Original Article

Discourse Cohesion and Coherence and the Writing Quality of English Argumentative Essays: An Analysis Based on Coh-Metrix

Yaojing Xie

Guangxi University of Foreign Languages, Guangxi, China

***Abstract**:* The study aims to examine the relationship between discourse cohesion and coherence and the quality of English writing, leveraging Coh-Metrix to conduct a quantitative analysis of textual features of 386 English argumentative essays in the four dimensions of connectives, referential cohesion, latent semantic analysis, and situation models. The results reveal the presence of statistically significant correlations between a portion of the discourse cohesion and coherence indicators included in the study and the quality of the essays sampled, as well as their varied predictive effects on writing scores.

Science Insights Education Frontiers 2026; 32(2):5249-5264.
DOI: 10.15354/sief.26.or178

How to Cite: Xie, Y. (2026). Discourse cohesion and coherence and the writing quality of English argumentative essays: An analysis based on Coh-Metrix. Science Insights Education Frontiers, 32(2): 5249-5264.

***Keywords** English Argumentative Essay, Discourse Cohesion and Coherence, Writing Quality, Coh-Metrix*

***About the Author**: Yaojing Xie, Guangxi University of Foreign Languages, Nanning 530222, Guangxi, China. E-mail: 416381594@qq.com .*

***Correspondence to**: Yaojing Xie at Guangxi University of Foreign Languages in China.*

***Conflict of Interests**: None*

***Funding**: No funding sources declared.*

***AI Declaration**: The author affirms that artificial intelligence did not contribute to the process of preparing the work.*

© 2026 Insights Publisher. All rights reserved.

Creative Commons NonCommercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-

nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed by the Insights Publisher.

Introduction

THE THEORY of cohesion was first systematically discussed in Halliday and Hasan's (1976) *Cohesion in English*, which highlights cohesion and coherence between sentences as key to the formation of a text. They argued that a text must maintain a coherent register appropriate to the context, while coherence could be achieved structurally through cohesion (Halliday & Hasan, 1976). In their work, cohesion is defined as a semantic relation within the text, which is realized through grammatical devices (such as reference, ellipsis, substitution, and conjunction) and lexical devices (such as repetition and collocation). Hu (1994) expanded on this theory, contending that cohesion is manifested not only at lexical and grammatical levels but also through semantic, syntactic, and phonetic layers as a facilitator of the delivery of textual intent. Meanwhile, Richards et al. (2005) asserted that cohesion was about the grammatical and lexical relationships within a text aimed at maintaining coherence of its overt structure.

On the basis of the theory of cohesion, researchers have explored the essence of text coherence. Brooks and Warren (2004) argued that coherence meant the parts of a discourse should contribute to its unification, and that, in addition to cohesive devices, logical, spatial, and temporal sequences must be adjusted according to the theme. Richards et al. (2005) noted that coherence concerned the overall representation of the text's meaning, highlighting the logical and conceptual relationships of information within it and relying on shared knowledge and content organization. Zhu (2007) emphasized the multi-layered nature of coherence, viewing it as the holistic effect of a text within a situational context, which means the presence of cohesion at internal, external, linear, and other levels, as well as a close connection with the situational context. McNamara et al. (2010) considered coherence as more related to semantic logic and intelligibility, a result of the interaction between the situational context and linguistic form. Wen (2013) further argued that coherence was the intrinsically logical relationships between sentences and between paragraphs, and that cohesion was necessary for the achievement of coherence but not its sole determinant. Wen's arguments resonate with Zhang and Liu's (2018) theory, which emphasizes that textual coherence represents not only the cohesion of meaning within the text but also the connection between the text and its context, and that cohesion is not a full guarantee of textual coherence.

To recap, the existing literature tends to treat cohesion as organization of text through explicit linguistic links while relating coherence to the integrity of information and logical clarity. On the other hand, cohesion does not always ensure coherence, and texts with less apparent cohesive devices

may still be coherent in content (Hu, 1994). It is evident that the relationship between textual cohesion and coherence and the quality of writing deserves further exploration.

