

Written Communication

<http://wcx.sagepub.com>

Linguistic Features of Writing Quality

Danielle S. McNamara, Scott A. Crossley and Philip M. McCarthy
Written Communication 2010; 27; 57 originally published online Nov 24,
2009;

DOI: 10.1177/0741088309351547

The online version of this article can be found at:
<http://wcx.sagepub.com/cgi/content/abstract/27/1/57>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Written Communication* can be found at:

Email Alerts: <http://wcx.sagepub.com/cgi/alerts>

Subscriptions: <http://wcx.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://wcx.sagepub.com/cgi/content/refs/27/1/57>

Linguistic Features of Writing Quality

Written Communication

27(1) 57–86

© 2010 SAGE Publications

Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>

DOI: 10.1177/0741088309351547

<http://wc.sagepub.com>



**Danielle S. McNamara,¹ Scott A. Crossley,²
and Philip M. McCarthy¹**

Abstract

In this study, a corpus of expert-graded essays, based on a standardized scoring rubric, is computationally evaluated so as to distinguish the differences between those essays that were rated as high and those rated as low. The automated tool, Coh-Metrix, is used to examine the degree to which high- and low-proficiency essays can be predicted by linguistic indices of *cohesion* (i.e., coreference and connectives), *syntactic complexity* (e.g., number of words before the main verb, sentence structure overlap), the *diversity of words* used by the writer, and *characteristics of words* (e.g., frequency, concreteness, imaginability). The three most predictive indices of essay quality in this study were syntactic complexity (as measured by number of words before the main verb), lexical diversity (as measured by the Measure of Textual Lexical Diversity), and word frequency (as measured by Celex, logarithm for all words). Using 26 validated indices of cohesion from Coh-Metrix, none showed differences between high- and low-proficiency essays and no indices of cohesion correlated with essay ratings. These results indicate that the textual features that characterize good student writing are not aligned with those features that facilitate reading comprehension. Rather, essays judged to be of higher quality were more likely to contain linguistic features associated with text difficulty and sophisticated language.

¹University of Memphis, Memphis, TN

²Mississippi State University

Corresponding Author:

Danielle S. McNamara, Psychology/Institute for Intelligent Systems, University of Memphis, Memphis TN 38152.

Email: dsmcnamara1@gmail.com

Keywords

writing proficiency, cohesion, coherence, essay quality, computational linguistics, assessment

Writing well is a significant challenge for students and of critical importance for success in wide variety of situations and professions. For high school seniors, writing skills are among the best predictors of success in course work during their freshmen year of college (Geiser & Studley, 2001). For professionals, writing skills are essential for their day-to-day work and critical for entry and promotion within their disciplines (Light, 2001). Writing provides the ability to articulate ideas, argue opinions, and synthesize multiple perspectives. Thus, effective writing is essential to communicating persuasively with others, including teachers, peers, colleagues, coworkers, and the community at large (Crowhurst, 1990).

Despite such evidence of the importance of writing, the 2002 National Assessment of Educational Progress (NAEP) report painted a dismal picture of the writing preparedness of American students. Less than a third of students in Grade 4 (28%), Grade 8 (31%), and Grade 12 (21%) scored at or above proficient levels, and only 2% wrote at advanced levels for all three samples. Moreover, only 9% of Grade 12 Black students and only 28% of Grade 12 White students were able to write at a proficient level (National Center for Educational Statistics, 2003).

This educational problem points to a need to better understand writing proficiency. One step in that direction is to develop a better understanding of the linguistic features that characterize proficient writing (e.g., Witte & Faigley, 1981). This study pursues that goal by computationally analyzing essays written by freshman college students and examining the degree to which linguistic features predict essay quality. Our definition of essay quality is premised on essays scores provided by expert raters using a standardized scoring rubric. We analyze the scored essays using Coh-Metrix, an automated text analysis tool that provides a large array of linguistic indices (Graesser, McNamara, Louwerse, & Cai, 2004). Specifically, we examine the ability of five classes of Coh-Metrix indices to predict essay quality. These comprise two classes of measures that assess cohesion (i.e., coreference, connectives), and three classes of measures that are indicative of language sophistication, including syntactic complexity, lexical diversity, and word characteristics (e.g., word frequency, concreteness, imaginability). Our objective is to evaluate the relative importance of these measures to essay proficiency.

One of the goals of Coh-Metrix is to improve our ability to measure text difficulty. This goal is achieved with computational indices of text cohesion as well as an assortment of indices focusing on characteristics of words, sentences, and discourse. Coh-Metrix generates these indices by combining lexicons, a syntactic parser, and several other components that are widely used in computational linguistics. For example, the MRC database Medical Research Council (MRC) Psycholinguistic Database provides psycholinguistic information about words (Coltheart, 1981); WordNet provides linguistic and semantic features of words as well as semantic relationships between words (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990); Latent Semantic Analysis (LSA) provides a statistical representation of world knowledge based on corpus analysis to compute the semantic similarities between words, sentences, and paragraphs (Landauer, McNamara, Dennis, & Kintsch, 2007). Coh-Metrix also provides indices related to syntax using a parser based on Charniak (2000). Graesser et al. (2004) provide an extensive overview of many of the language features reported by Coh-Metrix (see also McNamara, Louwerse, McCarthy, & Graesser, in press).

