

# *A Study on the Development of Cohesion in English Majors' Expository Writing*

**Hong Zeng**

*Foreign Languages College, Jiangxi Normal University, Nanchang, 330022, China*

**Keywords:** Cohesion, expository writing, English majors, Coh-Metrix

**Abstract:** This study analyzed 54 expository writings of 27 English majors through 20 cohesive indices provided by Coh-Metrix, aiming to explore the development of cohesion in their expository writings after one academic year. Results showed that: 1) English majors are able to flexibly use various cohesive devices to achieve textual cohesion in expository writing. The frequency of using connectives is the highest while the use of referential cohesion is relatively low. And the global coherence of their writing is at a low level as well. 2) With the increase of learning time, 10 cohesive indices show significant changes. Among them, 6 indices such as adjacent argument overlap, adjacent stem overlap, argument overlap, stem overlap, LSA similarity between adjacent paragraphs, and LSA Given/New have increased, while the other 4 indices such as LSA similarity between adjacent sentences, causal connectives, extended temporal connectives and causal cohesion have decreased. Through enriching longitudinal research on cohesion, it is hoped to promote English writing teaching and learning.

## **1. Introduction**

Cohesion is a fundamental prerequisite in the formation of texts and a crucial factor influencing discourse coherence as well. Appropriate application of cohesive devices in writing can enhance the coherence, thereby achieving clarity of expression, logical consistency, and structural integrity. However, during the writing process, second language (L2) learners often prioritize linguistic form and accuracy over cohesion. Their writings may exhibit a lack of coherence between sentence and sentence, or paragraph and paragraph, resulting in loosely structured texts.

In recent years, both domestic and foreign scholars have investigated differences in the use of various cohesive devices among students of different grades or proficiency levels, or explored the relationship between the use of cohesive devices and writing quality. It is evident that most of them are cross-sectional studies. But comprehensive description of how the use of cohesive devices changes over time within the same individuals would better enable teachers to grasp the developmental characteristics or patterns in students' writing. Therefore, this study will use expository writings by English majors as the corpus and utilize various indices provided by Coh-Metrix to measure and analyze cohesion of these texts, in order to thoroughly explore the developmental trends of cohesion in L2 students' writing over time.

## 2. Literature Review

### 2.1 Research on Cohesive Features in Writing

Some scholars have utilized Coh-Metrix to analyze the cohesive features of writing texts. For instance, McNamara et al.<sup>[1]</sup> examine cohesive features of essays from multiple levels and find that Coh-Metrix metrics could significantly differentiate texts between high and low cohesion, thus confirming its effectiveness as a tool for measuring discourse cohesion. Zhang et al.<sup>[2]</sup> compare the cohesion in argumentative writing between Chinese English learners and native English speakers, discovering that the former has a significantly higher use of logical connectives and temporal connectives than the latter. Although there is no difference in local lexical cohesion of their writings, significant differences are observed in several measures of global lexical cohesion.

### 2.2 Influence of Cohesive Features on Writing Quality

Numerous studies have emerged abroad exploring the relationship between cohesive devices employed by learners in writing and the quality of their writing. Scholars have found that features indicative of local cohesion show no correlation or a negative correlation with writing quality, whereas features indicative of global cohesion demonstrate a positive correlation with writing quality. Liang et al.<sup>[3]</sup> point out that certain measures of local cohesion, like content word overlap in adjacent sentences, positive connectives and LSA similarity between adjacent sentences, exhibit a negative correlation with writing scores. In contrast, LSA similarity between adjacent paragraphs is positively correlated with writing quality. And Crossley et al.<sup>[4]</sup> indicate that the use of local, global and textual cohesion explains 42% of the variance in overall judgments of writing quality.

The conclusions drawn by domestic scholars show some differences from those in foreign research. Liang<sup>[5]</sup> conducts a study on the coherence of students' written language based on Coh-Metrix. The results suggest that some cohesive features cannot predict writing scores, but local cohesive features such as adjacent argument overlap, adjacent stem overlap and LSA similarity between adjacent sentences are significantly correlated with students' writing scores. And high-scoring essays employ global cohesive devices more effectively, whereas low-scoring essays rely more on local cohesive devices. Similarly, Du & Cai's<sup>[6]</sup> results are largely consistent with Liang's, but they also discover that 10 indices about connective are not correlate with writing scores. Thus they consider that connective, as a superficial cohesive device, do not reflect the coherence of mental representation of texts.

