INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

www.hrmars.com
ISSN: 2226-6348

L2 Writing Development and Tertiary Grade Level

Wang Hongwei, Li Qian

To Link this Article: http://dx.doi.org/10.6007/IJARPED/v3-i4/1325 DOI: 10.6007/IJARPED/v3-i4/1325

Received: *23 October 2014,* ***Revised:*** *26 November 2014,* ***Accepted:*** *07* December 2014

Published Online: 25 December 2014

In-Text Citation: (Hongwei & Qian, 2014)

To Cite this Article: Hongwei, W., & Qian, L. (2014). L2 Writing Development and Tertiary Grade Level. *International Journal of Academic Research in Progressive Education and Development*, *3*(4), 301–316.

Vol. 3(4) 2014, Pg. 301 - 316

http://hrmars.com/index.php/pages/detail/IJARPED | **JOURNAL HOMEPAGE**


INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT
HRMARS MANAGEMENT ACADEMIC RESEARCH SOCIETY EXPLORING INTELLECTUAL CAPITAL
www.hrmars.com
ISSN: 2226-6348

L2 Writing Development and Tertiary Grade Level

Wang Hongwei
Instructor, Unit 63721
Email: hwdell@sina.com

Li Qian
Instructor, Unit 63721
Email: athenaqianlee@163.com

Abstract
Second language writing assessment is of great significance for second language researchers, teachers and learners. Studies of writing of different grade levels will provide detailed information about textual features in each grade and student's writing development. In this study, a corpus of 124 second language writing texts (from freshmen to senior students) was used to find out the textual feature differences in terms of China's English major students of the four grade levels based on the computational tool Coh-Metrix2.1. Results show that significant differences were found across the four tertiary grade levels in terms of a multiple indices such as basic textual features, cohesion, lexical sophistication, syntactic complexity, word frequency, readability as well as topic sentence-hood. However, not all linear patterns were shown in these indices across the four grade levels, some linear, some non-linear, and the period from the second year to the third year can be treated as a very important turning point that deserves our attention.
Keywords: Writing Proficiency, Tertiary Grade Level, Textual Features, Computational Indices

Writing Assessment

ESL students' writing research is one key theme for SLA and can be assessed by student performance (e.g., McNamara, 1997; Shaw and Weir, 2007) in that such assessments can closely reflect what L2 learners have learned in courses, which English proficiency they will have reached as well as what they will possibly encounter in real-world contexts by making use of writing tasks. For instance, College English test 4 or 6 is required in a majority of universities in China Mainland for obtaining the bachelor's degree. Such test appears to provide evidence that L2 learners are achieving writing skills required in tertiary levels.

Some researches based on writing assessments in large-scale tests to study the relations between textual differences and student writings, for instance, identifying the linguistic features distinguishing writing levels determined by valued scores (e.g., Cumming, et al., 2005; Hinkel, 2003). The most comprehensive and updated studies were from (Ferris, 1994; Grant and Ginther, 2000; Ma Guanghui, 2002; Jarvis et al., 2003; Cumming et al., 2005). Ferris (1994) identified and counted 28 text variables (lexical and syntactic features) in a corpus of 160 ESL texts and compared with ESL writing proficiency with the holistic given scores to the

compositions in the corpus. Grant and Ginther (2000) studied 90 pieces of Test of Written English (TWE) essays, separated into three proficiency levels (i.e., scores 3, 4, or 5), in order to find out differences in L2 students' writing by analyzing general text features, lexical features, grammatical features and clause-level features. Jarvis and his collogues (2003) explored multiple profiles of highly rated timed compositions and described how they compared in terms of altogether 21 lexical, grammatical, and discourse features. Cumming et al (2005) assessed how the discourse written for TOEFL® integrated writing tasks in a corpus of 216 compositions written for six TOEFL tasks by 36 examinees at three different proficiency levels from a variety of discourse features (e.g., lexical complexity, syntactic complexity, rhetoric, and pragmatics). Ma Guanghui (2002) compared 66 linguistic features of Chinese and American students in English writing, and significant differences were found in nine features, i.e., words, second personal pronoun, *because*-clause, verb "*do*", discourse article, sentence connector, persuasive verb, number of adjectives and *that*-clause. Generally speaking, Chinese learners used more second personal pronouns, connectives and adjectives, while American students were better in text length and clause initiators. The multi-dimensional approach to analyze texts has become the trend for discourse analysis.

