

The Changes in Text Cohesion of Senior High School Students Measured by Coh-Metrix as a Function of Grade Level

Jinxin Wen

School of Foreign Languages, China West Normal University, Sichuan Nanchong 637000, China

Abstract

National English Curriculum Standards for General High School sets requirements for text cohesion and text cohesion is an essential component of English writing and English writing evaluation. However, previous studies explored text cohesion mainly from perspectives of college students, horizontal comparison and traditional tools. In this study, a corpus of 300 essays by grade level (senior one and senior three) are analyzed computationally by Coh-Metrix to examine whether text cohesion improves as grade level rises. During the specific process, this paper aims to identify characteristics and discriminate differences between the text cohesion created by the two grades. The main instrument is Coh-Metrix—a type of computational tool that offers over 200 measurements to quantify texts. This study mainly employs the measures under the category of referential cohesion, LSA and connectives of it to assess text cohesion. The results demonstrate that as the grade level increases, text cohesion gets better. Specifically, these students produce more noun overlap of adjacent sentences, argument overlap of adjacent sentences, noun overlap of all sentences, argument overlap of all sentences, stem overlap of all sentences, LSA overlap of adjacent sentences, LSA overlap of adjacent paragraphs, LSA overlap of given/new sentences all connectives and additive connectives.

Keywords

Text cohesion, Coh-Metrix, English writing, Senior high school, Grade level.

1. Introduction

As the beginning part of this study, there are some basic information: background, purpose and significance as well as the layout of this paper.

National English Curriculum Standards for General High School (2017 edition, 2020 revision) requires: senior high school students should be equipped with abilities of using discourse cohesion means to form written discourse which express meanings and embody the logical relevance of meanings. Obviously, text cohesion plays an important role in English writing. When the text cohesion is good in one text, readers can understand meanings and ideas better which the author wants to convey. At the same time, it is widely seen as an important part of writing evaluation. However, in real teaching practices, many teachers just consider students' text cohesion from an abstract angle. Fortunately, the development of computational linguistics and corpus open up the way to more comprehensive research of text cohesion. Particularly, the appearance of Coh-Metrix brings possibility for quantifying text cohesion.

The purpose of this study is using Coh-Metrix to examine whether text cohesion improves with increasing grade level. Specifically, this paper identifies characteristics and differences of text cohesion on the same topic in two senior high school grades. Based on the findings, this paper is of interest to senior high school English teachers, because it can offer some new ideas for them to teach and assess English writing.

2. Literature Review

For a research, it is necessary to explain some concepts and learn about previous studies of the topic at the beginning.

(1) Text Cohesion

Text cohesion is one of the logic relationships which can effectively combine propositions and thoughts of text (Liang, 2006). Usually, it can be divided into two categories: local cohesion and global cohesion. Local cohesion refers to the use of noun phrases in place of pronouns, the definition of new terms, the addition of argument overlap between sentences, and the addition of connectives to make the relationships between ideas more clear. The term "global cohesion" refers to the practice of starting each paragraph with a topic sentence and adding topic headers to sections. The degree of difficulty and the reader's comprehension of a text are both crucially influenced by cohesion (McNamara et al., 2014).

A large proportion of previous studies on text cohesion are theoretical research. The origin, ideas, cognitive foundation, and relationship to other terms of text cohesion are all thoroughly explored (Guo, 2003; Miao, 1998). Certainly, scholars have not ignored empirical research. Liang Maochen measured students' text cohesive ability of 120 English writing divided into high grades group and low grades group through Coh-Metrix. He concluded that while there are substantial differences in students' overall writing abilities, the development of students' discourse cohesion ability and overall writing ability is essentially synchronous (Liang, 2006). Shang Yanzhi (2015) summarized cohesive errors frequently made by students in English writing and proposed teaching strategies to cultivate students' writing ability starting from textual cohesion issues. Beck, MCKeown, Omanson and Pople (1984) investigated the benefits of increasing the ease of texts for children. They revised two narrative texts aiming to solve surface, knowledge and content problems. Finally, their paper showed that all of the readers could benefit from the processing of the texts and text cohesion could improve readers' comprehension and memory of texts.