### 2.3 Research on the Development of Cohesion in Writing

Studies on the longitudinal development of cohesion in students' writing are limited. Crossley et al.<sup>[4]</sup> investigate the development of cohesive features in descriptive essays written by 57 L2 learners at the beginning, middle, and end of a semester. Their findings reveal that 13 indicators exhibit significant linear trends. And the level of cohesive devices used in students' essays during middle and end of the semester is higher compared to the initial stage, but there is no significant difference between the middle and end of the semester. Granados and Lorenzo<sup>[7]</sup> find that there is a 15% increase in the overall use of connectives in students' writing after three years, indicating an improvement in their writing proficiency. Abdi Tabari & Wind<sup>[8]</sup> employ Complex Dynamic Systems Theory to explore the changes in cohesive devices of L2 writing in more detail. They identify eight cohesive measures that significantly improved but display non-linear trends.

In domestic research, Zhang<sup>[9]</sup> conducts a study on the development of coherence in writing, analyzing English essays written by 30 non-English majors over four semesters with four

quantitative indices of cohesion provided by Coh-Metrix — the ratio of connectives, content word overlap, LSA similarity between all possible pairs of sentences, tense and aspect repetition. It is found that these four indices improved with increased learning time. And he emphasizes that students should not excessively use cohesive devices merely to achieve so-called coherence but should focus on the synchronous and coordinated development of discourse coherence with other linguistic features. In the same year, Zhang <sup>[10]</sup> also conducts a case study over two semesters, revealing that development in different cohesive dimensions shows non-linear trends and that individual differences in development of coherence are influenced by various factors, demonstrating complexity and unpredictability. And Zhu et al. <sup>[11]</sup> investigate the characteristics of development and variation of cohesive devices with stem overlap, connectives, LSA, tense and aspect repetition.

## 2.4 Summary

A review of the existing literature reveals that current empirical studies using Coh-Metrix tool to analyze cohesion in writing are predominantly cross-sectional in nature. And research exploring the longitudinal development of cohesion in writing is relatively scarce. Accordingly, this study will undertake a longitudinal investigation over two semesters into cohesion in expository writing by English majors, which aims to explore the dynamic developmental characteristics and trends of cohesion in students' writing using Coh-Metrix. Thus, following research questions are formulated:

- 1) What are the current characteristics of cohesion in expository writing by English majors?
- 2) After one academic year, whether the cohesion of English majors' expository writing has changed? If so, what are the trends?

## 3. Method

### 3.1 Data Collection

The data were sourced from English essays written by 32 English majors from a university in Jiangxi Province over the span of one academic year (two semesters). These essays were part of the students' unlimited writing exercises, all falling under the genre of expository writing and required to be no less than 200 words. Given the extended duration of the investigation, essays completed at the beginning of the first semester and at the end of the second semester were collected. As 5 students did not submit their final writing exercise on time, actual corpus for this study consisted of 54 expository essays from 27 students.

### 3.2 Tools and Measures

The tools employed in this study are Coh-Metrix and SPSS. Coh-Metrix, developed under the direction of Professor McNamara from the University of Memphis, is a web-based text analysis tool designed to perform automatic analyses of linguistic features in both spoken and written texts. The latest version, Coh-Metrix 3.0, encompasses 11 modules and 106 measures related to lexical, grammatical and cohesive features.

Based on the aim of this study and the commonly used measures in existing research, 20 cohesive measures from Coh-Metrix are selected to analyze students' writing texts, as follows: (1) adjacent argument overlap, adjacent stem overlap, argument overlap and stem overlap from "Referential Cohesion" module; (2) LSA similarity between adjacent sentences, LSA similarity between all possible pairs of sentences in a paragraph, LSA similarity between adjacent paragraphs and LSA Given/New from "Latent Semantic Analysis (LSA)" module; (3) all connectives, causal connectives, logical connectives, adversative and contrastive connectives, temporal connectives,

extended temporal connectives, additive connectives, positive connectives and negative connectives from “Connectives” module; (4) causal cohesion and intentional cohesion from “Situational Model” module; (5) second language readability score from “Text Readability and Easability” module.

### 3.3 Data Analysis

Initially, the 54 collected texts are manually transcribed into .txt files and imported them one by one into Coh-Metrix, then save the output data. Subsequently, relevant values of the aforementioned 20 measures are extracted and organized into an Excel spreadsheet in preparation for further statistical analysis. To examine the overall characteristics of cohesion in students' expository writings, descriptive statistical analyses of the selected measures are conducted using SPSS software. This analysis yielded the mean and standard deviation of these 20 measures for the initial writings. Following this, paired-samples T tests are performed on the values of these 20 measures from the two sets of expository writings, comparing the differences between them in order to investigate changes or developmental trends in cohesion in students' expository writing over the course of one academic year.