Due to the previously slow development of computer technology, linguistic features of the studies above are mainly based on hand counting and calculation that would cause computational problems. With the recent rapid information technology development, computational analyses for a large number of texts are taking root and have been attracted by researchers, practitioners, and teachers. The current state-of-the-art computational tool is Coh-Metrix, which can retrieve and extract multiple textual features, e.g., cohesion, lexical sophistication and syntactic complexity, etc, for much deeper level analysis.

Coh-Metrix

Coh-Metrix has been developed and upgraded by Institute for Intelligent Systems, Department of Psychology at the University of Memphis. The rapid technological development and development of different disciplines like corpus linguistics (e.g., Biber, Conrad & Reppen, 1998), discourse processing (e.g., Graesser, Gernsbacher, & Goldman, 2003), computational linguistics (e.g., Moore & Wiemer-Hastings, 2003), etc., have made it possible. It has two versions, one for public use and the other for private use. It has embedded with a large number of multi-level linguistic indices, for instance, indices of cohesion, lexical diversity and syntactic complexity, etc. Each linguistic level has many specific index measurements. The public version Coh-Metrix2.1 is used in this paper, which can retrieve 56 scores of textual features. More information can be seen through the website (http://cohmetrix.Memphisedu/cohmetrixpr/index.html) and the textual indices are shown in Appendix.

As Coh-Metrix has integrated the advanced development in different areas, it has been used for various purposes. For instance, many studies used Coh-Metrix to explore textual differences in L2 discourse studies (e.g., Crossley et al., 2007; Crossley et al., 2007; Crossley & McNamara, 2008; Liang, 2006) and L1 discourse studies (e.g., McCarthy et al., 2006), some analyzed cohesion (e.g., Crossley, Greenfield, & McNamara, 2008), lexis (e.g., Crossley et al., 2009) and text genre (Louwerse et al., 2004). In addition, Coh-Metrix has been validated by many studies (e.g., McNamara et al., 2006).

With regard to the relationships between linguistic indices and writing proficiency by employing Coh-Metrix, researchers and practitioners have carried out meaningful explorations and their studies enabled us empirical evidence and deep thought. For instance,

McNamara and her colleges (2010) adopted Coh-Metrix to detect what linguistic feature differences can predict writing proficiency (writing scores were rated by experts) and found out that syntactic complexity, lexical diversity and word frequency were predictive indices of writing proficiency and none indices of cohesion showed correlation with writing proficiency. Crossley and McNamara (2009) found out the lexical differences in L1 and L2 writings, indicating that L1 writers produced more cohesive writings, and employed more words with meaningful expressions, more infrequent words that make lexical variation and sophistication, than L2 writers. In Crossley, et al (2011), researchers wanted to find out the relationship between linguistic features and human judgment of writing proficiency by first language and second language writers using Coh-Metix. Results indicated that human judgment of writing proficiency was highly correlated with language sophistication such as lexical diversity, word frequency and syntactic complexity, but cohesion was not predictive of writing proficiency.

Meanwhile, some experts tried to explore the relations between linguistic features and grade levels. Crossley and his parterns (2010) attempted to find out linguistic feature differences (lexical sophistication, syntactic complexity and cohesion) across 9th, 11th and college freshmen by adopting Coh-Metrix. Results showed that linguistic features can serve the function of grade level, that is, as grade level increases, students produced more sophisticated words and more complex sentence structures but fewer cohesive devices. From studies taking the computational tool Coh-Metrix above, we can find that textual indices like lexical sophistication, word frequency and syntactic complexity can predict writing proficiency, cohesion have little impact on writing proficiency and the notion that writing proficiency is decided on cohesion can be rejected. This paper imitated Crossley and his colleagues (2010)'s study, in order to find the patterns of linguistic feature of Chinese learners across the four tertiary grade levels.