More than 50 published studies have demonstrated that Coh-Metrix indices can be used to detect subtle differences in text and discourse. Some of these studies used Coh-Metrix to distinguish different types of texts. For example, Louwerse, McCarthy, McNamara, and Graesser (2004) identified significant differences between spoken and written samples of English. McCarthy, Lewis, Dufty, and McNamara (2006) reported that Coh-Metrix could successfully detected authorship even though individual authors recorded significant shifts in their writing style. Graesser, Jeon, Yang, and Cai (2007) identified differences between physics content that occurred in textbooks, texts prepared by researchers, and conversational discourse in tutorial dialogue. McCarthy, Briner, Rus, and McNamara (2007) showed that Coh-Metrix could differentiate sections in typical science texts, such as introductions, methods, results, and discussions. Lightman, McCarthy, Dufty, and McNamara (2007) distinguished the beginnings, middles, and ends of chapters in a corpus of history and science text books for high school. Hall, Lewis, McCarthy, Lee, and McNamara (2007) demonstrated that Coh-Metrix could distinguish between American-English law texts and British-English law texts. Crossley, Louwerse, McCarthy, and McNamara's (2007) investigations of second language learner texts revealed a wide variety of structural and lexical differences between texts that were *adopted* (or authentic) versus *adapted* (or simplified) for second language learning purposes. Collectively, these studies demonstrate that Coh-Metrix is an extremely powerful text analysis tool, capable of assessing and differentiating an enormous variety of text types from the genre level to the sentence level.

The power of Coh-Metrix affords our examination of the linguistic features that characterize good and poor writing (see, for example, Halliday & Hassan, 1976). One of the central purposes of Coh-Metrix is to examine the role of cohesion in distinguishing text types and in predicting text difficulty. Indeed, one underlying assumption of Coh-Metrix is that cohesion is an important component of text difficulty. Cohesion arises from a variety of sources, including explicit referential overlap and causal relationships (Givón, 1995; Graesser, McNamara, & Louwerse, 2003). Referential cohesion refers to the degree to which there is overlap or repetition of words or concepts across sentences, paragraphs, or the entire text. Causal cohesion refers to the degree to which causal relationships are expressed explicitly, most often using connectives (e.g., because, so, and therefore) as linguistic cues. These are only two of many sources of cohesion, but they are the most widely investigated in psychological studies of discourse processing.

Greater cohesion in text has been shown to facilitate comprehension for many readers (Gernsbacher, 1990) and is particularly crucial for low-knowledge readers (McNamara, 2001). When there is a lack of referential or causal cohesion, an idea, relationship, or event must often be inferred by the reader. Low-knowledge readers lack the world knowledge needed to make these inferences. They lack sufficient knowledge to interpret explicit text constituents and to make the inferences needed to meaningfully connect these constituents. Whereas low-knowledge readers are *unable* to make such inferences, high-knowledge readers (who have more background knowledge but do not know the information in the text) are more likely to be successful. Thus, higher knowledge readers can benefit from cohesion gaps because lower cohesion forces readers to generate inferences to fill in the conceptual gaps in the texts and high-knowledge readers have sufficient knowledge to generate meaningful inferences (McNamara, 2001; O'Reilly & McNamara, 2007). Successfully generating inferences aids memory and learning because prior knowledge and the information in the text are more likely to be connected in the readers' mental representation of the text, and the reader's mental representation is likely to be more coherent as a result.

Such interactive effects between readers' knowledge and text cohesion have necessitated a distinction between cohesion and coherence. Whereas cohesion refers to the presence or absence of cues in the text, coherence refers to a quality of the mental representation of the text that is created by the reader. In general, cohesion is highly correlated with coherence because cohesion facilitates the process of developing a coherent mental representation. However, if the reader has sufficient background knowledge, the mental representation of a low-cohesion text may be coherent. Indeed, as coherence depends on generating inferences to connect the information in the text with

prior knowledge, low cohesion can even lead to a more coherent mental representation because the cohesion gaps can induce the reader to generate inferences (Graesser et al., 2003; Louwerse, 2001).

In sum, previous studies demonstrate that cohesion is important for ease of reading comprehension, but whether this facilitation benefits the reader depends on the needs of the reader. However, little is known about the relationship between cohesion and writing. Is cohesion (or coherence) an aspect of essays judged to be of higher quality? Many assume that it is (e.g., Collins, 1998; DeVillez, 2003), and a few studies have found some evidence in that regard (e.g., Witte & Faigley, 1981). Assumedly, cohesive devices cue the reader how to form a coherent representation of the text (Graesser et al., 2003; Louwerse, 2001). Thus, it is assumed by many that a cohesive text is a necessary condition for the text to communicate effectively the writer's intended message to the reader. Accordingly, cohesion within and across the text should facilitate the writer's goal of conveying the thesis of the composition.

Although the importance of cohesion in writing is widely assumed, there is scant evidence to support this notion. Empirical evidence either supporting or rejecting this notion appears to be available solely for second language (L2) writers. For example, Liu and Braine (2005) found a moderate relationship between referential cohesion (e.g., repetition of words) and the quality of writing for 50 students enrolled in a basic writing course at a Chinese university. Such a finding had been supported in previous research as well (e.g., Connor, 1984). However, there is also some opposing evidence. For example, Todd, Khongputb, and Darasawanga (2007) examined the relationship between comments on essays provided by tutors and the cohesive elements identified in essays written by eight postgraduate students at a Thai university. Contrary to expectations, there was no relationship found. The lack (or presence) of cohesive cues in the L2 essays did not seem to influence the tutors' comments.