Research Background

Recent decades have seen an increased application of the Coh-Metrix software in the field of discourse cohesion and coherence research. Researchers have conducted numerous studies on discourse cohesion and coherence measurement using Coh-Metrix, to validate the significance of its metrics for text coherence (Graesser et al., 2007; McNamara et al., 2010; Graesser et al., 2011). Applying Coh-Metrix's cohesion and coherence indicators, Liang (2006) investigated the criteria for measuring coherence of writing of Chinese EFL Learners to find that both local variables for coherence, such as adjacent argument overlap, adjacent stem overlap, and latent semantic analysis (LSA) between adjacent sentences, and global variables, including argument overlap, stem overlap, LSA across all sentences, and LSA between paragraphs, have strong explanatory power for essay coherence. Jiang (2016) affirmed the value of Coh-Metrix, stressing that it provided a comprehensive set of objective metrics for textual analysis, offering a reference framework for the automated measurement of textual cohesion. Zhou (2023) argued that Coh-Metrix could be applied to analyze discourse cohesion and coherence in dimensions of form and content: at the level of form, linguistic cohesion could be quantified through the measurement of frequencies of keywords and chunks using argument overlap and stem overlap as metrics, and at the level of content, semantic consistency could be evaluated through LSA across all sentences to ensure overall coherence. Well-cited empirical research in this area includes but is not limited to studies by Zhang (2016), which uses four indicators of discourse coherence from Coh-Metrix 3.0—connective incidence, lexical overlap, LSA across all sentences, and tense-aspect—to examine the development of coherence in college students' English essays; Peng (2017), which validated, through the use of Coh-Metrix, the role of continuation writing tasks in enhancing linguistic and content coherence in second-language writing; and Liang (2019), which conducted a comparative analysis of English news articles from China and the United States from the perspectives of local and global coherence, further verifying the operationality of quantifying discourse coherence.

In the meantime, research on the role of Coh-Metrix in predicting writing quality has also produced valuable findings. McNamara et al. (2010) analyzed 120 essays by native English speakers to find that grammatical and lexical complexity were stronger predictors of writing quality than text cohesion. Jia and Zhang (2020), based on the analysis of 90 timed essays from TEM-8 (tests for English majors-band 8 in China) discovered that features

like textual fluency, syntactic complexity, coherence, and the situation model could effectively predict writing quality. Wang and Zhang (2020), through analyzing 386 untimed essays, identified lexical complexity and the situation model as the strongest predictors of writing scores. Liang (2006), applying Coh-Metrix to analyze argumentative essays by college students, discovered that high-scoring essays were typically marked by use of global coherence devices, whereas low-scoring essays overly relied on local coherence devices, and drew the conclusion that there was a significantly positive correlation between coherence and writing quality. Du and Cai's (2013) study, based on multidimensional indicators from Coh-Metrix, supported this conclusion. According to Zhou's (2023) research findings, while writing coherence is closely related to language proficiency, high-quality writing not only relies on language proficiency but also entails deeper textual organization and semantic coherence. Kong's (2023) study, a Coh-Metrix-based evaluation, found that the diversity of cohesive devices employed by the student was indicative of their ability to apply knowledge on discourse cohesion.

Despite Coh-Metrix-based research on the relationship between discourse cohesion and coherence and the quality of essay writing being abundant, few studies have combined human and automated evaluation to measure writing quality in addressing the said relationship. Given this research background, the present study seeks to further examine the relationship, leveraging Coh-Metrix's assessment of the text cohesion and coherence of 386 English argumentative essays by Chinese college students, and combining human scoring and automated rating by the JuKu Grading platform (an online intelligent evaluation system for English compositions).

Research Design

Research Questions

- Q 1: What is the relationship between discourse cohesion and coherence and the writing quality of English argumentative essays?
- Q 2: How and to what extent do discourse cohesion and coherence affect the results of human scoring and JuKu's automated rating for these essays?

Research Subjects

A total of 386 argumentative essays were selected from the *Written English Corpus of Chinese Learners* (WECCL 2.0) as the study's research subjects. These essays share the same genre and topic, each with a word count of approximately 300 (this arrangement is meant to minimize the effects of varia-

bles like the subject and length of text on discourse cohesion and coherence). The WECCL 2.0 contains 4,950 timed and untimed essays by English ma-

Table 1. Variables and Indicators from Coh-Metrix Relevant to the Study.

Variables	Indicators from Coh-Metrix 3.0	Description	Numbers of Indicators
Connectives	CNCAll	All connectives incidence	7
	CNCCaus	Causal connectives incidence	
	CNCLogic	Logical connectives incidence	
	CNCADC	Adversative and contrastive connectives incidence	
	CNCTemp	Temporal connectives incidence	
	CNCTempx	Expanded temporal connectives incidence	
	CNCAdd	Additive connectives incidence	
Referential Cohesion	CRFNO1	Noun overlap, adjacent sentences, binary, mean	8
	CRFAO1	Argument overlap, adjacent sentences, binary, mean	
	CRFSO1	Stem overlap, adjacent sentences, binary, mean	
	CRFCWO1	Content word overlap, adjacent sentences, proportional, mean	
	CRFNOa	Noun overlap, all sentences, binary, mean	
	CRFAOa	Argument overlap, all sentences, binary, mean	
	CRFSOa	Stem overlap, all sentences, binary, mean	
	CRFCWOa	Content word overlap, all sentences, proportional, mean	
LSA (Latent Semantic Analysis)	LSASS1	LSA overlap, adjacent sentences, mean	4
	LSASSp	LSA overlap, all sentences in paragraph, mean	
	LSAPP1	LSA overlap, adjacent paragraphs, mean	
	LSAGN	LSA given/new, sentences, mean	
Situation Model	SMCAUSv	Causal verb incidence	8
	SMCAUSvp	Causal verbs and causal particles incidence	
	SMINTEp	Intentional verbs incidence	
	SMCAUSr	Ratio of casual particles to causal verbs	
	SMINTEr	Ratio of intentional particles to intentional verbs	
	SMCAUSlsa	LSA verb overlap	
	SMCAUSwn	WordNet verb overlap	
	SMTEMP	Temporal cohesion, tense and aspect repetition, mean	

jors and some non-English majors from various levels of higher education institutions across China. The essays were completed on paper media, collected, and transcribed into digital files without any modifications to them. With its extensive scope of sources, the corpus constitutes a relatively accurate reflection of the authentic states of English writing among Chinese college students (Wen et al., 2008).

