INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 2, Issue 4, (2013) E-ISSN: 2226-6348

A Computational Analysis of Textual Features and L2 Writing Proficiency

Wang Hongwei
Unit 63721
Email: hwdell@sina.com

Yan Liqin
Associate Professor, Department of English, School of Humanities and Social Sciences,
National University of Defense Technology
Email: lqyan305@126.com

To Link this Article: http://dx.doi.org/10.6007/IJARPED/v2-i4/396 DOI:10.6007/IJARPED/v2-i4/396

Published Online: 29 August 2013

Abstract
It is of great significance to discriminate good writing from poor writing and Coh-Metrix, a computational tool of reliability and validity, was used to find out the textual features distinguishing the proficiency of students' timed compositions in this study. The findings indicate that indices of number of words, Flesch Reading Ease Score (0-100) and word frequency, logarithm, mean for content words (0-6) in Coh-Metrix 2.1 have an entire 25.6% predictive power for teacher evaluation of writing proficiency in this study. Coh-Metrix will provide researchers and teachers more detailed statistical information for students' writing products, which can be used for effectively teaching second language writing; students can find out their writing strengths and weaknesses in use of textual features and thus modify their writing strategies and styles based on those retrived information.
Keywords: Coh-Metrix, Textual Features, Writing Proficiency

Introduction

In the second language acquisition, writing is the most critical and challenging aspect of the four language skills. Writing well can enable writers to express their ideas, elaborate their arguments well, which can make the communication successful with other people. However, the status of writing home and abroad is never of no problem.

Despite the importance of writing, the results were of no satisfaction at all. Report of 2002 National Assessment of Educational Progress (NAEP) indicated writing dissatisfaction of American students: less than 1/3 students were rated at or above proficient levels with Grade 4 (28%), Grade 8 (31%), and Grade 12 (21%) respectively; only 2% were rated at advanced levels for all three samples (National Center for Educational Statistics, 2003). Report of 2011 National Assessment of Educational Progress (NAEP), new computer-based assessment of students' writing skills, also indicates the similar results, that is, 20% students of Grade 8 and 21% students of Grade 12 were rated below Basic level; 54% students of Grade 8 and 52% students of Grade 12 were at Basic level; 24% students of Grades 8 and 12 were at the

Proficient level and only 3% students of Grades 8 and 12 were at Advanced level (National Center for Educational Statistics, 2012). In the College English Test (CET) band 4, the average score of students' writing is less than 7 (15 in total). Only 15% to 25% students of key universities passed in CET bands 4 and 6 writing (Li Hong et al. 2007). Due to the writing problems faced with learners and educational circles, it is necessary for us to have a better understanding of L2 writing. A great deal of resaerchers tried to find the relationship between linguistic features and proficient writing (e.g., Ellis 1994; Oxford 1990; Witte and Faigley, 1981). This paper aims to realize the goal by analyzing writings of second year college students in a university of China and to explore what textual feature differences exist between high- and low-proficiency writing and the predictive powers of textual features for writing proficiency by employing the state-of-the-art computational tool Coh-Metrix (Graesser et al., 2004). The definition of writing proficiency was based on the reliable raters' (here is the teachers) scores using the set rubric of College English Test band 4 (CET-4). The computational tool Coh-Metrix, which is embedded with a series of textual indices for analysis, can provide detailed textual indices representation in writing products.