Corpus of Compositions

This corpus of writings selected was from the sub-corpus –Written English Corpus of Chinese Learners (WECCL) in Spoken and Written English Corpus of Chinese Learners (SWECCL) created and established in 2005 by Nanjing University and Foreign Language Teaching and Research Press, granted by Foreign Language Education and Research Center, Beijing Foreign Language University. WECCL has a million words, with writing texts coming from English majors of four grades in nine schools with different English levels, which ensures the representativeness of the corpus. The corpus is mainly argumentative writings, with small proportion of narrative and expositional writings students' writing texts. Writing variables are controlled, like writing time, writing condition, writing style, writing length, writing level, writing title and student type. Detailed information can be found in (Lifei and Qiufang, 2007). A corpus of 124 writing texts in WECCL were selected for our analysis, i.e., 32 texts of freshmen, 30 texts of sophomores, 32 texts of juniors and 30 texts of seniors. The writing prompt here was:

> *Some people see education simply as going to schools or colleges, or as a means to secure good jobs; most people view education as a lifelong process. In your opinion, how important is education to a modern adult person? Write a composition of about 300 words on the following topic: Education as a Lifelong Process.*

The detailed information was in Table 1. Results suggest that there was no significant difference in average words in the writing texts of the four-grade students with the mean text

words of 313 in freshmen' writing, 320 in sophomores' writings, 299 in juniors' writings and 306 in seniors' writings (Table 1).

Results

Before entering those texts, spellings correction was firstly conducted. In this study, the average spelling mistake per text was two, which would extert a very little impact on the retrival result that can be neglected. Then all the corrected texts were submitted into Coh-Metrix2.1 and all the 56 scores of textual indices came out in EXCEL format and we recorded all the scores of 124 writing texts into another whole EXCEL format for further analysis. Analysis of variance (ANOVA) was used to investigate whether significant difference was found in textual features across the four tertiary grades.

4.1 Basic Textual Features

Of the six indices testing the basic textual features, indices of average paragraphs, average sentences, average words per sentence, average syllables per word showed significant differences in ANOVA (Table 2-4).

Results indicate that first-grade students wrote more paragraphs than second- and fourth-grade students with the *p* value of .005, and no significant differences were found between other grades. Only the third grade showed a little ascending trend, all grades showed descending trend in average paragraphs. In index of average words per sentence, it was found that an ascending trend was shown across the four grades except the third grade that is a bit lower than the second grade. In addition, an ascending trend was shown across the four grade students in index of average syllables per word. In all, given the fixed number of words in text, higher-grade students used fewer paragraphs in writing with longer words, and more words in each sentence.

Cohesion

After ANOVA, significant differences were found in cohesion indices of lexical coreference, connectives, and text cohesion in Coh-Metrix2.1.

Coreferential Cohesion

Three forms of coreference between sentences are currently measured in Coh-Metrix2.1. As stated in Document file as to coreferential cohesion, *noun overlapped*, *noun stem overlapped*, and *argument overlap*, etc. are explained for instance (More detailed information can be found at Coh-Metrix website). Results from Table 5 to Table 8 displayed that an ascending trend was shown in the four indices across the four grades, of which the first- and the fourth-grade showed significant difference in indices of argument overlap (adjacent) and argument overlap (all distances); every two grades showed significant differences in indices of stem overlap (adjacent) and stem overlap (all distances).

Connectives

Researchers have recognized the importance of cohesion and cohesion is analyzed in many fields such as linguistics, discourse processing, psychology, education, etc. Connectives are importantly categorized in the group of cohesion relations in text (Halliday & Hasan, 1976). Of the indices assessing connectives, indices of ratio of pronouns to noun phrases, personal pronoun incidence score and negations showed significant differences through ANOVA. A descending trend was found across four grades in index of ratio of pronouns to noun phrases with the significant differences between the first and the third grade as well as between the

first and the fourth grade. Four grades showed a descending trend in index of personal pronoun incidence score with the significant differences between the first grade and the other three grades. Besides, a descending trend was detected across the four grades in index of number of negations except the third grade, with the significant difference between the first and the fourth grade. Specific results of ANOVA were from Table 9 to Table 11.