Research Instruments

Based on its specific needs, the study adopted seven connective indicators, eight referential cohesion indicators, four LSA indicators, and eight situation model indicators from Coh-Metrix 3.0's output of 106 indicators (McNamara et al., 2014), to evaluate discourse cohesion and coherence of the essays selected, each of which was assigned 27 values of variables (**Table 1**).

Research Processes

First off, automated rating of the essays was conducted using the JuKu platform, and two senior instructors in college English writing were engaged to

Table 2. Correlations between Discourse Cohesion and Coherence Indicators and Scores of the Essays.

Variable Categories	Indicators	Human Rating		JuKu Rating	
		Correlation Coefficients	Significance (2-tailed)	Correlation Coefficients	Significance (2-tailed)
Connectives	CNCAll	0.098	0.054	-0.007	0.892
	CNCCaus	-0.218**	0.000	-0.228**	0.000
	CNCLogic	-0.084	0.098	-0.208**	0.000
	CNCADC	-0.103*	0.044	-0.052	0.310
	CNCTemp	0.029	0.568	0.073	0.154
	CNCTempx	-0.119*	0.020	-0.045	0.374
	CNCAdd	0.224**	0.000	0.122*	0.017
Referential Cohesion	CRFNO1	0.123*	0.016	0.171**	0.001
	CRFAO1	0.125*	0.014	0.158**	0.002
	CRFSO1	0.178**	0.000	0.205**	0.000
	CRFCWO1	-0.023	0.649	-0.088	0.085
	CRFNOa	0.083	0.103	0.181**	0.000
	CRFAOa	0.075	0.140	0.176**	0.001
	CRFSOa	0.118*	0.021	0.217**	0.000
	CRFCWOa	-0.065	0.203	-0.053	0.295
LSA	LSASS1	0.057	0.267	0.074	0.145
	LSASSp	0.059	0.251	0.126*	0.014
	LSAPP1	0.159**	0.002	0.039	0.446
	LSAGN	0.062	0.226	-0.067	0.190
Situation Models	SMCAUSv	-0.123*	0.015	-0.111*	0.029
	SMCAUSvp	-0.189**	0.000	-0.271**	0.000
	SMINTEp	0.067	0.191	-0.113*	0.026
	SMCAUSr	-0.062	0.224	-0.180**	0.000
	SMINTEr	-0.172**	0.001	-0.056	0.272
	SMCAUSlsa	0.102*	0.046	-0.120*	0.019
	SMCAUSwn	0.197**	0.000	-0.014	0.788
	SMTEMP	-0.042	0.406	0.035	0.493

*Note: ** denotes a correlation at the significance level of 0.01 (2-tailed); * denotes a correlation at the significance level of 0.05 (2-tailed).*

score them as well. JuKu, as an online automated rating system based on corpora and cloud computing, can measure the gap between the students' English composition and exemplary ones from corpora to instantly generate scores and comments for the composition in evaluation. The two human raters received pre-scoring training and uniformly applied the scoring rubric proposed by Jacobs et al. (1981) to ensure criterion consistency. The Cronbach's Alpha coefficient of the scoring results is 0.92, indicating high reliability of the scores. The average of the two raters' scores represents the results of human scoring for each essay, which, together with the JuKu's rating results, served as the measure of the quality of the essay. Subsequently, Coh-Metrix was applied to analyze the textual features of all these argumentative essays, and the analysis results were stored in Excel. Finally, SPSS 26.0 was employed for statistical analysis: Pearson correlation analysis was first conducted to examine the relationship between discourse cohesion and

coherence indicators and writing quality in these essays, followed by multiple linear regression analysis on those indicators showing statistically significant correlations, which led to the establishment of a regression model of the relationships among the variables.

Findings and Discussions

The Relationship between Discourse Cohesion and Coherence and Writing Quality in Argumentative Essays

Table 2 shows that among the aforementioned 27 indicators of discourse cohesion and coherence, 14 exhibited significant correlations with the results of human rating. These included four connective indicators, four referential cohesion indicators, one LSA indicator, and five situation model indicators. The correlation coefficients between each essay's measures of these indicators and the scores from the human raters reached statistical significance, but the correlations were relatively weak, with absolute values ranging from 0.103 to 0.224. In comparison, 15 indicators showed significant correlations with the scores from JuKu. These included three connective indicators, six referential cohesion indicators, one LSA indicator, and five situation model indicators. Similarly, the correlation coefficients between the essays' measures of these indicators and scores from JuKu were statistically significant; nevertheless, the correlations were not strong, with absolute values ranging from 0.111 to 0.271. There were no statistically significant correlations between the other indicators and writing scores in these essays.