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 et al., 1998), discourse processing (e.g., Graesser et al., 2003), and computational linguistics (e.g., Moore and 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 texual level has many specific index measurements. The public version Coh-Metrix 2.1 is used in this paper, which can retrieve 56 scores of textual features. More information can be found at 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 and McNamara, 2008; Liang, 2006) and L1 discourse studies (e.g., McCarthy et al., 2006). Some analyzed cohesion (e.g., Crossley et al., 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 textual indices and writing proficiency in use of Coh-Metrix, researchers and practitioners have carried out meaningful explorations and their studies enabled us empirical evidence and deep thought. For instance, in Crossley et al (2010), researchers attempted to find out linguistic feature differences (lexical sophistication, syntactic complexity and cohesion) across 9^{th}, 11^{th} and college freshmen by adopting Coh-Metrix. Results showed that linguistic features can serve the function for grade level, that is, as grade level increases, students produced more sophisticated words and more complex sentence structures but fewer cohesive devices. In Crossley et al (2011), researchers wanted to find out the relationship between linguistic features and human judgment of writing proficiency by analyzing L1 and L2 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. McNamara and her colleges (2010) adopted Coh-Metrix to detect what linguistic feature differences can predict writing proficiency (writing scores 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 in predicting writing proficiency. In addition, 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 enable lexical variation and sophistication, than L2 writers. From studies taking the computational tool Coh-Metrix above, we can find that textual indices like lexical sophistication, word frequency and syntactic complexity can predicate writing proficiency, cohesion does not have impact on writing proficiency and the notion that writing proficiency is decided on cohesion thus can be rejected. Based on the results of other researchers using Coh-Metrix, questions are posed – what textual feature differences will display across the large number of indices in Coh-Metrix in terms of L2 writers (Chinese L2 learners), and which textual features are predicative of writing proficiency. We will seek answers by adopting Coh-Metrix.

Writing Collection

All the students involved in the study were the second year college students from a university in China who took English general courses for one and half a year and fulfilled the English exam at the end of the semester. The writing task was for students to write a piece of writing named “Prediction for the future of the Internet” in half an hour. The writing task followed the standard of College English Test band 4 (CET4) set by the Ministry of English in China. Raters of three experienced English teachers were trained to evaluate students’ essays based on the standardized rubric commonly used in CET band 4 (15 scores in total) in assessing these writings and gave scores for every piece of writing and it was found that their inter-rater reliability was high with the $r = .857$, $r = .839$, $r = .807$, $p < .001$ for 10 writing samples tried. Then we collected 844 essays in text forms by correcting word spellings and recorded each score of each piece of writing in EXCEL format for later analysis.

The mean score of the writings was 10.440 ($SD = 1.42$) with the minimum score of 5.0 and the maximum of 14.0 and all the scores were distributed in normal condition. The 844 graded essays were separated into two groups based on a Z score split (above +0.5 and below −0.5) of writing scores resulting in high-proficiency group of 219 writings that received scores of no less than 11.5 and low-proficiency group of 222 writings that received scores of no more than 9.5 (Field 2005). In the high-proficiency group, the mean score was 12.2 ($SD = 0.47$) and in the low-proficiency group, the mean score was 8.6 ($SD = 0.81$) (Table 1 and Table 2).

Results

Independent Samples T Test

Through Independent Samples T Test in SPSS 16.0, we found some textual differences between high- and low-proficiency groups in some indices (Table 3).

Basic Counts of the Two Groups

Along the basic counts of text characteristics, all showed significant differences between the two groups, i.e., high-proficiency writings with more words, longer sentences and more paragraphs but low-proficiency writings with fewer words, shorter sentences and fewer paragraphs (Table 4).

Cohesion of the Two Groups

Of the 13 coreference indices and the 13 indices assessing the incidence of connectives, only indices of CVLP, NAC, and IIAEP could function as indicators of high- and low-proficiency writing proficiency. Of all the three indices assessing cohesion, high-proficiency writings had fewer cohesive devices than low-proficiency writings (Table 5).

In the forms of lexical coreference between sentences, argument overlap and stem overlap were significantly different between the two groups, with high-proficiency group higher lexical coreference between sentences (Table 6).

In addition, noun and pronoun cohesion were significant different along the two groups with the result that there was lower noun phrase incidence score, lower personal pronoun incidence score and lower ratio of pronouns to noun phrases in high-proficiency writings than those in low-proficiency writings (Table 7).

Syntactic Complexity of the Two Groups

Among the indices assessing within-sentence and sentence-sentence syntactic complexity, six indices showed significant differences between the two groups.