The jury is still out concerning the role of cohesion in L2 writing. However, writing in a foreign or second language is not the focus of this study, and there is good reason to postulate that the factors driving first (L1) and second language writing may be quite different in that L2 writers spend less time planning and ignore deep-level structures such as cohesion when revising (Leki, 1993; Raimes, 2001). Moreover, we are aware of no studies that have empirically shown that the presence or absence of cohesive cues is directly related to judgments of the quality of the writing for native English writers. The importance of cohesion in L1 writing is most often substantiated with reports of a relationship between subjective judgments of coherence and the quality of the writing. For example, a 1975 NAEP report showed a drop in writing quality from 1969 to 1974 that has been judged to be related in large measure to the lack of coherence in writing (Bamberg, 1983). Similarly,

the poor performance in the 1998 and 2002 NAEP evaluations has been assumed to arise partly from coherence failures. In these studies, the NAEP evaluators rated informative writing as “excellent” when it presented information effectively and consistently with well-chosen details. Failures in coherence, along with errors in grammar and mechanics, resulted in “inadequate” and “poor” scores. Notably, these attributions refer to coherence rather than cohesion, largely because they are based on subjective judgments of the quality of the writing, rather than measures of cohesive cues in the essays.

Given the predominate assumption that cohesion is related to essay quality and a lack of sufficient evidence in either direction, we further investigate this issue in this study by examining whether cohesive cues are more predominant in essays judged to be high quality as opposed to those judged to be of lower quality. In addition to indices of cohesion (i.e., coreference and connectives), we examine three other types of linguistic features. First, we examine indices related to syntactic complexity (e.g., number of words before the main verb, sentence structure overlap). Second, we examine the diversity of words used by the writer. Third, we examine indices on characteristics of words (e.g., frequency, concreteness, imagability).

We examine these indices because of their relationship to text difficulty for comprehension, or alternatively, to language sophistication in terms of production. How syntactic complexity relates to reading comprehension has long been of interest to researchers (Just & Carpenter, 1992; Raynor & Pollatsek, 1994). Although many early studies and most readability formulas have measured syntactic complexity through sentence length (Bormuth, 1969; Chall & Dale, 1995), sentence length is not considered to provide a valid measure of syntactic complexity (e.g., Davison & Kantor, 1982). Instead, most psycholinguistic theories of reading comprehension focus on syntactic structure (Just & Carpenter, 1987; Raynor & Pollatsek, 1994), with the notion that syntax helps the reader link underlying relationships between concepts. Thus, readers segment sentences into phrases and constituents (parsing) and determine relationships between them. These relationships serve as a temporary structure upon which to organize ideas into concepts (Just & Carpenter, 1987). If the syntax of a sentence is complex, higher demands are placed on working memory processing, especially for less-skilled readers (Just & Carpenter, 1992). These higher demands are likely because less-skilled readers cannot immediately construct the appropriate syntactic structures (Raynor & Pollatsek, 1994). This failure may be the result of less-skilled readers' inability to successfully parse sentences, which can lead them to process texts word by word (Field, 2004).

Successful text comprehension is a multilevel process that depends to a large extent on word identification (Perfetti, Wlotko, & Hart, 2005). As such, greater text difficulty and increased comprehension challenges are also

associated with lexical diversity and word frequency. Greater lexical diversity of words in a text means that a wider variety of words are used across the text, which is associated with more challenging text. Lower frequency words in a text means that the words are less familiar and less accessible to the reader. Words that are less common and thereby less frequently encountered by readers have longer eye fixation times and are more difficult to decode (Just & Carpenter, 1987). By contrast, frequent words are processed more quickly and better understood than infrequent words (Haberlandt & Graesser, 1985).

By contrast, less complex sentences, less diversity of words across the text, and more familiar words will generally facilitate reading. Therefore, if proficient writers strive to produce text that facilitates comprehension for the reader, essays judged as higher quality are likely to have these characteristics (i.e., contain less complex sentences, lower diversity of words, and more familiar words). This having been said, as portrayed in Figure 1, we must also recognize that more complex syntax, greater lexical diversity, and less frequent words may be reflective of more sophisticated, skilled language production. Whereas complex syntax, lexical diversity, and infrequent words may result in text that is difficult to process, it also may be reflective of more sophisticated, skilled language production. For example, consider the highly regarded verbal skills of American President Barack Obama, and then consider the opening sentence of his November, 2008 victory speech: "If there is anyone out there who still doubts that America is a place where all things are possible, who still wonders if the dream of our fathers is alive in our time, who still questions the power of our democracy, tonight is your answer." This 45-word sentence features 32 different word-types, eight clauses, a 41-word if-clause, and a main verb arriving after the sentence is 95% complete, a combination unlikely to facilitate working memory resources during comprehension. However, once said, the reaction of the nearly quarter-million people present was elated approval.¹

In sum, better orators and writers may know and use both more complex syntax and less frequent words in their speech. Likewise, lexical diversity is indicative of the range of vocabulary available to a speaker or writer (McCarthy & Jarvis, 2007). Greater lexical diversity in speech or writing is commonly thought to reflect greater linguistic skills, speaker competence, or even a speaker's socioeconomic status (e.g., Ransdell & Wengelin, 2003). If proficient writing, and thereby essay quality, is judged largely by the sophistication of the writing rather than on the ease of processing, then essays judged to be of higher quality are likely to be characterized by more complex sentences, less frequent words, and a greater diversity of words.