Connective Indicators and the Quality of Writing

Different connective indicators exhibit distinct levels of correlations with the quality of writing. Specifically, CNCCaus ($r_{human\ rating} = -0.218^{**}, p < 0.01$), CNCADC ($r_{human\ rating} = -0.103^{*}, p < 0.05$), and CNCTempx ($r_{human\ rating} = -0.119^{*}, p < 0.05$) showed significant negative correlations with the quality of the essays as evaluated by human raters, while CNCAdd ($r_{human\ rating} = 0.224^{**}, p < 0.01$) showed a significant positive correlation with it. Under JuKu's evaluation system, CNCCaus ($r_{JuKu\ rating} = -0.228^{**}, p < 0.01$) and CNCLogic ($r_{JuKu\ rating} = -0.208^{**}, p < 0.01$) were also negatively related to the quality of the essays, and CNCAdd ($r_{JuKu\ rating} = 0.122^{*}, p < 0.05$) was positively related to it. This suggests that frequent use of causal, adversative, and logical connectives may weaken the natural fluency of the text, making the argumentation appear formulaic. The finding echoes the research results of Crossley and McNamara (2011) and Kong (2023). On the other hand, appropriate use of additive connectives helps enhance inter-sentence cohesion and

improve the overall quality of writing. This finding aligns with the research results of Jia and Zhang (2020), which revealed a significant positive correlation between the incidence of additive connectives and writing scores. It also supports Tang's (2006) argument that the instructor needs to direct students on how to legitimately use cohesive devices, particularly connectives. What distinguishes the present study is that it reveals a more complex relationship between connective usage and writing quality. Its research data clearly evidences that appropriate cohesive devices can improve coherence, but over-reliance on them may compromise the quality of discourse (McNamara et al., 2010).

Referential Cohesion Indicators and the Quality of Writing

There were significant positive correlations between the recurrence of nouns, arguments, and stems and the quality of essays. As per both human and JuKu rating results, CRFNO1 ($r_{human\ rating} = 0.123^*$, $p < 0.05$; $r_{JuKu\ rating} = 0.171^{**}$, $p < 0.01$), CRFAO1 ($r_{human\ rating} = 0.125^*$, $p < 0.05$; $r_{JuKu\ rating} = 0.158^{**}$, $p < 0.01$), and CRFSO1 ($r_{human\ rating} = 0.178^{**}$, $p < 0.01$; $r_{JuKu\ rating} = 0.205^{**}$, $p < 0.01$) were all significantly positively related to the quality of the essays. This suggests that moderate lexical overlap helps enhance text coherence, making the representation of content more consistent. Also, regardless of the rating actor (human rater or JuKu), CRFSOa ($r_{human\ rating} = 0.118^*$, $p < 0.05$; $r_{JuKu\ rating} = 0.217^{**}$, $p < 0.01$) was significantly positively related to the quality of the essays. Particularly in JuKu evaluations, CRFNOa ($r_{JuKu\ rating} = 0.181^{**}$, $p < 0.01$) and CRFAOa ($r_{JuKu\ rating} = 0.176^{**}$, $p < 0.01$) showed significant positive correlations with essay scores, further foregrounding this trend. Together, these findings demonstrate that a higher level of referential cohesion contributes to the overall coherence of essays, predicting stronger structure and logic. Yet, they are opposed to the findings of Jia and Zhang's study (2020), which revealed a significant negative correlation between referential cohesion indicators and writing performance. This discrepancy may be the result of multiple factors such as differences in the source of the essay sample, sample size, grading criteria, research design, and methods.

LSA Indicators and the Quality of Writing

There was a positive correlation between LSA indicators and the quality of the essays. Specifically, LSAPP1, an indicator of semantic coherence of adjacent paragraphs, was significantly and positively related to human-rated writing performance ($r_{human\ rating} = 0.159^{**}$, $p < 0.01$), while LSASSp, an indicator of the semantic coherence across all sentences in a text, was significantly and positively related to writing scores from Juku ($r_{JuKu\ rating} = 0.126^*$,

$p < 0.05$). This indicates that the semantic coherence within an essay, i.e., the consistency of content and logical connection between paragraphs or sentences, is a factor affecting writing quality. A higher level of semantic overlap may help enhance the overall coherence of the essay, producing a more explicit representation of the theme. Therefore, optimizing semantic overlap among paragraphs or all sentences could be an effective tactic for enhancing the quality of essays.