In within-sentence level, there were more modifiers per noun phrase, more words before main verb in main clause but fewer higher level constituents per word in high-proficiency group, whereas there were fewer more modifiers per noun phrase, fewer words before main verb in main clause and more higher level constituents per word in low-proficiency group (Table 8).

In sentence-sentence level, low-proficiency writings showed higher sentence syntax similarity score than high-proficiency writings whether sentences are adjacent, sentences are across paragraphs or sentences are in paragraphs (Table 9).

Lexical Diversity of the Two Groups

Of the indices measuring lexical diversity, none index was enough to separate the two groups even through Independent Samples T-Test.

Word Frequency of the Two Groups

Of the five indices testing lexical frequency, three indicated significant differences between the two groups, i.e., lower CELEX (raw) for content words, lower CELEX (logarithm) for content words in high-proficiency group than in low-proficiency group (Table 10).

Readability Index of the Two Groups

Besides, the two groups showed significant difference in Flesch Reading Ease Score (0-100) and Flesch-Kincaid Grade Level (0-12). As stated in Coh-Metrix document file (at its website), the higher the Flesch Reading Ease Score is, the easier the text is to read; by contrast, the higher the Flesch-Kincaid Grade Level is, the more difficult the text is to read. Thus the two indices of readability both informed that writings of higher-proficiency group were more difficult to read and comprehend that those of lower-proficiency group.

Correlation

Correlation was conducted to select the predictors of Coh-Metrix indices showing signigficant differences as the dependent variables and scores of high- and low-proficiency group as the fixed factor, and we found some indices functioning writing proficiency as well as correlating with writing scores (Table 12).

From the many correlation coefficients between textual features with writing proficiency, we found that basic textual feature (measured by average words, average words per sentence, average sentence, average syllables per word, average paragraph) had a positive correlation with writing proficiency; cohesion (measured by personal pronoun incidence score, noun phrase incidence score, ratio of pronouns to noun phrases, stem overlap) had a negative correlation; word frequency (measured by Celex for content word) had a negative correlation; syntactic complexity (measured by average words before the main verb) showed a positive correlation while sentence syntax similarity score showed a negative correlation; readability, measured by Flesch Reading Ease Score, had a negative correlation, while Flesch-Kincaid Grade Level showed a positive correlation.

Multiple Regression

A stepwise regression analysis was conducted to examine which of the variables examined in the t test and correlation analyses are predictive for teacher holistic writing ratings.

A linear multiple regression analysis was calculated including the variables. These variables were regressed against the holistic scores for the 441 evaluated writings. Not wanting to run the risk of collinearity between variables, which would waste potential model power, we did not simply select the variables with the remaining highest effect sizes. Instead, we tested the variables for collinearity to ensure that no index pair correlated above r = >.70. If variables that correlated above .70 were used in the model, it would make interpretation difficult because it would be unclear which variables were contributing to the model, as many of the variables might be redundant (Brace et al., 2006; Tabachnick and Fidell, 2001). Thus the seven variables were checked for outliers and multicollinearity.

The outliers' values demonstrated that there were no independent errors caused by residuals (the absolute value of residuals is less than three). Coefficient values demonstrated that the model's data did suffer from multicollinearity, that is, the correlation coefficient between average words per sentence and Flesch Reading Ease Score was −.861, p < .01; average words per sentence and Flesch-Kincaid Grade Level (0-12) was −0.735, p < .01; Flesch Reading Ease Score and Flesch-Kincaid Grade Level was −.845, p < .01. Besides, the VIF values of average words per sentence were higher than 1, which was beyond the threshold for multicollinearity of VIF value under 1 (Field, 2005).

Correlations between the raters' writing evaluations and the three indices are significant (N = 441): average words (r = .444, p < .001), Flesch Reading Ease Score (0-100) (r = –.249, p < .001), and Celex, logarithm, mean for content words (0-6) (r = –.212, p < .01). As shown in t test and correlation analyses, high-proficiency writers used more words and more infrequent words, which would increase reading difficulty and comprehension.