To be short, text cohesion of English writing received a slightly low attention (Xie, 2020). Although, some researchers have investigated text cohesion, their research hardly employs computational tools. There is still a room to investigate text cohesion with advanced instruments.

(2) Coh-Metrix

Coh-Metrix, a web-based text analysis tool developed by McNamara et al. from the University of Memphis in the United States, provides more than 200 measures and can evaluate surface and deep characteristics of texts (McNamara et al., 2014). Domestic and foreign studies based on Coh-Metrix have focused on difficulty of reading discourse, automated writing evaluation, prediction of writing quality and texts' characteristics.

For some scholars, Coh-Metrix is a useful tool to research discourses' difficulty of exams or textbooks. Jiang Jinlin and Han Baocheng (2018) made use of Coh-Metrix to compare difficulty of 163 reading texts from CET 6, TOEFL and IELTS. Their findings was that texts of CET 6 were easier than others and different from the two tests in four dimensions: narrative, connectivity, lexical specificity and referential cohesion. Chen Anni and Guo Aiping (2019) researched the change in difficulty of the first four volumes of READING AND WRITING of NEW HORIZON from lexical, syntactic and cohesive dimensions through Coh-Metrix. They found that difficulty progression was not obvious both from the perspective of macroscopic view and language characteristics. These studies can provide reference basis for exam development, reading grading, and textbook compilation.

The studies of automated writing evaluation based on Coh-Metrix specifically consider the reliability and validity of a certain automated essay scoring system or its application for

constructing a scoring system. Zhang Guoqiang (2022) utilized Coh-Metrix and some other tools to analyze lexical complexity, syntactic complexity and text cohesion of 826 texts from CET-6. He constructed predictive models and compared the predictive variables to examine whether the reliability and validity of automated essay scoring were different between different types of writing tasks. As for S Latifi and M Gierl (2021), they researched nine feature categories of Coh-Metrix for developing prompt-specific AES scoring models. Their findings demonstrated that Coh-Metrix features could be used to support educational goals and scoring goals in large-scale language assessments.

As for prediction of writing quality, researchers always construct models to achieve their goals. McNamara et al. (2010) investigated 120 native speakers' essays and pointed out that high quality writings are with various and uncommon words as well as more complex sentences. From Du Huiying and Cai Jinheng (2013), they utilized Coh-Metrix to discover language characteristics which can affect English major writing quality of arguments. Consequently, a predictive score model was carried out. Jia Wenfeng and Zhang Peixin (2020) discussed the relationships between different text features and writing quality. Their conclusion was that text fluency, syntactic complexity, cohesion, situational models and other characteristics were relevant to essays' quality. More importantly, discourse features could effectively predict writing quality.

Coh-Metrix is also employed to identify texts' characteristics. Arthur C. Graesser, Danielle S. McNamara, and Jonna M. Kulikowich (2011) discussed five major factors that accounted for most of the variance in texts. They found out that texts could be quantified by Coh-Metrix and it was helpful for assigning the right text to the right student at the right time. Based on their findings, Andrew Elfenbein (2011) researched the usefulness of Coh-Metrix measuring textual features from a broader way. In his opinion, Coh-Metrix provided more possibilities for researchers to identify the roles of text features in comprehension for different populations at different moments in development.

Briefly speaking, although some researchers have investigated text characteristics utilizing Coh-Metrix, text cohesion is just seen as a small part in their papers. Their description or analysis of text cohesion is relatively general.

In conclusion, the research on text cohesion based on Coh-Metrix is still not enough. Although some scholars have studied this topic, the participants of previous studies are almost university students while senior high school students are ignored. Meanwhile, previous studies of text cohesion have payed attention to theoretical aspects, relation with other variances or horizontal comparison between different participants or genres while longitudinal characteristics and differences are ignored. Therefore, the research, which uses Coh-Metrix to identify features and differences of text cohesion of high school students in a more specific and comprehensive way to finally examine whether text cohesion improve with increasing grade level, is still needs to be done.