Situation Model Indicators and the Quality of Writing

The level of correlation between different situation model indicators and the quality of the essay varies. SMCAUSv ($r_{human\ rating}$ = -0.123*, $p < 0.05$), SMCAUSvp ($r_{human\ rating}$ = -0.189**, $p < 0.01$), and SMINTEr ($r_{human\ rating}$ = -0.172**, $p < 0.01$) were significantly and negatively related to the essay quality rated by human raters, as opposed to SMCAUSlsa ($r_{human\ rating}$ = 0.102*, $p < 0.05$) and SMCAUSwn ($r_{human\ rating}$ = 0.197**, $p < 0.01$) with significant positive correlations with it. At the same time, SMCAUSv ($r_{JuKu\ rating}$ = -0.111*, $p < 0.05$), SMCAUSvp ($r_{JuKu\ rating}$ = -.271**, $p < 0.01$), SMINTEp ($r_{JuKu\ rating}$ = -.113*, $p < 0.05$), SMCAUSr ($r_{JuKu\ rating}$ = -.180**, $p < 0.01$), and SMCAUSlsa ($r_{JuKu\ rating}$ = -.120*, $p < 0.05$) were all negatively related to the writing scores from JuKu. This suggests that overuse of causal and intentional expressions in writing may adversely affect the essay's grade, possibly because higher frequencies of these expressions add a narrative tone to the text, which should be avoided in the argumentative genre. On the other hand, a higher degree of verb semantic overlap is favorable to the coherence of the text, thereby generating higher rating results. These results not only corroborate previous findings on the negative correlation between the frequency of intentional verbs and the quality of the essay (Wang & Zhang, 2020; Jia & Zhang, 2020) but also reveal the potentially adverse effects of causal vocabulary on writing scores. Compared with existing research, this study identifies more situation model indicators that are significantly related to writing quality; it also highlights that the use of causal and intentional expressions and verb overlaps has varied effects on writing scores under different rating frameworks.

To sum up, the results of human and JuKu rating both showed positive correlations between relevant indicators in the dimension of referential cohesion and the quality of the essays, which is consistent with the findings of Liang (2006) and Du and Cai (2013) mentioned in the section of Research Background. In the dimension of connectives, all indicators exhibit negative correlations with the quality of the essays, except for the incidence of additive connectives. In the dimension of LSA, both semantic overlap between adjacent paragraphs and semantic overlap across all sentences were positively related to the quality of writing. In the dimension of situation models, all

relevant indicators showed negative correlations with the quality of the essays as per Juku's evaluations. However, human scoring results showed significant positive correlations of LSA verb overlap and WordNet verb overlap with writing quality. The differences observed between human and Juku rating results may derive from their different evaluation criteria, which is a question deserving further and deeper investigations.

Table 3. Multiple Linear Regression Models of Human Rating.

Models	R	R^2	AdjustedR^2	SE of Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	0.224[a]	0.050	0.048	4.6197100	0.050	20.347	1	384	0.000
2	0.297[b]	0.088	0.083	4.5326862	0.038	15.887	1	383	0.000
3	0.343[c]	0.118	0.111	4.4637940	0.030	12.913	1	382	0.000
4	0.387[d]	0.150	0.141	4.3880669	0.032	14.299	1	381	0.000
5	0.415[e]	0.172	0.161	4.3362607	0.022	10.158	1	380	0.002
6	0.430[f]	0.185	0.172	4.3079722	0.013	6.007	1	379	0.015

a. Predictor: (constant), CNCAdd
b. Predictors: (constant), CNCAdd, CNCCaus
c. Predictors: (constant), CNCAdd, CNCCaus, LSAPP1
d. Predictors: (constant), CNCAdd, CNCCaus, LSAPP1, SMCAUSwn
e. Predictors: (constant), CNCAdd, CNCCaus, LSAPP1, SMCAUSwn, CNCADC
f. Predictors: (constant), CNCAdd, CNCCaus, LSAPP1, SMCAUSwn, CNCADC, CRFSO1

Predictive Effects of Discourse Cohesion and Coherence Indicators on Writing Scores

The study employs multiple linear regression analysis to investigate the predictive effects of discourse cohesion and coherence indicators on the quality of the essay. First, indicators without a significant correlation with writing scores were excluded from the regression analysis, based on the correlation analysis results. Those indicators with a significant correlation with writing scores were treated as independent variables, and the writing score was treated as the dependent variable, for multiple stepwise regression analysis. This method can effectively eliminate multicollinearity between indicators, ensuring an optimal regression equation (Xu, 2013). Correlation matrix analysis revealed significant correlations between some indicators ($r > 0.7$), which might cause biased model estimates or reduced predictive effects. Therefore, to enhance the models' generalizability, those with multicollinearity issues were excluded through collinearity diagnostics. Model selection criteria include a tolerance greater than 1 - R^2 and a variance inflation factor (VIF) less than 2 (Zhang & He, 2022).