The stepwise regression analysis showed that the combination of three indices had significantly predicted writing ratings**,** F (3, 437) = 50.130, p < .001, r = .506, r^2 = .256, adjusted r^2 = .251. Thus, the three indices combined (number of words, Flesch Reading Ease Score (0-100) and Celex, logarithm for content words) accounted for 25.6% of the variance in the evaluation of the 441 writings. Average number of words was a significant predicator (t = 10.010, p < .001), accounting for 19.8% of the variance. Flesch Reading Ease Score (0-100) was also a significant predictor (t = –4.448, p < .001), accounting for 4.7% of the variance. Celex, logarithm, mean for content words (0-6), another significant predicator (t = –2.569, p < .05), accounted for another 1.1% of the variance (see Table 13 for additional information). Hence our model is:

Holistic writing score =

12.629 + Number of words × .018
+ Flesch Reading Ease Score × (–.027)
+ Celex, logarithm, mean for content words × (– 1.565)

Discussion

The three most predictive textual features from Coh-Metrix2.1 for writing proficiency were number of words, Flesch Reading Ease Score (0-100) and Celex, logarithm mean for content words (0-6). The combination of the three indices significantly predicted writing ratings, F (3, 337) = 50.130, p < .001, r = .506, r^2 = .256, adjusted r^2 = .251. Thus, the three indices combined (number of words, Flesch Reading Ease Score (0-100) and CELEX, logarithm for content words) accounted for 25.6% of the variance in the evaluation of a collection of 441 writings. These features consistently indicated that writings of higher scores were characterized by textual features for comprehension difficulty while writings of lower scores were characterized by textual features easier for reading. High-proficiency writers employed much longer texts in due time which indicated they had a long-term memory and were able to retrieve more information in the given time; whereas low-proficiency students faced the struggle in thinking out what he/she had to write in such a short time. High-proficiency writers used more words per sentence in which more information was given and stretched with more infrequent words with more conceptually and semantically deep information. Meanwhile, due to raters' wide knowledge and cue skills in text comprehension, their reading speed and easiness were not confined to the surface textual features (e.g., cohesion), which could possibly reversely impede their reading if there were too many unnecessary cohesive devices. Thus higher scores were given to those high-proficiency writings with longer texts and more infrequent words indicating a much difficult text for reading.

Such results will give insights as to the status of students' L2 writing learning and how L2 writing should be taught and trained and how students learn to write a more sound piece of writing. There goes a notion that teaching cohesion can make a coherent piece of writing but the results show that cohesion has little impact on teachers' judgment of students' L2 writing in this study. One possible reason to explain the fact of too many cohesive devices represented in student writings is that English teachers pay too much attention to cohesion in writing with the thought that cohesion can enhance the connections for idea and argument progression. However, the finding rejects the previous notion.

Results of previous studies that adopted Coh-Metrix inform that lexical sophistication and syntactic complexity are predicative of writing quality, but our study has not found the similar result. That means indices of lexical diversity or syntactic complexity have not distinguished higher-rated writings from lower-rated writings, the targeted students being of the same level in indices of lexical diversity and syntactic complexity. One question is emerged: why the results of other researchers using Coh-Metrix have not occurred across the writings of the targeted population? Is it due to the significant differences of text lengths of the two groups? If so, we need to explore much further and deeper. In addition, there occurs a certain conflicting result as compared to what is stated in Coh-Metrix document file at the website: Structurally dense sentences tend to have more high order syntactic constitutes per word. However, it is found that high-proficiency group had a lower value than low-proficiency group in index of the mean number of high order syntactic constitutes per word (the third column in Table 8). Therefore, whether or not Coh-Metrix can accurately extract the value of high order syntactic constitutes per word from L2 writings needs to be explored because it is

possibly difficult for Coh-Metrix to process and then extract this textual feature of L2 writing. Thus its measurement should be evaluated and testified.