3. Methodology

(1) Research questions

This essay seeks to determine whether text cohesion varies with increasing grade level of senior high school students. Specifically, the research questions are as follows:

1. What are the characteristics of text cohesion in the two senior high school classes?

Are there any differences of text cohesion in the two senior high school grades and if so, what are they?

(2) Research Materials

The corpus of this paper consists of 300 texts from Pigaiwang which are on the same topic and belong to two senior high school grades in Jiangsu province (150 texts each grade). The topic is "Work Hard To Make Our Dreams Come True" and students need to write essays with a length from 300 to 500 words without limit on specific topic and genre.

(3) Research Tools and Measures

Developments in computational linguistics and discourse processing provide possibilities for researchers to measure text with a wide range of indices. These indices have been gathered in Coh-Metrix 3.0 (Graesser et al., 2004). It can quantify lexical, syntactic and discourse characteristics and covers 11 parts: descriptive statistics, text easability principle component scores, referential cohesion, LSA (Latent Semantic Analysis) (Landauer et al., 2007), lexical diversity, connectives, situation model, syntactic complexity, syntactic pattern density, word information and readability. Also, Coh-Metrix overcomes the shortcomings that manual analysis are too tedious and subjective and improve quality of research results to some extent. To just measure text cohesion, this paper generally adopts the idea of Jiang Jinlin (2016) that referential cohesion, LSA and connectives are directly relevant to text cohesion. First, referential cohesion refers to overlap in content words between local sentences or coreference which vary along local cohesion and global cohesion in Coh-Metrix (McNamara et al., 2014). Second, LSA is a advantage of Coh-Metrix, which surmounts the statistics of surface features and constructs a hidden semantic space of text. Third, connectives matters in creating cohesive connections between ideas and clauses and provide clues to the organization of texts (Cain & Nash, 2011).

Table 1. Categories and indices of Coh-Metrix directly related to text cohesion

Category	Indices	Number
Referential cohesion	Noun overlap mean of adjacent sentences	8
	Argument overlap mean of adjacent sentences	
	Stem overlap mean of adjacent sentences	
	Noun overlap mean of all sentences	
	Argument overlap mean of all sentences	
	Stem overlap mean of all sentences	
LSA	Content word overlap mean of adjacent sentences	4
	Content word overlap mean of all sentences	
	LSA overlap mean of adjacent sentences	
	LSA overlap mean of all sentences in paragraphs	
	LSA overlap mean of adjacent paragraphs	
Connectives	LSA overlap mean of given/new sentences	9
	All connectives incidence	
	Casual connectives incidence	
	Logical connectives incidence	
	Adversative and contrastive connectives incidence	
	Temporal connectives incidence	
	Expanded temporal connectives incidence	
	Additive connectives incidence	
	Positive connectives incidence	
	Negative connectives incidence	

(4) Research Procedures

Corpus collection

Nearly ten thousand texts from different provinces under the topic of “Work Hard To Make Our Dreams Come True” were downloaded from Pigaiwang. Among the several provinces, only texts from Jiangsu had relatively large and equal capacity between two grades. Then, 300 texts (150 texts each grade) were selected after eliminating texts that did not meet the word count requirements, language or content fragmentation, and punctuation confusion. More than 100 texts of each grade is a relatively sizeable and representative sample capacity, which can make the findings of this study more reliable and persuasive. Finally, 300 essays were input in the form of txt. , numbered, and constitute a corpus.

Data Collection and Analysis

The 300 texts were respectively input into Coh-Metrix 3.0 and 106 indices were displayed. Data was output in the form of txt.. Then, it was cleaned and stored into Excel. What’ more, descriptive statistics, independent sample t-test and Mann-Whitney U test of SPSS27.0 were employed to analyze data of text cohesion.