Following the selection process, the multiple linear regression analysis of human rating yielded six scoring models (**Table 3**). The data showed that the indicators weighing in human scoring, in descending order, were

CNCAdd, CNCCaus, SMCAUSwn, LSAPP1, CNCADC, and CRFSO1, explaining 5%, 3.8%, 3.2%, 3%, 2.2%, and 1.3% of the variance, respectively

Table 4. Model Coefficients from Linear Regression Model 6 for Human Rating.

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model 6	B	SE	Beta	t	Sig.	Tolerance	VIF
(Constant)	71.389	1.706		41.847	0.000		
CNCAdd	0.079	0.017	0.234	4.689	0.000	0.862	1.160
CNCCaus	-0.072	0.020	-0.177	-3.701	0.000	0.939	1.065
LSAPP1	4.308	1.294	0.160	3.328	0.001	0.930	1.075
SMCAUSwn	7.357	2.124	0.163	3.464	0.001	0.971	1.030
CNCADC	-0.072	0.024	-0.149	-2.977	0.003	0.861	1.161
CRFSO1	2.866	1.169	0.117	2.451	0.015	0.942	1.061

Table 5. Multiple Linear Regression Models of JuKu Rating.

Models	R	R^2	Adjusted R^2	SE of Estimate	R^2 Change	F Change	df1	df2	Sig. F Change
1	0.271[a]	0.073	0.071	4.1945353	0.073	30.450	1	384	0.000
2	0.335[b]	0.112	0.108	4.1108028	0.039	16.803	1	383	0.000
3	0.390[c]	0.152	0.145	4.0230299	0.040	17.895	1	382	0.000
4	0.404[d]	0.163	0.154	4.0022647	0.011	4.974	1	381	0.026
5	0.416[e]	0.173	0.162	3.9836608	0.010	4.567	1	380	0.033
6	0.431[f]	0.186	0.173	3.9575251	0.013	6.036	1	379	0.014

a. Predictor: (constant), SMCAUSvp
b. Predictors: (constant), SMCAUSvp, CRFSOa
c. Predictors: (constant), SMCAUSvp, CRFSOa, SMCAUSr
d. Predictors: (constant), SMCAUSvp, CRFSOa, SMCAUSr, SMINTEp
e. Predictors: (constant), SMCAUSvp, CRFSOa, SMCAUSr, SMINTEp, CNCLogic
f. Predictors: (constant), SMCAUSvp, CRFSOa, SMCAUSr, SMINTEp, CNCLogic, CNCAdd

(see R^2 Change). Through model comparison, Model 6 was identified as the optimal one for predicting writing quality graded by human raters, with the highest coefficient of determination (R^2), explaining 18.5% of the variance.

Table 4 shows model coefficients from regression Model 6 for human rating, which contains six indicators that reached statistical significance ($| t | > 2$, $p < 0.05$) and thus, had significant effects on the writing scores. Additionally, the tolerance and VIF for each variable in Model 6 approximated 1, suggesting no multicollinearity issues among the predictors (Qin, 2003). Based on the unstandardized coefficients in the table, the equation for this model is:

The Essay Score from Human Rating (on a 100-point scale) = 71.389 + 0.079 * CNCAdd – 0.072 * CNCCaus + 4.308 * LSAPP + 7.357 * SMCAUSwn – 0.072 * CNCADC + 2.866 * CRFSO1

Furthermore, regression analysis was also applied to JuKu rating to yield six scoring models (**Table 5**). The data show that the indicators weighing in JuKu scoring, in descending order, were SMCAUSvp, SMCAUSr,

Table 6. Model Coefficients from Linear Regression Model 6 for JuKu Rating.

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model 6	B	SE	Beta	t	Significance	Tolerance	VIF
(constant)	84.296	1.347		62.568	0.000		
SMCAUSvp	-0.061	0.016	-0.182	-3.704	0.000	0.888	1.126
CRFSOa	5.364	1.228	0.208	4.369	0.000	0.947	1.056
SMCAUSr	-2.083	0.595	-0.178	-3.500	0.001	0.835	1.198
SMINTEp	-0.047	0.022	-0.103	-2.138	0.033	0.921	1.086
CNCLogic	-0.047	0.016	-0.157	-2.914	0.004	0.742	1.348
CNCAdd	0.038	0.016	0.124	2.457	0.014	0.888	1.126

CRFSOa, CNCAdd, SMINTEp, and CNCLogic, accounting for 7.3%, 4%, 3.9%, 1.3%, 1.1%, and 1% of the variance, respectively. Model 6, with the largest coefficient of determination (R^2), was identified as the optimal model for JuKu rating, which could explain 18.6% of the variance.