Along these same lines, many theories of writing focus on the effects of working memory, skill, and knowledge on writing ability (e.g., Kellogg, 2008;

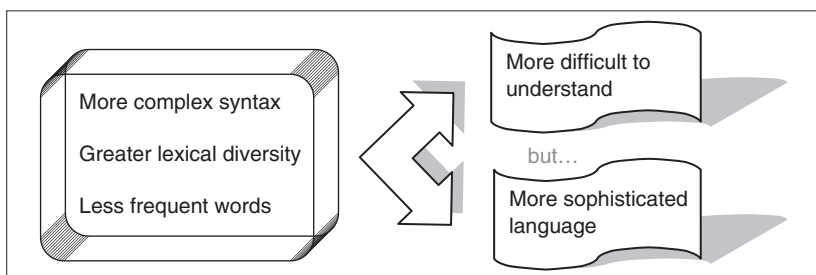


Figure 1. Complex syntax, lexical diversity, and lower frequency words may increase text difficulty, but may also be reflective of more sophisticated language

McCutchen, 2000; Swanson & Berninger, 1996). Accordingly, more-skilled writers have greater working memory capacity to devote to the writing process because they possess more skill and knowledge about language and writing. Therefore, better writers would be more likely to use more sophisticated language in their writing because greater working memory capacity or greater skill and knowledge should facilitate the writer in retrieving less familiar words as well as a more diverse range of words. Similarly, proficient writers would be assumed to have the ability to write more complex sentences because there would be either greater working memory capacity to do so or because proficient writers may have more knowledge of syntactic structures. Thus, proficient writers would be expected to have the capacity to write in more complex or sophisticated language.

At the same time, those who judge writing quality may be looking for more sophisticated language as a signature of writing proficiency. Moreover, whether or not these linguistic features increase the difficulty of the text will also depend on the readers' skill in text comprehension. Skilled readers process complex syntax and less frequent words more quickly than less-skilled readers and assumedly are more familiar with a greater range of syntactic structures and words (Just & Carpenter, 1992; Raynor & Pollatsek, 1994). Therefore, skilled readers (of the essays) may not be affected by these textual attributes. In addition, features of writing associated with language sophistication may overshadow features of writing that facilitate processing, such as higher cohesion and less difficult text. If so, then essays judged as higher quality would be characterized by more complex sentences, a more diverse use of words, and less familiar words. If, on the other hand, the presence of cohesion or the facilitation of the reading process predominates in judgments of writing quality, then the opposite pattern can be expected.

The goal of this study is to examine linguistic differences related to cohesion and linguistic sophistication between high- and low-proficiency writers, as indicated by their score on an essay. To accomplish this goal, we first collected a corpus of argumentative essays and then had expert raters evaluate them using a holistic rubric. We conducted two types of analyses. In the first, our objective was to examine whether Coh-Metrix indices successfully distinguished between high- and low-proficiency essays. To this end, we divided the essays into high- and low-proficiency groups based on human evaluations. We then conducted a discriminant analysis, which allows us to determine if linguistic features of the essays captured by Coh-Metrix are predictive of group membership (in this case, high- and low-proficiency essays). In this first analysis, the goal was essentially to use linguistic features to distinguish between good and poor writers. In the second analysis, our goal was to examine the success of Coh-Metrix indices in predicting the essay grade. For this second analysis, we conducted a regression analysis using essay rating as the dependent variable and Coh-Metrix indices as the predictor variables to determine which linguistic features measured by Coh-Metrix were most predictive of essay ratings and accounted for the largest amount of variance associated with essay quality.

Method

Corpus Collection

We collected a corpus of essays from undergraduate students at Mississippi State University (MSU). The MSU corpus was designed to account for learner variables such as age (adult students) and learning context (freshman college composition class). The corpus was also designed to consider task variables such as medium (writing), first language (English), genre (argumentative essays), essay length (between 500 and 1,000 words), and topics (4 prompts: see Table 1). The final corpus consisted of 120 essays. The essays were untimed and written outside the classroom. Thus, referencing of outside sources was allowed but was not required. Students were allowed to select the essay prompt. Therefore, there is an unequal number of students per prompt.

Essay Evaluation

Five writing tutors with at least 1 year's experience working in a large university writing center rated the 120 essays from the MSU corpus. The raters evaluated the essays based on a standardized rubric commonly used in

Table 1. Essay Prompts Provided to Student

	Number of essays
Some people say that in our modern world, dominated by science, technology, and industrialization, there is no longer a place for dreaming and imagination. What is your opinion?	30
Marx once said that religion was the opium of the masses. If he was alive at the end of the 20th century, he would replace religion with television.	30
In his novel "Animal Farm," George Orwell wrote "all men are equal: but some are more equal than others." How true is this today?	40
Feminists have done more harm to the cause of women than good.	20

assessing SAT essays. The rubric (see Appendix A) was used to holistically assess the quality of the essays and had a minimum score of 1 and a maximum score of 6. Raters were informed that the distance between each score was equal. Accordingly, a score of 5 is as far above a score of 4 as a score of 2 is above a score of 1. The raters were first trained to use the rubric with 20 essays. A Pearson correlation for each essay evaluation was conducted between all possible pairs of raters' responses. If the correlations between all raters did not exceed $r = .80$ (which was significant at $p < .001$), the ratings were reexamined until scores reached the $r = .80$ threshold. After the raters had reached an interrater reliability of at least $r = .80$, each rater then evaluated an additional 20 essays. In this manner, the five tutors evaluated the 120 essays analyzed in this study.