4. Results

This paper aims to identify characteristics and differences of text cohesion with increasing grade level. In order to address this goal, this study conducted descriptive statistics, independent sample t-test and Mann-Whitney U test through SPSS. Two indices (Positive connectives incidence and Negative connectives incidence) are all showed with zero of every essay, so the following part will not consider them anymore

Table 2. Analysis of One-sample T Test results for selected indices of grade one

Variable	K9-10		Grade one		P
	Mean	Standard deviation	Mean	Standard deviation	
Referential Cohesion					
Noun overlap mean of adjacent sentences	0.397	0.186	.227	.128	.169
Argument overlap mean of adjacent sentences	0.537	0.186	.555	.161	.010
Stem overlap mean of adjacent sentences	0.501	0.195	.332	.157	.127
Noun overlap mean of all sentences	0.281	0.147	.137	.097	.211
Argument overlap mean of all sentences	0.398	0.167	.404	.151	.005
Stem overlap mean of all sentences	0.381	0.166	.219	.134	.168
Content word overlap mean of adjacent sentences	0.102	0.045	.136	.045	.090
Content word overlap mean of all sentences	0.070	0.032	.093	.034	.089
LSA					
LSA overlap mean of adjacent sentences	0.360	0.100	.180	.058	.205
LSA overlap mean of all sentences in paragraphs	0.300	0.098	.157	.060	.193
LSA overlap mean of adjacent paragraphs	0	0	.286	.114	.500
LSA overlap mean of given/new sentences	0.376	0.050	.299	.033	.072
Connectives					
All connectives incidence	86.130	21.215	99.260	17.663	.045
Casual connectives incidence	26.200	11.606	38.953	13.386	.123
Logical connectives incidence	36.058	15.587	45.541	13.172	.074
Adversative and contrastive connectives incidence	15.875	10.126	18.603	8.977	.050
Temporal connectives incidence	18.087	9.025	20.205	8.553	.035
Expanded temporal connectives incidence	18.193	9.807	15.665	8.623	.047
Additive connectives incidence	44.462	14.981	43.882	12.573	.004

(1) Research Question one: Characteristics of the two grade levels' text cohesion

To discover the characteristics of text cohesion, comparisons between the research's data and the norms of social studies (see Appendix) will be made. A subset of a substantial corpus of

texts compiled by the Touchstone Applied Science Associates (TASA) were analyzed to produce these standards. The TASA corpus was translated into the respective Common Core State Standards grade levels for grades K through 1, 2 through 3, 4 through 5, 6 through 8, 9 through 10, and 11 and higher. Kindergarten is the lowest grade, and "K" stands for it. The first year of senior high school in China is equivalent to grade 10 (McNamara et al., 2014).

On the basis of the previous description of grade, the data of senior one should be compared with 9-10 grades' norms of social studies. One-sample T Test was run to distinguish differences between the norms and grade one's data. According to Table 1, significant differences were showed on indices of argument overlap mean of adjacent sentences ($p=0.010$), argument overlap mean of all sentences ($p=0.005$), all connectives incidence ($p=0.045$), temporal connectives incidence ($p=0.035$), expanded temporal connectives incidence ($p=0.047$) and additive connectives incidence ($p=0.004$). Compared with the norms, this study obtained higher value of argument overlap mean of adjacent sentences, argument overlap mean of all sentences, all connectives incidence and temporal connectives incidence, while gained the lower value of expanded temporal connectives incidence and additive connectives incidence.

Table 3. Analysis of One-sample T Test results for selected indices of grade three

Variable	K11-CCR		Grade three		P
	Mean	Standard deviation	Mean	Standard deviation	
Referential Cohesion					
Noun overlap mean of adjacent sentences	0.399	0.197	.267	.159	.125
Argument overlap mean of adjacent sentences	0.527	0.194	.604	.157	.043
Stem overlap mean of adjacent sentences	0.523	0.212	.359	.176	.117
Noun overlap mean of all sentences	0.289	0.146	.188	.141	.133
Argument overlap mean of all sentences	0.399	0.157	.453	.162	.040
Stem overlap mean of all sentences	0.405	0.168	.274	.175	.121
Content word overlap mean of adjacent sentences	0.092	0.045	.144	.046	.138
Content word overlap mean of all sentences	0.064	0.028	.101	.038	.140
LSA					
LSA overlap mean of adjacent sentences	0.382	0.107	.195	.066	.200
LSA overlap mean of all sentences in paragraphs	0.332	0.109	.170	.068	.199
LSA overlap mean of adjacent paragraphs	0	0	.339	.127	.500
LSA overlap mean of given/new sentences	0.382	0.053	.312	.039	.064
Connectives					
All connectives incidence	90.993	18.121	94.197	17.584	.011
Casual connectives incidence	26.776	10.524	40.174	13.961	.126
Logical connectives incidence	37.279	14.150	43.739	11.878	.051
Adversative and contrastive connectives incidence	17.618	9.610	17.978	8.206	.006
Temporal connectives incidence	18.169	9.035	18.226	9.549	.001
Expanded temporal connectives incidence	17.083	9.492	15.884	8.340	.023
Additive connectives incidence	48.488	14.460	39.628	11.617	.064