Table 6 shows model coefficients from regression Model 6 for JuKu rating. This model included six predictors, all of which reached statistical significance ($| t | > 2$, $p < 0.05$). Based on the unstandardized coefficients in the table, the regression equation for the model is as follows:

The Essay Score from JuKu Rating (on a 100-point scale) = 84.296 − 0.061 * SMCAUSvp + 5.364 * CRFSOa − 2.083 * SMCAUSr − 0.047 * SMINTEp − 0.047 * CNCLogic + 0.038 * CNCAdd

The results of empirical analyses reveal that discourse cohesion and coherence indicators included in Model 6 had significant predictive effects in assessing writing quality, as they accounted for over 18% of the variance in writing scores (18.5% for human rating and 18.6% for Juku rating). This finding aligns with Du and Cai's (2013) research results, which suggested an evident connection between text coherence and writing scores, with the coherence indicators from Coh-Metrix explaining 17% of the variance in writing scores.

Conclusion

The study demonstrates the correlation between discourse cohesion and coherence and the quality of English argumentative writing, providing insights for further improving EFL writing instruction. Based on its findings, the study proposes instructional recommendations as follows. First, increase training on discourse cohesion and coherence literacy in students. Students

should be guided to develop the understanding that effective discourse relies not only on explicit cohesive devices but also on deeper semantic links and logical coherence. Even in texts lacking obvious cohesive markers, coherence can still be achievable through implicit contextual clues. Moreover, overuse of cohesive devices may lead to redundancy and undermine the quality of the text. In writing instruction, teachers should focus on fostering students' ability to appropriately employ cohesive tactics, teaching them legitimate and accurate use of connectives, references, substitutions, and other cohesive devices. Second, strengthen the teaching of higher-order discourse cohesion and coherence. Teachers can help students comprehend more deeply the functions and applicable contexts of various cohesive and coherent devices by designing and implementing targeted exercises, such as "comparative analysis of cohesive devices" and "discourse coherence remedies." Also, students should be encouraged to use metacognitive strategies to regularly reflect on cohesion and coherence issues in their own writing to progressively cultivate the ability to construct coherent texts. Third, raise students' awareness of other textual features aside from discourse cohesion and coherence (Zhang, 2016). Despite the significant correlations between discourse cohesion and coherence indicators and writing quality, as confirmed by the study, they are far from being the only predictors of writing scores. Therefore, teachers should integrate training on discourse cohesion and coherence with that on other core textual features, such as lexical diversity and syntactic complexity, to construct a multi-dimensional writing skill cultivation framework. It is highly recommended that teachers adopt diverse teaching methods and design integrated writing tasks to foster students' capacity to coordinate vocabulary, syntax, discourse cohesion and coherence in writing essays.

Despite its potential to serve as a valuable reference for future research, the study has its limitations: (i) It addresses only discourse cohesion and coherence, without considering other factors influencing writing quality, such as lexical diversity and syntactic complexity; (ii) The analysis primarily relied on quantitative measures from Coh-Metrix, without fully incorporating the perceptions of human raters or feedback from the automated rating mechanism; (iii) The subjective nature of human rating may have affected the reliability of writing evaluations, even though they were made by two independent raters; (iv) The sample is confined to argumentative essays by Chinese EFL learners, which may constrains the generalizability of the study's findings. Future research should look at a broader scope of text features, increase the number of raters, use a larger sample size with more diverse genres, and combine quantitative and qualitative methods to provide a more comprehensive understanding of the factors that can affect discourse effectiveness, offering deeper insights for the improvement of essay scoring systems and writing instruction.

References

Brooks, C. & Warren, R. P. (2004). Understanding Fiction. Beijing: Foreign Language Teaching and Research Press.

Crossley, S. A. & McNamara, D. S. (2011). Shared features of L2 writing: Intergroup homogeneity and text classification. Journal of Second Language Writing, 20(4), 271-285. DOI: https://doi.org/10.1016/j.jslw.2011.05.007

Du, H. & Cai, J. (2013). A Coh-Metrix-based prediction model of the quality of argumentative essays by Chinese English learners. Modern Foreign Languages, 36(3), 293-300. Available at: https://kns.cnki.net/kcms2/article/abstract?v=-w8R4t09JVUlQYuGlQOBOrVmxzz6Htk0Ra9Sj-LUcM3JOoYb1AnVNuERWW2DOTxamSZT9SeKsOcbWRLxYwrqdR37J5zMeljKm1s0VzFv70UT0rtzDwYSMtgStg9bgJJs8u-CODNt_Kp9D3dP4ldGEpMMUx9bWxq9vsS_FgkwfFQ=&uniplatform=NZKPT&language=CHS

Graesser, A. C., Jeon, M., Yan, Y., & Cai, Z. (2007). Discourse cohesion in text and tutorial dialogue. Information Design Journal, 15(3), 199-213. DOI: https://psycnet.apa.org/doi/10.1075/idj.15.3.02gra

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: providing multilevel analyses of text characteristics. Educational Researcher, 40(5), 223-234. DOI: https://psycnet.apa.org/doi/10.3102/0013189X11413260

Halliday, M. A. K. & Hasan, R. (1976). Cohesion in English. London: Longman.