Text Cohesion

In Coh-Metrix2.1, text cohesion is assessed by Latent Semantic Analysis (LSA). As found in its website, Latent Semantic Analysis (LSA) is a mathematical, statistical technique for representing world knowledge, based on a large corpus of texts. Singular value decomposition, a general form of principle component analysis, is adopted to condense a very large corpus of texts to 100-500 dimensions (Deerwester, et al., 1990; Landauer et al., 1997; Landauer, et al., 1998). In our study, the ascending trend was demonstrated in the four grades, with significant differences found between the first grade and the third grade as well as between the first grade and the fourth grade (Table 12).

Lexical Sophistication

Through ANOVA analysis, some indices were found significantly different in the four grades, i.e., the word abstractness, type-token ratio for content words (TTR), and content word concreteness. Word abstractness means that the word is characterized with few distinctive features that are difficult for the reader to shape image in the mind, and is measured by hypernym value in WordNet (Fellbaum, 1998; Miller, et al., 1990). Generally speaking, a descending trend was exhibited in index of mean hypernym values of nouns across the four grades, the first grade the highest value and the fourth grade the lowest. However, no significant differences were found between every two grades (Table 13). In index of mean hypernym values of verbs, an increasing trend was found, however, only significant difference was found between the first and the fourth grade (the first grade the lowest value and the fourth grade the highest) with the p <.000 (Table 14). In index of type-token ratio for content words (TTR), results exhibited a mixed curve, i.e., an increasing trend from the first to the second grade, and a decreasing trend from the second to the third grade, and an increasing trend from the third to the fourth grade – the second grade with the highest value, of which the significant difference was found between the first and the second grade with the p value of .044 (Table 15). Index of content word concreteness also showed significant difference in the four grades. Content word concreteness is about how concrete a word is and is extracted by the use of the MRC Psycholinguistics Database (Coltheart, 1981) from which particular characteristics of a word can be scaled. As elaborated in the website, high numbers lean toward concrete and low numbers to abstract with the values varying between 100 and 700 (see Document file for more information). This study found the increasing trend in the four grades except the second grade with the lowest value. Significant difference was found between the second and the fourth grade with the p value of .019, which indicated that higher grade students used more abstract words, whereas lower grade students resorted to more concrete words in their writing (Table 16).

Syntactic Complexity

As stated in Document at the website, syntactic complexity is measured by Coh-Metrix2.1 in several ways such as the mean number of modifiers per noun-phrase, the mean number of higher level constituents per sentence and number of words that appear before the main verb

of the main clause in the sentences of a text. ANOVA analyses of this paper found significant differences in indices of the mean number of modifiers per noun-phrase and the mean number of higher level constituents per word. As for the first, modifiers are to describe the property or nature of the phrase head. The more modifiers are in the phrase, the more complex the phrase is. As for the latter, higher level constituents is defined in Document as that structurally dense sentences tend to have more high order syntactic constitutes per word. The higher the value, the more complex the sentence structure. Thus sentences with more modifiers per noun phrase and higher order syntactic constitutes per word can render the sentence more complex and difficult. Results showed that there was an increasing trend in index of the mean number of modifiers per noun-phrase throughout the four grades and significant differences were found between the first and all the other grades, i.e., higher level students used more modifiers in noun phrase (Table 17). Meanwhile, significant difference was found in index of the mean number of higher level constituents per word across the four grades, a general decreasing trend demonstrated with the third grade higher than the second grade, and significant difference was only found between the first and the fourth grade with the p value of .000 (Table 18).

Word Frequency

Indices of word frequency measure how often particular words occur in the English language and reported indices are from CELEX (Baayen, et al., 1993), a corpus of 17.9 million words. In our study, four indices of word frequency showed significant differences in ANOVA analysis. A decreasing trend was found in index of mean Celex for content words, that is, higher grade students used more content words than lower grade students. Significant differences were also found in the log of frequency as the log of the frequencies is compatible with research on reading time implied in studies of Haberlandt et al. (1985) and Just et al. (1980). The results indicated that all the four indices showed a decreasing trend throughout the four grades, i.e., higher grades students used more content words and more infrequent words which will increase reading time and difficulty (Table 19-22).

Readability

Two indices of text difficulty are measured in Coh-Metrix2.1: the Flesch Reading Ease Score and the Flesch Kincaid Grade Level. Values of the Flesch Reading Ease Score and the Flesch Kincaid Grade Level entitle different meanings, i.e., the higher the Flesch Reading Ease Score, the easier the text is to read; the higher the Flesch Kincaid Grade Level, the more difficult the text is to read (detailed information as to how two indices are measured and their differences can also be found in Document at the website). In our study, a decreasing trend was found in index of Flesch Reading Ease Score rather than the Flesch Kincaid Grade Level throughout the four grades, significant differences between the first and the other grades (Table 23).