In the context of L2 writing in China, the theme of L2 writing is usually augmentative in that argumentative writings can train students to think and express their ideas well. Such writings are usually told in three paragraphs – the first part as introduction, introducing the very topic; the second part as main body, elaborating the idea and arguments which main include at least two evidences to support the idea; the last part as the conclusion which is of two or three sentences. Under such writing set in L2 writing classes, students are taught much more familiar with cohesive expressions (*first of all*, *meanwhile*, *on one hand, on the other hand*, *however*, etc), rather than more lexical choice and complex syntactic structure. Students are taught less about writing techniques, e.g., how to elaborate the arguments and evidences, how to use lexical variation and sophistication, how to use more "advanced" words and expressions, which result in the situation that students remember a great number of cohesive expression, lacking in content elaboration.

The results above have some teaching implications that L2 writing teaching in China should have some change and adjustment. Teachers give higher scores to writing texts of higher reading difficulty indicate that texts of higher reading difficulty are highly valued and attention should be transformed from concentrating on cohesion to lexical diversity and syntactic complexity which increase reading difficulty. Teachers should encourage and train students to write longer texts, use more infrequent words and more complex sentence structures. For instance, students are trained as to how shorter words change into longer words, frequent words into less frequent words, shorter sentences into longer sentences, only one argument into alternative arguments, etc. After such training and practice, students will store much more words and expressions into their mind and will retrieve much more and quickly when given a writing task during the limited time, and what's more, their writings will be highly evaluated.

Conclusion

The study aims to tap into the relationship between writing proficiency and textual features in light of Chinese L2 learners through the use of computational automatic tool Coh-Metrix, of which the results show the significant differences in multiple textual features in high and low L2 writing groups. More detailed information in regard to writing and textual features will provide researchers and teachers with statistical feedback for the betterment of L2 research, teaching pedagogy, and will provide learners detailed information about their strengths and weaknesses in L2 writing.

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.

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

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

Field, A. (2005). *Discovering statistics using SPSS*. London: Sage.

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.

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

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., Louwerse, M. M., Cai, Z., & Graesser, A. (2008). Coh-Metrix version 2.1. Retrieved on Sept. 25, 2011 from http//:cohmetrix.memphis.edu.

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.

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.

National Center for Educational Statistics. (2003). *The nation's report card: Writing 2002, NCES 2003-529*. Retrieved from http://nces.ed.gov/nationsreportcard/pubs/main2002/2003529.asp

National Center for Educational Statistics. (2012). *The nation's report card: Writing 2011, NCES 2012-470*. Retrieved from http://nces.ed.gov/nationsreportcard/pubs/main2011/2012470.asp

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

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

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion and writing quality. *College Composition and Communication*, (3), 189-204.

Table 1.

Mean Score of High- And Low-Proficiency Groups

Group	Mean score	SE of Mean Score
High-proficiency	12.2	.47
Low-proficiency	8.6	.81

Table 2.

Corpus of The Composition

Group	Texts in all	Words in all
High-proficiency	219	260,785
Low-proficiency	222	214,465

Table 3.
Indices As A Function of Distinguishing Writing Proficiency

Indices of basic textual features
1 Average words (AW) 2 Average sentences (AS) 3 Average paragraphs (AP) 4 Average words per sentence (AWPS) 5 Average syllables per word (ASPW)
Indices of causal and intentional cohesion
1 Incidence of causal verbs, links, and particles (ICVLP) 2 Incidence of positive causal connectives (IPCC) 3 Incidence of intentional actions, events, and particles (IIAEP)
Indices of lexical coreference
1 Argument Overlap, all distances, unweighted (AO) 2 Stem Overlap, adjacent, unweighted (SOA) 3 Stem Overlap, all distances, unweighted (SOAD)
Indices of pronoun anaphors
1 Noun Phrase Incidence Score (per thousand words) (NPIS) 2 Ratio of pronouns to noun phrases (P/NP) 3 Personal pronoun incidence score (PPIS)
Indices of syntactic complexity
1 Mean number of modifiers per noun-phrase (MNMPNP) 2 Mean number of higher level constituents per word (MNHLCPW) 3 Mean number of words before the main verb of main clause in sentences (MNWBMV) 4 Sentence syntax similarity, adjacent (SSSA) 5 Sentence syntax similarity, all, across paragraphs (SSSAP) 6 Sentence syntax similarity, sentence all, within paragraphs (SSSSAP)
Indices of word frequency
1 Celex, raw, mean for content words (0-1,000,000) (CRMCW) 2 Celex, logarithm, mean for content words (0-6) (CLMCW) 3 Celex, logarithm, minimum in sentence for content words (0-6) (CLMSCW)
Indices of readability
1 Flesch Reading Ease Score (0-100) (FRES) 2 Flesch-Kincaid Grade Level (0-12) (FKGL)