The mean essay score was 3.26 ($SD = 1.23$). The 120 graded essays were separated into two groups based on a median split resulting in a low-proficiency group of essays that received scores between 1 and 3 ($n = 67$) and a high-proficiency group of essays that received scores between 3.1 and 6 ($n = 53$). Among those essays in the low-proficiency group, the majority had scores of either 2 ($n = 32$) or 3 ($n = 25$); among the high-proficiency group, the majority had scores of 4 ($n = 22$) or 5 ($n = 14$).

Results

Descriptive Statistics

Descriptive statistics for the low-proficiency (i.e., scored 1-3) and high-proficiency (i.e., scored 3.1-6) essays are reported in Table 2. The average

Table 2. Descriptive Statistics as a Function of Low- and High-Proficiency Essays and Tests of Differences Between Them

Variable	Low proficiency	High proficiency	<i>F</i> (1, 78)
Raters' essay evaluations	2.33 (.62)	4.41 (.73)	186.84*
Number of words	700.11 (114.38)	748.65 (106.12)	3.77
Number of sentences	38.40 (9.50)	39.48 (7.55)	0.30
Number of paragraphs	5.42 (1.42)	5.63 (1.41)	0.41
Number of words per sentence	18.87 (3.79)	19.40 (3.34)	0.42
Number sentences per paragraph	7.90 (5.75)	8.11 (7.07)	0.02

Note: Standard deviations are in parentheses.

* $p < .001$.

scores of the low- and high-proficiency essays were significantly different, as expected. The low- and high-proficiency essays were not significantly different in terms of number of sentences, number of paragraphs, number of words per sentence, and number of sentence per paragraph. Although the high-proficiency essays were slightly longer in terms of number of words, this difference was not significant. The descriptive statistics indicate that the essays were approximately one page (i.e., typed, double-spaced) consisting of 5 paragraphs, with relatively typical sentence and paragraph lengths.

Prompts

We conducted an analysis of variance (ANOVA) including prompt as a between-subjects variable and essay score as the dependent variable to examine whether the prompts accounted for variance in the essay evaluations. The results from this analysis indicated that the prompts had no significant effect on the raters' evaluations of the essays ($p > .05$). We also examined differences between the four essay prompts according to Coh-Metrix indices. None of the Coh-Metrix indices showed significant differences as a function of prompt.

Discriminant Analysis

The purpose of the discriminant analysis was to examine whether features related to the five classes of variables from Coh-Metrix (i.e., word, sentence, lexical overlap, connectives, lexical diversity) are predictive of low versus high essay quality. A discriminant analysis is a common approach used in many previous studies that have sought to distinguish text types (e.g., Biber,

2003; McCarthy, Lewis, Dufty, & McNamara, 2006). A discriminant analysis is a statistical procedure that predicts group membership (in this case, high- and low-proficiency essays) using a series of predictor variables (in this case, the selected Coh-Metrix variables). A training set is used to generate a model. The model acts as the algorithm that predicts group membership. The model is then applied to a test set to calculate the accuracy of the analysis. Thus, we randomly divided the corpus into two sets: a training set ($n = 80$) and a test set ($n = 40$). The purpose of the training set was to identify which of the Coh-Metrix variables best distinguished the low-proficiency essays from the high-proficiency essays. These selected variables were later used to predict the low-proficiency essays from the high-proficiency essays in the training and test sets using the generated model.

Reducing the likelihood that the model is overfitted is important because if too many variables are used in the discriminant analysis, the model fits not only the signal of the predictors but also the unwanted noise. When overfitting occurs, the training model fits the data well, but when the model is applied to new data the fit lacks accuracy because the noise will not be the same from data set to data set (Tabachnick & Fidell, 1989). With this consideration, a ratio of 20 observations to 1 predictor is standard for analyses of this kind (Field, 2005). Given that the training set contained 80 essays and using the standard estimate of one predictor per 20 variables, we determined that four indices would be an appropriate number of predictors for the discriminant analysis that would not create problems of overfitting.

Coh-Metrix version 2.0 was used to calculate the scores for each essay on 53 indices. The indices considered were the 53 indices from the five classes: coreference ($n = 13$ indices), connectives ($n = 13$ indices), syntactic complexity ($n = 8$ indices), lexical diversity ($n = 5$ indices) word characteristics ($n = 14$ indices). An analysis of variance (ANOVA) was conducted to select the predictors for the discriminant analysis using Coh-Metrix indices as the dependent variables and the high and low essay quality from the training set as the fixed factor. Neither of the 13 coreference indices and the 13 indices assessing the incidence of connectives showed significant differences as a function of essay quality nor were they significantly correlated with the essay scores. These results indicate that cohesion as measured by coreference and the use of connectives did not distinguish between the good and poor essays. Thus, no indices from those two categories would be useful if included in the discriminant analysis.