Topic sentence-hood

Index of topic sentence-hood was found significantly different in the four grades. The result suggested an ascending trend throughout the four grades and significant difference was only found between the first and the fourth grade with the *p* value of .036, which indicates higher level students pay more attention to topic sentence writing, which will give the reader topic/theme of the paragraph and make the reader better understand what is expressed in the writing (Table 24).

Discussion

What is the L2 writing developmental trend for Chinese learners at the tertiary level? Is it linear or non-linear? What textual differences are demonstrated in indices of Coh-Metrix2.1, cohesion, lexical sophistication and/or syntactic complexity? Such questions are explored in this paper.

Though significant differences were found in a multitude of indices represented in Coh-Metrix2.1, results represented across the indices were not all linear given the text length controlled. For instance, complete ascending trends were found in indices of average syllables per word (basic textual indices), in four indices of coreferential cohesion and in index of text cohesion (cohesion), in index of mean hypernym values of verbs (lexical sophistication), in index of the mean number of modifiers per noun-phrase (syntactic complexity), and in index of topic sentence-hood. Some significant differences were found between the first and the third and between the first and the fourth grade, for instance, which means one year's learning is insufficient to discriminate L2 writing level and two or more time is possible to do so. In addition, complete descending trends were found in index of ratio of pronouns to noun phrases, personal pronoun incidence score (cohesion), all the four indices of word frequency, in index of Flesch Reading Ease Score (readability).

Some graphs were not demonstrated in a complete ascending or descending trend, for instance, an ascending trend was shown in index of average words per sentence across the four grades except the third grade that was a bit lower than the second grade; and a mixed curve was exhibited in index of type-token ratio for content words (TTR), i.e., an increasing trend from the first to the second grade, and a decreasing trend from the second to the third grade, and an increasing trend from the third to the fourth grade. As can be found, the representation of the third grade student should draw our attention in that the textual feature trends are, more or less, determined by the presentation of the third grade students. Since in China, English accreditation – College English Test 4 or 6 (CET 4 or CET 6) is required so as to obtain bachelor's degree so that students will spend one or two years' learning taught and guided by L2 teachers. When passing CET 4 or CET 6, students will lose interest in English and spend less time on English (probably only students majored in English will continue English learning, however, they are not immersed in learning writing skills but advanced learning such as introduction of linguistics, statistics, translation and interpretation, etc). Therefore, in the transition from the second to the third year, textual representation of the third year is not parallel to the general tendency.

In addition, there was a certain conflicting point found in this study as compared to that stated in Document file at the website: Structurally dense sentences tend to have more high order syntactic constitutes per word. However, in this paper, a general decreasing trend was demonstrated in index of the mean number of higher level constituents per word across the four grades but the third grade higher than the second grade. Therefore, whether Coh-Metrix can accurately extract the value of high order syntactic constitutes per word from L2 writings needs to be explored because textual features of L2 writing will possibly bring difficulty for Coh-Metrix to process and then extract. Thus its measurement should be evaluated and testified.

This study demonstrated that textual features like basic textual features, lexical sophistication, syntactic complexity and cohesive devices distinguish tertiary grade levels. This conforms to the notion that L2 writing development still continue to function in college. It is found that a variety of textual features, i.e., cohesion, lexical sophistication and syntactic complexity can be used to distinguish L2 learners of tertiary levels in China. Higher level

students used more words in sentences and longer words than lower level students. In cohesion, higher level students used fewer grammatical cohesive devices but higher lexical cohesive devices, whereas lower level students used more grammatical cohesive devices. In lexical sophistication, higher level students showed higher verb hypernymy, higher TTR, more concrete words and lower noun hypernymy than lower level students. Of the indices of word frequency, higher level students used more infrequent, more difficult words in their writings than lower level students. Additionally, indices of syntactic complexity also indicate the difference in higher level and lower level students. If we assume that higher grade students mean higher English proficiency, thus we can conclude that high proficiency students use sophisticated words and complex syntax which will increase reading difficulty, low proficiency students use familiar words and easier syntactic structures in their writing which speed up the reading and compression.