The senior three's data should be compared to TASA's 11-CCR grades in accordance with the description of grade. According to table 3, the indices of argument overlap mean of adjacent sentences ($p=0.043$), argument overlap mean of all sentences ($p=0.040$), all connectives incidence ($p=0.011$), adversative and contrastive connectives incidence ($p=0.006$), temporal connectives incidence ($p=0.001$) and expanded temporal connectives incidence ($p=0.023$) demonstrated significant differences. Compared with the norms, this study's value of argument overlap mean of adjacent sentences, argument overlap mean of all sentences, all connectives

incidence, adversative and contrastive connectives incidence and temporal connectives incidence was higher, while the value of Expanded temporal connectives incidence was lower.

Research Question two: Differences of text cohesion between the two grade levels

This research ran Independent Sample T Test and Mann-Whitney U test to determine whether differences existed for each measure. The detailed information was provided in table 4.

Table 4. Analysis of Independent Sample T Test and Mann-Whitney U test results for selected indices of grade one and three

Variable	Levene test (p)	p	Mann-Whitney U test (p)
Referential Cohesion			
Noun overlap mean of adjacent sentences	.373	.015	
Argument overlap mean of adjacent sentences	.460	.008	
Stem overlap mean of adjacent sentences	.510	.162	
Noun overlap mean of all sentences	.003		.000
Argument overlap mean of all sentences	.395	.007	
Stem overlap mean of all sentences	.016		.003
Content word overlap mean of adjacent sentences	.961	.124	
Content word overlap mean of all sentences	.543	.053	
LSA			
LSA overlap mean of adjacent sentences	.562	.037	
LSA overlap mean of all sentences in paragraphs	.858	.076	
LSA overlap mean of adjacent paragraphs	.289	.000	
LSA overlap mean of given/new sentences	.266	.002	
Connectives			
All connectives incidence	.904	.013	
Casual connectives incidence	.799	.440	.071
Logical connectives incidence	.217	.214	
Adversative and contrastive connectives incidence	.126	.529	
Temporal connectives incidence	.077	.060	
Expanded temporal connectives incidence	.925	.823	
Additive connectives incidence	.432	.003	

The Levene Test showed that noun overlap mean of all sentences and stem overlap mean of all sentences violated the assumption of homogeneity of variance. Therefore, Mann-Whitney U Test was run on these two indices, while other variables were analyzed by Independent Sample T Test.

Each dimension had variables exhibiting differences. In the category of referential cohesion, noun overlap mean of adjacent sentences ($P=.015$), argument overlap mean of adjacent sentences ($p=.008$), noun overlap mean of all sentences ($p=.000$), argument overlap mean of all sentences ($p=.007$) and stem overlap mean of all sentences ($p=.003$) demonstrated significant differences between two grades. Under the category of LSA, significant differences were showed in LSA overlap mean of adjacent sentences ($p=.037$), LSA overlap mean of adjacent paragraphs ($p=.000$) and LSA overlap mean of given/new sentences ($p=.002$). About the

category of connectives, significant differences were found in all connectives incidence ($p=.013$) and additive connectives incidence ($p=.003$).

Table 4 also presented information of mean and standard deviation. It could be seen that the value of each variable in senior three is higher than that in senior one.

5. Discussion

This paper explored characteristics and differences of text cohesion of two grade levels. The following part will show the discussion about the previous results.

