of CSL writing quality based on grammatical richness. *Language Teaching and Linguistic Studies*, 5, 10–22.