Conclusion

We found that cohesion, lexical sophistication, and syntactic complexity can be used to distinguish high proficiency writers from low proficiency writers throughout the four tertiary grade levels in this study based on Coh-Metrix results. Thus computational analyses of student writing texts will inform us of the second language develoment pattern based on textual feature development, the characteristics of high proficiency writers and tell us how to teach learners to become mature writers. Besides, more attention should be given to Chinese Learners' lingustic development and further studies are ahead for detailed elaboration.

References

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Brace, N., Kemp, R., & Snelgar, R. (2006). *SPSS for psychologists: A guide to data analysis using SPSS for windows* (3rd edition). London: Palgrave.

Coh-Metrix Document file at website: http://cohmetrix.Memphisedu/cohmetrixpr/index.html

Coltheart, M. (1981). The MRC psycholinguistic database quarterly. *Journal of Experimental Psychology*, *33A*, 497-505.

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42, 475-493.

Crossley, S. A., Louwerse, M. M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal*, 91(2), 15-30.

Crossley, S. A., McCarthy, P. M., & McNamara, D. S. (2007). Discriminating between second language learning text-types. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the 20th international Florida artificial intelligence research society* (pp. 205-210). Menlo Park, California: AAAI Press.

Crossley, S. A., & McNamara, D. S. (2008). Assessing second language reading texts at the intermediate level: An approximate replication of Crossley, Louwerse, McCarthy, and McNamara (2007). *Language Teaching*, 41, 409-429.

Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18, 119-135.

Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of writing quality: Coh-Metrix analyses of first and second language writing. *Int. J. Cont. Engineering Education and Life-Long Learning*, Vol. 21, Nos. 2/3, 170-191.

Crossley, S. A., McNamara, D. S., Weston, J., & McLain Sullivan, S. T. (2010). The development of writing proficiency as a function of grade level: A linguistic analysis. *Technical Report*. The University of Memphis.

Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of writing quality: Coh-Metrix analyses of first and second language writing. *Int. J. Cont. Engineering Education and Life-Long Learning*, Vol. 21, Nos. 2/3, 170-191.

Crossley, S. A., McNamara, D. S., Weston, J., & Sullivan, M. S. T. (2010). The development of writing proficiency as a function of grade level: A linguistic analysis. *Technical Report*. The University of Memphis.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & Mark, J. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10, 5-43.

Deerwester, S. S., Dumais, T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society For Information Science*, 41, 391-407.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Ferris, D. (1994). Lexical and syntactic features of ESL writing by students at different levels. *TOSEL Quarterly*, 28(2), 414-420.

Graesser, A. C., Gernsbacher, M. A., & Goldman, S. R. (2003). *Introduction to the Handbook of discourse processes*. In A.C. Graesser, M. A., Gernsbacher, M. A., & S.R. Goldman (Eds.), Handbook of discourse processes (pp.1-24). Mahwah, NJ: Erlbaum.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36, 193-202.

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

Haberlandt, K., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, 114, 357-374.

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

Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, 37, 275-301.

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions of L2 proficiency. *Journal of Second Language Writing*, 12, 377-403.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284.

Maocheng, L. (2006). Research of Learners' Written Discourse Coherence. *Modern Foreign Language Quarterly*, 29(3), 284-292.

Ma Guanghui. (2002). Comparative analysis of linguistic features in writing of Sino-U.S. colleges students. *Foreign Language Teaching and Research*, 34(5), 345-349.

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. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp.843-848). Mahwah, NJ: Erlbaum.

McCarthy, P. M, Lewis, G. A., Dufty, D. F., & McNamara, D. S. (2006). Analyzing writing styles with Coh-Metrix. *In Proceedings of the Florida Artificial Intelligence Research Society International Conference*. FLAIRS, Melbourne, Florida.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57-86.

McNamara, D. S., Ozuru, Y., Graesser, A. C., & Louwerse, M. (2006). Validating Coh-Metrix. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp.573-578). Mahwah, NJ: Erlbaum.