There were two indices of syntactic complexity that showed significant differences between high- and low-proficiency essays: the mean number of higher level constituents per word and the number of words before the main

verb. The index that displayed the largest effect size for syntactic complexity was the number of words before the main verb. Three of the five indices of lexical diversity showed significant differences as a function of essay quality. The index that displayed the largest effect size for lexical diversity was the measure of textual lexical diversity (MTLD). The index that displayed the largest effect size among those related to word characteristics was Celex logarithm frequency including all words. To assure that the latter result did not occur by chance, we confirmed that there was a significant difference of a function of essay quality for 36 out of the 75 potential word frequency indices provided by Coh-Metrix. These measures were not included in the 14-word characteristic indices because they would be redundant with the selected Celex measure.

Tests for homogeneity of variance confirmed that the amount of variance in the high-proficiency essays and in the low-proficiency essays for each of three indices selected based on the ANOVA results was not significantly different ($p > .05$). We also analyzed collinearity between the variables to ensure that none of the three indices correlated at $r \geq .70$ (Brace, Kemp, & Snelgar, 2006). The correlations between the indices are provided in Table 3, which shows that the indices are not strongly correlated. The ANOVA results for the three indices are provided in Table 4, and the indices are described in the following sections.

Syntactic complexity. The index of syntactic complexity that showed the largest difference between high- and low-proficiency essays was the number of words before the main verb. For example, one type of simple sentence structure is noun phrase + verb (e.g., "The dog ate"; "The girl walked"; "She laughs"). These simple sentences contrast with the sentence, "Thus, in syntactically simple English sentences there are few words before the main verb," for which there are seven words before the main verb (i.e., are). More complex syntactic structures are more difficult for readers to process (Just & Carpenter, 1987; Raynor & Pollatsek, 1994), likely because they impose greater demands on cognitive resources (Perfetti et al., 2005). The results here indicate that high-proficiency writers use more complex syntax than low-proficiency writers.

Lexical diversity. The index of lexical diversity that showed the largest difference between the high- and low-proficiency essays was the Measure of Textual and Lexical Diversity (MTLD; McCarthy, 2005). Unlike other lexical diversity indices, MTLD values do not vary as a function of text length. MTLD also allows for comparisons between text segments of considerably different lengths (at least 100 to 2,000 words) and produces reliable results over a wide range of genres while strongly correlating with other lexical

Table 3. Correlations Between Variables in Training Set

	Syntactic complexity	Word frequency
Lexical diversity	−0.06	−0.51**
Syntactic complexity		−0.07

** $p < .001$.

Table 4. Descriptive and ANOVA Statistics for Low- and High-Proficiency Essays

Variables	Low proficiency	High proficiency	$F(1, 78)$	h_p^2
Syntactic complexity	4.06 (0.92)	4.89 (1.24)	11.87*	0.13
Lexical diversity	72.64 (10.89)	78.71 (13.19)	5.07**	0.06
Word frequency	3.17 (0.08)	3.13 (0.09)	4.64**	0.06

* $p < .001$. ** $p < .05$.

diversity indices (McCarthy, 2005). Indices of lexical diversity assess a writer's range of vocabulary and are indicative of greater linguistic skills (e.g., Ransdell & Wengelin, 2003). The ANOVA results from the current study demonstrate that more proficient writers use a greater range of lexical diversity in their essays.

Word frequency. The word characteristic index that showed the largest difference between the high- and low-proficiency essays was word frequency as measured by CELEX word frequency (logarithm including all words). The CELEX (Baayen, Piepenbrock, & Gulikers, 1995) database consists of frequencies taken from the early 1991 version of the COBUILD corpus, a 17.9 million-word corpus. Frequent words have been shown to facilitate decoding. Rapid or automatic decoding is important because they are strong predictors of text readability (Just & Carpenter, 1987). The results suggest that high-proficiency writers use words that occur less frequently in language.

Accuracy of model. We first conducted a discriminant analysis using the training set. The discriminant analysis model from the training set was then used to predict group membership of the essays in the test set. If the results of the discriminant analysis are statistically significant, then the findings validate the predictions of the analysis (that linguistic indices can be used to classify low- and high-proficiency essay).

The accuracy of the analysis is first evaluated by plotting the correspondence between the actual text type (either low- or high-proficiency essays) in

Table 5. Predicted Text Type Versus Actual Text Type Results

Actual text type	Predicted text type	
	Low proficiency	High proficiency
Training set		
Low proficiency	29	16
High proficiency	12	23
Test set		
Low proficiency	16	6
High proficiency	6	12

the testing and training sets and the predictions made by the discriminant analysis (see Table 5). The results show that the discriminant analysis, using the three variables, correctly predicted 52 of the 80 essays in the training set ($df = 1, n = 80$) $\chi^2 = 7.16, p < .01$). For the test set, the discriminant analysis correctly predicted 28 of the 40 essays ($df = 1, n = 40$) $\chi^2 = 6.20, p < .05$). The model provides 67% accuracy.