## 4. Results and Discussion

### 4.1 Current Characteristics of Cohesion in Expository Writing

The results show that English majors are able to employ various cohesive devices in expository writing, with connectives being used most frequently but significant individual difference. Among these, causal connectives (e.g., because, so) are the most frequently used, followed by logical connectives (e.g., and, or). This phenomenon indicates that students tend to use relatively simple causal and logical connectives to link different facts, viewpoints and explanations, while less employing other semantic relations such as adversative or additive connectives. Variation in the use of connectives and transitional words in essays may result from different writing purposes and prompt requirements. When it requires students to explain personal opinions rather than compare different objects, students are inclined to utilize causal, logical and temporal connectives to ensure clarity and coherence in their writings.

And students are less adept at using referential cohesion (adjacent argument overlap, adjacent stem overlap, argument overlap and stem overlap), with the mean values for these four indices all being less than 0.5 (the maximum value being 1). Chinese lacks morphological inflections, which could not utilize a rich array of affixes to form new words centered on arguments or stems like English does for enhancing cohesion between sentences<sup>[12]</sup>. Consequently, due to interference from mother-tongue thinking or limited proficiency in the target language, students are unable to flexibly change vocabulary forms and retrieve words from their mental lexicon to achieve textual cohesion, resulting in medium-low performance in these cohesive devices.

Additionally, the mean values of metrics within LSA module, which reflects textual cohesion, are relatively lower. Specifically, the maximum values for LSA similarity between all possible pairs of sentences in a paragraph, LSA similarity between adjacent paragraphs and LSA Given/New are all less than 0.5. It indicates that even the student with the highest proficiency exhibits relatively low level of coherence in their writing. This may be due to a deficiency in the knowledge of discourse cohesion and coherence or insufficient language proficiency, which restricts the allocation of cognitive resources and attention to semantic coherence and discourse structure, and then resulting in an overall lower level of coherence<sup>[5]</sup>.

## 4.2 Trends of Cohesion in Expository Writing

After an academic year, there has been an increase in the use of cohesive devices in students' expository writing, with their essays showing stronger cohesion. Specifically, the most notable improvement is referential cohesion. And there has been a slight increase in LSA similarity between adjacent paragraphs and LSA Given/New, which is in line with previous studies<sup>[4, 8, 11]</sup>. However, LSA similarity between adjacent sentences, causal connectives, extended temporal connectives and causal cohesion have all shown a declining trend, as shown in Table 1. And the remaining ten measures exhibit no significant differences.

Table 1: Paired-samples T tests of cohesive measures of initial and final writing.

Measures	Initial writing		Final writing		Mean difference	t	Sig.
	Mean	SD	Mean	SD			
Adjacent argument overlap	0.484	0.141	0.632	0.138	0.149	-4.494	.000**
Adjacent stem overlap	0.419	0.153	0.662	0.179	0.243	-7.205	.000**
Argument overlap	0.315	0.072	0.556	0.132	0.241	-8.752	.000**
Stem overlap	0.282	0.094	0.577	0.175	0.295	-9.514	.000**
LSA similarity between adjacent sentences	0.370	0.123	0.254	0.157	-0.116	3.273	.003*
LSA similarity between adjacent paragraphs	0.194	0.075	0.232	0.071	0.038	-2.791	.010*
LSA Given/New	0.187	0.062	0.252	0.063	0.065	-3.943	.001*
Causal connectives	105.494	21.880	89.716	15.942	-15.778	3.589	.001*
Extended temporal connectives	23.901	8.612	15.716	8.528	-8.185	3.984	.000**
Causal cohesion	0.543	0.450	0.300	0.200	-0.243	2.585	.016*

Note: \*  $p < 0.05$ ; \*\*  $p < 0.01$ .

The significant increase in the indices of “Referential Cohesion” module may be attributed to two factors. First, through continuous depth of the professional knowledge, students gradually overcome the negative transfer of Chinese in English writing. They increasingly adapt to and become accustomed to English expressions, and can flexibly using derivation and inflection of words to represent semantic relationship. As a result, they employ more argument and stem overlap to enhance cohesion. Second, due to the improvement in their L2 proficiency and the accumulation of new knowledge, the growing depth and breadth of their vocabulary also contribute to a more flexible use of referential cohesive devices.

And there are significant differences between initial and final writings in the LSA similarity between adjacent sentences, LSA similarity between adjacent paragraphs and LSA Given/New. However, there is no notable change in LSA similarity between all possible pairs of sentences in a paragraph, which contrasts with Zhang's<sup>[9]</sup> research. Although the change of this index is minimal, the significant increase in the mean values of LSA similarity between adjacent paragraphs and LSA Given/New indicates a development of cohesion and coherence in the essays by English majors. It suggests that students are increasingly focusing on semantic connections and global coherence in their writing. The observed decrease in LSA similarity between adjacent sentences may be attributed to students' jumping mind, leading to reduced coherence between adjacent sentences.