McNamara, T. F. (1997). Performance testing. In Clapham, C., Corson, D. (Eds.), *Encyclopedia of language and education: Language testing and assessment*. Kluwer Academic Publishers, Norwell, Massachusetts, pp.131-139.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Five papers on WordNet. Cognitive Science Laboratory, Princeton University, No.43.

Moore, J. D., & Wiemer-Hastings, P. (2003). Discourse in computational linguistics and artificial intelligence. In A. C. Graesser, M. A. Gernsbacher, and S.R. Goldman (Eds.), *Handbook of discourse processes* (pp.439-486). Mahwah, NJ: Erlbaum.

Oxford, R. (1990). *Language learning strategies: What every teacher should know*. NY: Newbury House/Harper & Row.

Shaw, S. D., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing*. Cambridge University Press, Cambridge.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Needham Heights, MA: Allyn & Bacon.

Lifei, W., & Qiufang, W. (2007). The Establishment and literature review of Spoken and Written English Corpus of English Learners. *Foreign Language Community*, (1), 22-28.

Table 1. Compositions of the four tertiary grade levels

Grade	Grade 1	Grade 2	Grade 3	Grade 4	In Total
Number of compositions	32	30	32	30	124
Total Words	9,830	9,481	9,269	9,332	37,912

Table 2. Difference in average paragraph

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.724	.166	6.015	0.001

Table 3. Difference in average words per sentence

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.384	0.251	7.171	0.000

Table 4. Difference in average syllables per sentence

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.475	0.225	29.083	0.000

Table 5. Difference in argument overlap, adjacent

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.195	0.315	3.144	0.028

Table 6. Difference in argument overlap, all distances

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
2.485	0.064	3.663	0.014

Table 7. Difference in stem overlap, adjacent

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.977	0.121	13.418	0.000

Table 8. Difference in stem overlap, all distances

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
2.167	0.095	11.874	0.000

Table 9. Difference in ratio of pronouns to noun phrases

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.642	0.183	8.636	0.000

Table 10. Difference in personal pronoun incidence score

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.124	0.342	11.043	0.000

Table 11. Difference in number of negations, incidence score

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.714	0.168	4.243	0.007

Table 12. Difference in mean LSA, Paragraph to Paragraph

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.805	0.15	6.531	0.000

Table 13. Difference in mean hypernym values of nouns

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
0.067	0.977	2.812	0.042

Table 14. Difference in mean hypernym values of verbs

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
0.354	0.786	6.68	0.000

Table 15. Difference in type-token ratio for all content words

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.014	0.389	2.858	0.040

Table 16. Difference in content word concreteness

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
0.289	0.833	3.824	0.012

Table 17. Difference in mean number of modifiers per noun-phrase

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
1.147	0.333	9.376	0.000

Table 18. Difference in mean number of higher level constituents per word

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
0.792	0.501	7.217	0.000

Table 19. Difference in Celex, raw, mean for content words

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
0.406	0.749	14.341	0.000

Table 20. Difference in Celex, logarithm, mean for content words

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
2.048	0.111	21.263	0.000

Table 21. Difference in Celex, raw, minimum in sentence for content words

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
2.533	0.06	7.147	0.000

Table 22. Difference in Celex, logarithm, minimum in sentence for content words

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
0.59	0.623	27.674	0.000

Table 23. Difference in Flesch Reading Ease Score (0-100)

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
2.667	0.051	35.022	0.000

Table 24. Difference in topic sentence-hood

Test of homogeneity of Variance		ANOVA	
Levene Statistic	Sig.	F	Sig.
0.68	0.566	3.349	0.021