Hu, Z. (1994). Cohesion and Coherence in Discourse. Shanghai: Shanghai Foreign Language Education Press.

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). Testing ESL composition: A practical approach. English composition program. Newbury House Publishers, Inc., Rowley, MA 01969.

Jia, W. & Zhang, P. (2020). The Relationship Between textual features and writing quality of essays from TEM-8 writing tests. Foreign Language Testing and Teaching, 2020(1), 1-9. DOI: https://doi.org/10.26970/j.cnki.fltt.1004-5112.2020.01.001.

Jiang, J. (2016). The application of Coh-Metrix in foreign language teaching and research. Foreign Language in China, 2016(5): 58-65.

Kong, W. (2023). Verification of the validity of the English Proficiency Scale for Chinese Learners based on the English Language Learner Corpus: Using the Cohesion Knowledge Application Ability Scale as an Example. Foreign Language Testing and Teaching, 2023(2), 29-38. DOI: https://doi.org/10.26970/j.cnki.fltt.1004-5112.2023.02.004.

Liang, J. (2019). A Coh-Metrix-based comparative study of discourse coherence in Chinese and American English News. Journal of Heilongjiang University of Technology, 2019(8), 139-143. DOI: https://doi.org/10.3969/j.issn.1672-6758.2019.08.027

Liang, M. (2006). Coherence of written discourse of foreign language learners. Modern Foreign Languages, 2006(3), 284-292. DOI: https://doi.org/10.3969/j.issn.1003-6105.2006.03.008

Louwerse, M. M, McCarthy, P. M., McNamara, D. S. & Graesser, A. C. (2004). Variation in language and cohesion across written and spoken registers. In K. D. Forbus, D. Gentner, & T. Regier (eds.). Proceedings of the 26th Annual Meeting of the Cognitive Science Society (pp. 843-848). Mahwah, NJ: Erlbaum.

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). Automated Evaluation of Text and Discourse with Coh-Metrix. New York: Cambridge University Press.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. Discourse Processes, 47, 292-330. DOI: https://doi.org/10.1080/01638530902959943?urlappend=%3Futm_source%3Dresear

chgate.net%26utm_medium%3Darticle
Peng, H. (2017). An Empirical Study on Coherence of Writing by EFL Learners. Journal of PLA University of Foreign Languages, 2017(7): 87-92.
Qin, X. (2003). Quantitative Data Analysis in Foreign Language Teaching Research. Wuhan: Huazhong University of Science and Technology Press.
Richards, J. C., Schmidt, R., Kendrick, H., & Kim, Y. (2005). Longman Dictionary of Language Teaching and Applied Linguistics (Third Edition). Beijing: Foreign Language Teaching and Research Press.
Tang, S. (2006). Issues about coherence in college students' English writing. Journal of Zhanjiang Normal University, 2006(2), 106-108. DOI: https://doi.org/10.3969/j.issn.1006-4702.2006.01.023
Wang, J. & Zhang, T. (2020). The relationship between quantitative indicators and machine-generated scores of second-language writing. Foreign Language Testing and Teaching, 2020(3), 12-20. DOI: https://doi.org/10.26970/j.cnki.fltt.1004-5112.2020.03.002
Wen, J. (2013). Integrated realization of cohesion and coherence in texts. Foreign Language Research, 2013(4), 107-112.
Wen, Q., Liang, M., & Yan, X. (2008). Spoken and Written English Corpus of Chinese Learners (SWECCL 2.0). Beijing: Foreign Language Teaching and Research Press.
Xu, H. (2013). Cases of Statistical Analysis in Second Language Learning Research. Beijing: Foreign Language Teaching and Research Press.
Zhang, D. & Liu, R. (2018). The Development of the Theory of Text Coherence and Cohesion and Its Application (Second Edition). Shanghai: Shanghai Foreign Language Education Press.
Zhang, G. & He, F. (2022). Reliability and validity of an automated scoring system for English compositions: A quantitative Analysis of textual features in different types of writing tasks. Foreign Language Testing and Teaching, 2022(01), 44-56. DOI: https://doi.org/10.26970/j.cnki.fltt.1004-5112.2022.01.006
Zhang, J. (2016). The development of discourse coherence in college English compositions: A study based on Coh-Metrix. The World and Chongqing, 33(03), 66-72. DOI: https://doi.org/10.13769/j.cnki.cn50-1011/d.2016.03.014
Zhou, W. (2023). Coherence in continuation writing in senior secondary English education: A study based on Coh-Metrix. Foreign Language Testing and Teaching, 2023(2), 39-47. DOI: https://doi.org/10.26970/j.cnki.fltt.1004-5112.2023.02.005.
Zhu, H. (2007). Discourse Coherence and Structural Cohesion. Journal of Ningde Teachers College (Philosophy and Social Sciences Edition), 83(4), 58-62.

Received: October 10, 2025
Revised: December 15, 2025
Accepted: January 30, 2026