(1) Characteristics of the two grade levels' text cohesion

The results above showed the characteristics of the two grades' text cohesion compared to native speakers are both similar and different. In terms of referential cohesion, the results indicated that there were more argument overlaps in Chinese students' writings. Argument overlap occurs when there is overlap between a noun in one sentence and the same noun (in singular or plural form) in another sentence; it also occurs when there are matching personal pronouns between two sentences (McNamara et al., 2014). Then, no significant difference was showed in the category of LSA of both grades suggested that the two grades' aspects of LSA were approaching native speakers. In addition, the above results revealed that expanded temporal connectives incidence of the two grades was inferior to native speakers.

With reference to differences, they were displayed in category of connectives. From results above, temporal connectives incidence of senior grade one was higher than native speakers, whereas additive connectives incidence was lower. At the same time, higher value of adversative and contrastive connectives incidence was obtained by senior grade three than native speakers, while lower value of temporal connectives incidence was gained.

In sum, the results indicated that students of China overuse argument overlap and connectives especially adversative and contrastive connectives and temporal connectives while the application of expanded temporal connectives and additive connectives was not enough. The reason may be due to the materials used in this paper. "Working hard to make our dreams come true" was a small issue that provided efficiency for students. For example, the word "hard" was used repeatedly in both individual and all essays.

(1) Differences of text cohesion between the two grade levels

The results above revealed that significant differences existed in noun overlap of adjacent sentences, argument overlap of adjacent sentences, noun overlap of all sentences, argument overlap of all sentences, stem overlap of all sentences, LSA overlap of adjacent sentences, LSA overlap of adjacent paragraphs, LSA overlap of given/new sentences, all connectives incidence and additive connectives incidence. Meanwhile, they indicated that senior three obtained better text cohesion than senior one. Namely, text cohesion of senior high school students improved as grade level increases.

Overall, this finding lent support to the research of Crossley et al. (2011), which emphasized the importance of cohesion features in identifying grade levels. They researched and analyzed essays computationally by level (9th grade, 11th grade and college freshman). Their findings, however, indicated that these students produced fewer cohesive characteristics in text as a function of grade level, in contrast to this paper's findings that higher cohesion was associated with rising grade level. For example, in Crossley et al.'s study, ninth-grade students were more likely to produce texts that had a larger incidence of positive logical connectives and more content word overlap, but these writings were judged to be of lesser quality. A possible explanation was that the participants of the two research were with different educational levels. Namely, Crossley et al.'s study both involved college students and senior high school students, but this study only touched on the latter. Raters' criteria and objectivity of their paper was the

other likely reason, because expert raters were likely high knowledge readers (McNamara, 2001). What's more, the results of this study were different from the findings of Crossley and McNamara (2012), which pointed out that high-quality L2 essays and low-quality L2 essays were distinguished from linguistic indices related to linguistic sophistication and not text cohesion. This vary may have resulted from the difference of the corpus between the two research. The essays of Crossley and McNamara were from Advanced Level Examination (HKALE) of graduating senior high school students at Hong Kong while this study's essays were from two senior grade levels at Jiangsu province. The English proficiency of these two regions' students was considerably different from one another since the former had a deeper accumulation because of historical and cultural factors. Moreover, Freedman and Pringle (1980) noted that as grade levels rise, improvements in writing cohesion appeared to slow down. In particular, there were no distinctions between essays produced by seniors in high school and those written by third-year college students in terms of textual unity, organization, development, or coherence. This difference from this study may also result from the different educational levels of the two studies' participants.

To be short, this paper's findings differed from many others', which found out text cohesion were worse with increasing grade levels. In nature, this kind of outcome was probably the product of working memory and prior knowledge (Kellogg, 2008). More specifically, a writer's working memory capacity increases with experience. The capacity along with a writer's background knowledge allows the writer to create writings a greater depth of information. Therefore, one possible explanation for the difference may be the inadequate working memory and background knowledge of this paper's participants.