Lastly, the majority of measures within “Connectives” module exhibit no significant difference, which purports the findings of Crossley et al.<sup>[4]</sup>. The excessive use of connectives does not contribute to the construction of textual coherence<sup>[5]</sup>, and these English majors might have become aware of this issue, thereby effectively monitoring their use of connectives. Similarly, Zhang's<sup>[9]</sup> analysis of the developmental trends in the ratio of connectives in students' writing also indicates that differences in this index across each semester are not significant. But he predicts that with increased learning time, there will be a noticeable trend in the use of connectives in students' essays.



Nonetheless, he also highlights the issue of overuse and even misuse of connectives in English writing by university students. Further analysis reveals that only causal connectives and extended temporal connectives exhibit a clear downward trend over time. But Granados & Lorenzo<sup>[7]</sup> find English majors reduce their use of extended temporal connectives while significantly increasing causal connectives after three years of bilingual instruction. The reason for this discrepancy could be the different genres of writing involved, with the former being expository and the latter narrative.

## 5. Conclusion

This study examines the development of cohesion in expository writing by 27 English majors with the help of Coh-Metrix. The findings reveal that there is an increase in the use of cohesive devices in students' essays after one academic year, with them showing greater cohesion. Out of the 20 cohesive measures, 10 indices have changed significantly. Among these, 6 indices (adjacent argument overlap, adjacent stem overlap, argument overlap, stem overlap, LSA similarity between adjacent paragraphs and LSA Given/New) demonstrate a growing trend, while 4 indices (LSA similarity between adjacent sentences, causal connectives, extended temporal connectives and causal cohesion) exhibit a declining trend. Overall, textual cohesion and discourse coherence of students' expository writings have been developed to a certain extent. In writing instruction, apart from focusing on the accuracy and complexity of linguistic forms, teachers should also guide students to learn and exert the knowledge of cohesion, particularly in the cultivation of students' latent semantic logic and coherence.

There are still certain limitations in this study. Firstly, the sample size is relatively small. Secondly, the time span of the study is relatively short. A longitudinal study with multiple time points over a longer period might yield more detailed and accurate results. Additionally, it only examines 20 cohesive measures from Coh-Metrix. Further studies should consider a broader range of quantitative indices to provide a more comprehensive reflection of the developmental trends or characteristics of cohesion in students' writing.

## References

- [1] McNamara, D.S., Louwerse, M.M., McCarthy, P.M. and Graesser, A.C. (2010) *Coh-Metrix: Capturing Linguistic Features of Cohesion*. *Discourse Processes*, 47, 292-330.
- [2] Zhang, M.M., Ding, Y.H. and Tao, Q.Q. (2024) *A Corpus-based Study on Discourse Cohesion in Chinese English Learners' Argumentative Essays*. *Journal of Hunan Institute of Engineering (Social Science Edition)*, 34(01), 65-69.
- [3] Liang, G., Crossley, S.A. and McNamara, D.S. (2013) *Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study*. *Assessing writing*, 18(3), 218-238.
- [4] Crossley, S.A., Kyle, K. and 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.
- [5] Liang, M.C. (2006) *Research on Coherence in EFL Learners' Writing*. *Modern Foreign Languages*, (03), 284-292+330.
- [6] Du, H.Y. and Cai J.T. (2013) *A Coh-Metrix-based model of linguistic features predicting argumentative writing quality of EFL learners*. *Modern Foreign Languages*, 36(03), 293-300+331.
- [7] Granados, A. and Lorenzo, F. (2021) *English L2 connectives in academic bilingual discourse: A longitudinal computerised analysis of a learner corpus*. *Revista signos*, 54(106), 626-644.
- [8] Abdi Tabari, M. and Wind, A.M. (2023) *Dynamic development of cohesive devices in English as a second language writing*. *International Review of Applied Linguistics in Language Teaching*.
- [9] Zhang, J.H. (2016) *Research on Development of Coherence in University English Writing Based on Coh-Metrix*. *World & Chongqing*, 33(03), 66-72.
- [10] Zhang, J.H. (2016) *A case study on the development of EFL learners' written discourse coherence from DST perspective*. *Foreign Language Education in China*, 9(03), 47-55+96.
- [11] Zhu, H.M., Tang, J.H. and Zhang, M.Y. (2022) *A Case Study of the Dynamic Development of Cohesive Devices in Second Language Writing*. *Journal of Linyi University*, 44(06), 119-129+144.
- [12] Li, D.D. (2015) *A Comparative Study of Affixations between English and Chinese from the Perspective of Pragmaticalization*. *Foreign Languages in Chin*, 12(04), 38-44.