Table 4.
Difference of The Basic Counts Between The Two Groups

Group	Indices of basic counts				
	AW	AS	AP	AWPS	ASPW
High-proficiency	218	12.3	3.67	21.70	5.97
Low-proficiency	179	11.3	3.40	16.59	6.73
p	.000	.010	.002	.000	.000

Table 5.
Difference of Causal And Intentional Cohesion Between The Two Groups

Group	Indices of cohesion		
	CVLP	NPCC	IIAEP
High-proficiency	6.36	2.85	1.53
Low-proficiency	7.09	3.18	1.74
p	.000	.010	.038

Table 6.
Difference of Argument And Stem Overlap Cohesion Between The Two Groups

Group	Indices of argument and stem overlap cohesion		
	AO	SOA	SOAD
High-proficiency	.69	.59	.58
Low-proficiency	.65	.52	.51
p	.009	.001	.001

Table 7.
Difference of Noun And Pronoun Cohesion Between The Two Groups

Group	Indices of noun and pronoun cohesion		
	NPIS	RPNP	PPIS
High-proficiency	280	.31	8.67
Low-proficiency	288	.33	9.60
p	.001	.019	.002

Table 8.
Difference of Phrasal Complexity of The Two Groups

Group	Indices of within-sentence complexity		
	MNMNP	MNHLCPW	MNWBMV
High-proficiency	.70	.76	4.41
Low-proficiency	.64	.77	4.02
p	.000	.001	.010

Table 9.
Difference of Sentential Complexity of The Two Groups

Group	Indices of sentence-sentence complexity		
	SSSA	SSSAP	SSSSAP
High-proficiency	.11	.11	.11
Low-proficiency	.12	.12	.13
p	.001	.006	.000

Table 10.
Difference of Word Frequency of The Two Groups

Group	Indices of word frequency		
	CRMCW	CLMCW	CLMSCW
High-proficiency	2.84E3	2.52	1.46
Low-proficiency	3.08E3	2.58	1.58
p	.031	.000	.000

Table 11.
Difference of Readability Between The Two Groups

Group	FRES (0-100)	FKGL (0-12)
High-proficiency	59.7	8.80
Low-proficiency	67.4	7.89
p	.000	.000

Table 12.
Indices As A Function of Writing Proficiency As Well As Correlating With Writing Scores

Indices	Writing proficiency
AW	.444***
FRES	–.249***
AWPS	.227***
CLMCW	–.212**
CLMSCW	–.196**
FKGL	.182**
MNMPNP	.179**
AS	.160**
ASPW	.149**
AP	.148**
PPIS	–.143**
SSSSAP	–.140**
SOAD	.135**
SOA	.135**
NPIS	–.131**
SSSA	–.116*
P/NP	–.116*

Note: * p < .01, **p < .01, ***p < .001.

Table 13.
Linear Regression To Predict Writing Proficiency

Entry	Variables Added	R	R^2	B	B	SE
Entry 1	Average words	.444	.198	.417	.018	.002
Entry 2	Flesch Reading Ease Score (0-100)	–.249	.047	–.190	–.027	.006
Entry 3	Celex, logarithm, mean for content words (0-6)	–.212	.011	–.110	–1.565	.609

Note: Estimated constant term is 12.629; B = unstandardized beta; B = standardized beta; SE = standard error.

Appendix Coh-Metrix 2.1 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