Appendix Coh-Metrix2.1 Indices

No.	Description	Measure	Full description
1	Title	Title	Title
2	Genre	Genre	Genre
3	Source	Source	Source
4	JobCode	JobCode	JobCode
5	LSASpace	LSASpace	LSASpace
6	Date	Date	Date
7	Causal content	CAUSVP	Incidence of causal verbs, links, and particles
8	Causal cohesion	CAUSC	Ratio of causal particles to causal verbs (cp divided by cv+1)
9	Pos. additive connectives	CONADpi	Incidence of positive additive connectives
10	Pos. temporal connectives	CONTPpi	Incidence of positive temporal connectives
11	Pos. causal connectives	CONCSpi	Incidence of positive causal connectives
12	Neg. additive connectives	CONADni	Incidence of negative additive connectives
13	Neg. temporal connectives	CONTPni	Incidence of negative temporal connectives
14	Neg. causal connectives	CONCSni	Incidence of negative causal connectives
15	All connectives	CONi	Incidence of all connectives
16	Adjacent argument overlap	CREFA1u	Argument Overlap, adjacent, unweighted
17	Adjacent stem overlap	CREFS1u	Stem Overlap, adjacent, unweighted
18	Adjacent anaphor reference	CREFP1u	Anaphor reference, adjacent, unweighted
19	Argument overlap	CREFAau	Argument Overlap, all distances, unweighted

20	Stem overlap	CREFSau	Stem Overlap, all distances, unweighted
21	Anaphor reference	CREFPau	Anaphor reference, all distances, unweighted
22	NP incidence	DENSNP	Noun Phrase Incidence Score (per thousand words)
23	Pronoun ratio	DENSPR2	Ratio of pronouns to noun phrases
24	Conditional operators	DENCONDi	Number of conditional expressions, incidence score
25	Negations	DENNEGi	Number of negations, incidence score
26	Logic operators	DENLOGi	Logical operator incidence score (and + if + or + cond + neg)
27	LSA sentence adjacent	LSAassa	LSA, Sentence to Sentence, adjacent, mean
28	LSA sentence all	LSApssa	LSA, sentences, all combinations, mean
29	LSA paragraph	LSAppa	LSA, Paragraph to Paragraph, mean
30	Personal pronouns	DENPRPi	Personal pronoun incidence score
31	Noun hypernym	HYNOUNaw	Mean hypernym values of nouns
32	Verb hypernym	HYVERBaw	Mean hypernym values of verbs
33	No. of paragraphs	READNP	Number of Paragraphs
34	No. of sentences	READNS	Number of Sentences
35	No. of words	READNW	Number of Words
36	Sentences per paragraph	READAPL	Average Sentences per Paragraph
37	Words per sentence	READASL	Average Words per Sentence
38	Syllables per word	READASW	Average Syllables per Word
39	Flesch Reading Ease	READFRE	Flesch Reading Ease Score (0-100)
40	Flesch-Kincaid	READFKGL	Flesch-Kincaid Grade Level (0-12)
41	Modifiers per NP	SYNNP	Mean number of modifiers per noun—phrase
42	Higher level constituents	SYNHw	Mean number of higher level constituents per word
43	Words before main verb	SYNLE	Mean number of words before the main verb of main clause in sentences
44	Type-token ratio	TYPTOKc	Type-token ratio for all content words
45	Raw freq. content words	FRQCRacw	Celex, raw, mean for content words (0-1,000,000)
46	Log freq. content words	FRQCLacw	Celex, logarithm, mean for content words (0-6)

47	Min. raw freq. content words	FRQCRmcs	Celex, raw, minimum in sentence for content words (0-1,000,000)
48	Log min. freq. content words	FRQCLmcs	Celex, logarithm, minimum in sentence for content words (0-6)
49	Concreteness content words	WORDCacw	Concreteness, mean for content words
50	Pos. logical connectives	CONLGpi	Incidence of positive logical connectives
51	Neg.logical connectives	CONLGni	Incidence of negative logical connectives
52	Intentional cohesion	INTEC	Ratio of intentional particles to intentional content
53	Intentional content	INTEi	Incidence of intentional actions, events, and particles.
54	Temporal cohesion	TEMPta	Mean of tense and aspect repetition scores
55	Syntactic structure similarity adjacent	STRUTa	Sentence syntax similarity, adjacent
56	Syntactic structure similarity all 1	STRUTt	Sentence syntax similarity, all, across paragraphs
57	Syntactic structure similarity all 2	STRUTp	Sentence syntax similarity, sentence all, within paragraphs
58	Content word overlap	CREFC1u	Proportion of content words that overlap between adjacent sentences
59	Spatial cohesion	SPATC	Mean of location and motion ratio scores.
60	Min. concreteness content words	WORDCmcs	Concreteness, minimum in sentence for content words
61	GNRPure	GNRPure	Genre purity
62	TOPSENr	TOPSENr	Topic sentence-hood