The accuracy of the model is also evaluated in terms of recall and precision. Recall scores are computed by tallying the number of hits (i.e., correctly predicted) over the number of hits plus misses (i.e., correctly predicted plus those not correctly predicted). Precision is the number of correct predictions divided by the number of correct predictions plus incorrect predictions (i.e., correctly predicted plus those falsely predicting). This distinction is important because if an algorithm predicted everything to be a member of a single group, it would score 100% in terms of recall but could only do so by claiming members of the other group. If this prediction occurred, then the algorithm would score low in terms of precision. By reporting both values, we can better understand the accuracy of the model (Tabachnick & Fidell, 1989). The accuracy of the model for predicting high- and low-proficiency texts are provided in Table 6. The overall accuracy of the model for the training set was .65. The overall accuracy for the test set was .70.²

Multiple Regression

A stepwise regression analysis was conducted to examine which of the three variables examined in the discriminant function analysis were predictive of human essay ratings using the continuous score (rather than the dichotomous split used in the discriminant analysis). A hierarchical multiple regression

Table 6. Discriminant Analysis Precision and Recall for the Training and Test Sets

Text set	Precision	Recall	FI
Training set			
Low proficiency	0.64	0.71	0.68
High proficiency	0.66	0.59	0.62
Test set			
Low proficiency	0.73	0.73	0.73
High proficiency	0.67	0.67	0.67

analysis was calculated including the three variables representing word frequency, syntactic complexity, and lexical diversity. These three variables were regressed against the holistic evaluations for the 120 evaluated essays. The variables were checked for outliers and multicollinearity. The outliers' values demonstrated that there were no independent errors caused by residuals. Coefficient values demonstrated that the model's data did not suffer from multicollinearity. All VIF values were under 1, the threshold for multicollinearity (Field, 2005).

Correlations between the raters' essay evaluations and the three indices were significant ($N = 120$): syntactic complexity ($r = .34, p < .001$), lexical diversity ($r = .20, p < .01$), and word frequency ($r = -.35, p < .001$). As shown in the ANOVA results, the correlations reflect the finding that lower evaluated essays contained higher frequency words, whereas higher evaluated essays had greater lexical diversity and more complex syntactic structures.

The stepwise regression analysis showed that the indices significantly predicted essay ratings, $F(1, 118) = 15.85, p < .001, r = .47, r^2 = .22$, adjusted $r^2 = .20$. Thus, the three indices combined (Celex logarithm frequency including all words, MTLTD, and number of words before main verb) account for 22% of the variance in the evaluation of the 120 essays examined.³ Celex word frequency was a significant predictor ($t = -3.43, p < .001$) and accounted for 7.9% of the variance. The number of words before the main verb was also a significant predictor ($t = 3.84, p < .001$) and accounted for 11.8% additional variance. MTLTD was not a significant predictor ($t = .19, p > .05$) but accounted for 2.5% of the variance (see Table 7 for additional information).

Discussion

What linguistic features characterize good writing? Some might assume that more proficient writers are more cohesive and thus produce more

Table 7. Linear Regression Analysis to Predict Essay Ratings

Entry	Variable added	R	R ²	B	B	SE
Entry 1	Syntactic complexity	.34	.11	0.29	0.32	0.07
Entry 2	Lexical diversity	.38	.14	0.01	0.02	0.00
Entry 3	Word frequency	-.47	.22	-4.44	-0.31	1.29

Notes: Estimated constant term is 15.81; B = unstandardized beta; B = standardized beta; SE = standard error.

coherent essays. Here, using 26 validated indices of cohesion from Coh-Metrix, we found no indication that higher scored essays were more cohesive. There were no cohesion indices that showed differences between high- and low-proficiency essays, and there were no indices of cohesion that correlated with the essay ratings. However, higher scored essays were more likely to contain linguistic features associated with text difficulty and sophisticated language.

The three most predictive features from Coh-Metrix of essay quality were syntactic complexity (as measured by number of words before the main verb), lexical diversity (as measured by MTLD), and word frequency (as measured by Celex, logarithm for all words). These features consistently indicated that essays that received high scores, and thus presumably higher scored essays, were characterized by linguistic features known to render comprehension more difficult, particularly for less-skilled readers. However, these features may be characterized as being prototypical of writers with more sophisticated language. This conclusion makes sense with respect to theories of writing that propose that more-skilled writers have greater working memory capacity to access and use less familiar words as well as more complex syntax in their writing.

On the other hand, there is an intuitive notion that better writing is more coherent. However, here we found no differences in terms of cohesion. One might suppose that the cohesion indices may simply be inadequate. However, Coh-Metrix indices have been shown to reliably distinguish between high- and low-cohesion texts for a corpus of published studies in which high-cohesion texts were associated with improved comprehension (McNamara et al., in press). The Coh-Metrix indices for cohesion have been validated across a number of other studies distinguishing text types (Crossley et al., 2007; McCarthy et al., 2006, 2007). Thus, we are confident that the Coh-Metrix cohesion indices reliably assess cohesion.

Although the cohesion indices reliably and validly assess cohesion in text, one possibility is that a distinction between cohesion and coherence is crucial

with respect to judgments of writing. As discussed earlier, cohesion comprises the cues that can be detected within the text, whereas coherence is in the mind of the reader. High-knowledge readers can gain from low-cohesion text that forces them to generate inferences while reading (e.g., McNamara, 2001); moreover, skilled readers who have sufficient knowledge about the domain are not affected by cohesion in the text (O'Reilly & McNamara, 2007). In this study, we can presume that the readers of the texts (i.e., the raters) were fairly skilled readers and possessed sufficient knowledge about the topics of the text. Hence, we can further assume that the raters who read the essays did not actually *need* cohesive cues in the essays in order to understand the text. In that sense, the coherence of the essay may emanate from some other aspects of the text that we are not able to measure.