6. Conclusion

This study sought to examine whether text cohesion improves as grade level increases in senior high school. One finding revealed that argument overlap, content word, LSA overlap of adjacent paragraphs and connectives except expanded temporal connectives were overused. Also, the noun overlap of adjacent sentences, the argument overlap of adjacent sentences, the noun overlap of all sentences, the argument overlap of all sentences, the stem overlap of all sentences, the LSA overlap of adjacent sentences, and the LSA overlap of adjacent paragraphs, LSA overlap of given/new sentences, all connectives and additive connectives all showed significant differences. They suggested that senior three had more cohesive text than senior one. In particular, senior high school students' text cohesion gets better with increasing grade level.

Although this paper research text cohesion with computational tool Coh-Metrix and senior high school students, there are still several limitations. First, the current study roughly considered grade level as proficiency level or quality level. Researchers can organize a pre-test to ensure participants' level in further study. Second, the sample size was small to some extent. Future study can research a larger sample size to make results more reliable and valid. Third, despite the fact that this study tries to show how text cohesion develops as grade levels rise, grade continuity was not covered. It is vital to include senior two in further research.

References

- [1] Andrew Elfenbein. (2011). Comments on Grasser, McNamara, and Kulikowich: Research in Text and the Uses of Coh-Metrix. *Educational Researcher*, 40(5), 246-248.
- [2] Beck, I. L., McKeown, M. G., Omanson, R. C., & Pople, M. T. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. *Reading Research Quarterly*, 19, 263-177.

- [3] Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology*, 103(2), 429-441.
- [4] Chen An-ni & Guo Ai-ping. (2019). Research on the Coefficient of Difficulty for College English Textbooks Based on Coh-Metrix — A Case Study of New Horizon College English (Third Edition). *Theory and Practice of Education*, 39(6): 46-48.
- [5] Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*.
- [6] DU Huiying & CAI Jinting. (2013). A Coh-Metrix-based model of linguistic features predicting argumentative writing quality of EFL learners. *Modern Foreign Languages (Quarterly)*, 36(3): 293-299.
- [7] Freedman, A., & Pringle, I. (1980). Writing in the college years: Some indices of growth. *College Composition and Communication*, 31, 311-324.
- [8] Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, 36, 193-202.
- [9] Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, 40(5), 223-234.
- [10] GUO Chunjie. (2003). Cognitive Foundations of Textual Coherence. *Modern Foreign Languages*, (01): 13-21.
- [11] JIA Wenfeng & ZHANG Peixin. (2020). Exploring the relationship between textual features and writing quality in TEM8 writing test. *Foreign Language Testing and Teaching*, (1): 1-8.
- [12] JIANG Jinlin. (2016). The Application of Coh-Metrix in Foreign Language Teaching and Research. *Foreign Languages in China*, 13(5): 59-60.
- [13] JIANG Jinlin & HAN Baocheng. (2018). A Study of Reading Text Difficulty of CET6, TOEFL AND IELTS Based on Coh-Metrix. *Foreign Languages in China*, 15(3): 86-93.
- [14] Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1, 1-26.
- [15] Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*.
- [16] Latifi, S., & Gierl, M. (2021). Automated scoring of junior and senior high essays using Coh-Metrix features: Implications for large-scale language testing. *Language Testing*, 38(1), 62-85.
- [17] LIANG Maocheng. (2006). A study of coherence in EFL learners' written production. *Modern Foreign Languages*, 29(3): 9..
- [18] McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55, 51-62.
- [19] McNamara D. S., A. C. Graesser, & P. M. McCarthy. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. New York: Cambridge University Press.
- [20] MIAO Xingwei. (1998). The relationship between cohesion and coherence. *Journal of Foreign Languages*, (4): 44.
- [21] Ministry of Education of the People's Republic of China. (2020). *National English Curriculum Standards for General High School (2017 edition, 2020 revision)*. Beijing: Beijing Normal University Publishing Group.

- [22] Scott A. Crossley, Jennifer L. Weston, Susan T. McLain Sullivan & Danielle S. (2011). The Development of Writing Proficiency as a Function of Grade Level :A Linguistic Analysis. *Written Communication*, 28(3): 282-311.
- [23] SHANG Yanzhi. (2015). Cultivating Students' English Written Expression Ability from Text Cohesion. *Teaching & Administration*, (27): 87-90.