Along these lines, one possible explanation is that the effect of cohesive cues is different in low- and high-cohesion essays. This point is well exemplified by an example provided by Witte and Faigley (1981, p. 201):

1. The quarterback threw the ball toward the tight end.
2. Balls are used in many sports.
3. Most balls are spheres, but a football is an ellipsoid.
4. The tight end leaped to catch the ball.

Whereas each of these sentences are tied together by the coreferent *ball*, sentences 2 and 3 lead to a less coherent mental representation, particularly for a reader who is knowledgeable about American football. Thus, in addition to cohesive ties such as those measured by Coh-Metrix, writing must also maintain a pragmatic coherence.

Halliday and Hassan (1976) refer to a similar notion, the *texture* of a text. While Halliday and Hassan are well known for providing a structured description of potential cohesive cues in text, they also acknowledged that there are other aspects of text meaning that cannot be captured by its cohesion.

The difference between cohesion and coherence can also be exemplified by examining examples of essays from our study. In Appendix B, we provide an example of a high- and low-proficiency essays, which were both written in response to the prompt: "Marx once said that religion was the opium of the masses. If he was alive at the end of the 20th century, he would replace religion with television."

The two essays differ in word frequency, syntactic complexity, and lexical diversity in the expected directions. These differences are also captured by Flesch-Kincaid or traditional measures. The low-proficiency essay has fewer words per sentence ($M = 18.0$) as compared to the high-proficiency essays

($M = 27.8$), and thus a lower Flesch-Kincaid Grade level estimate (grade 9.9 compared to 12.0). Beyond traditional readability scores, the essay that is scored as high proficiency contains more infrequent or rare words (e.g., apathy, sedate, stupor, and comatose), it includes a greater diversity of words (high proficiency: tokens = 751, types = 357, MTLD = 117.871; low proficiency: tokens = 773, types = 315, MTLD = 90.189), and it contains more complex syntax according to Coh-Metrix. An example of the latter is in the sentence, "With a television, not only will there be pseudo-religious overtones to a hefty portion of the channel listings but also an array of mindless chat shows, sitcoms, and reality TV all of which can be set to Marx's standards about the way religion negatively affects humanity." Notably, this 46-word sentence contains six clauses and there are seven words before the (difficult to discern) main verb.

What is not different between these essays is cohesion. According to the Coh-Metrix output, the essays are relatively equal in terms of referential overlap and the use of connectives. Nonetheless, there emerges from reading the two essays a clear sense that one is more coherent than the other. The higher scored essay simply addresses the question in a more coherent way. The higher scored essay is structured in such a manner that the main ideas and arguments are clearly presented and connected allowing for the development of a coherent point of view. Word choices in the higher scored essay are used appropriately and contextually allowing for the essay to stay on topic. In addition, the higher scored essay has a much clearer thesis and a conclusion that summarizes the article's perspective.

A more structured article, such as the highly scored example, should also allow relatively skilled readers (such as the raters used in this study) to process and perhaps look for more sophisticated language. It seems that more proficient writers were able to produce more sophisticated language. While sophisticated language use may have an effect on essay ratings, we are not arguing that writers should be instructed to use such language. Instead, our results only offer evidence that in order to become better writers, students may need to become familiar with and have a better command of a greater diversity of words, less frequent words, and more complex syntactical structures. Reaching this objective will, unfortunately, take time (reading and writing) and deliberate practice (Ericsson, 2006; Kellogg, 2008).

However, these aspects of language may also be helped along through the use of strategies. Strategy instruction across a variety of domains builds on the notion that less-skilled students should learn strategies that mimic those exhibited by skilled students or strategies that compensate for deficits in skills. Providing instruction and practice to use strategies have been found to be

highly beneficial to both comprehension and learning (e.g., McNamara, 2007; Palincsar & Brown, 1984). Strategy instruction is particularly needed and effective for those students who are struggling most, namely, those with less knowledge and lower reading skills (Magliano et al., 2005; O'Reilly & McNamara, 2006). Using strategies can also help to increase working memory resources (e.g., McNamara & Scott, 2001). As such, helping students to learn writing strategies, and thereby scaffolding the writing process, may help them to also improve the sophistication of their language if they are less focused on processes associated with planning, drafting, and revising the essay.

Conclusion

In sum, the results of this study indicate that more-skilled writers use more sophisticated language. The study also offers evidence that some of the textual features of good student writing may not be the same as those features that are considered to be facilitative for reading. Indeed, the results offer researchers some insight into where these differences between quality in reading texts and quality of written text may lie.

Future work will examine to what degree these linguistic differences influence essay scores, and to what degree strategy training benefits the writers. In addition, future work should look at larger corpora of graded essays that would allow for the examination of additional linguistic variables. While the corpus analyzed in our study was well designed and allowed for the presented statistical model, a larger corpus of graded essays would allow for the examination of additional linguistic variables without overfitting the model.

In reference to the graded essays presented in this article, it is important to recognize that our definition of writing quality rested on human judgments by expert raters. While this is the norm in studies such as ours, the expert raters were likely skilled readers who had read numerous essays on the same topic. They were also trained to reliability using a standardized rubric, meaning they were trained to rate the essays reliably in comparison to other essay raters on an exact scale. The raters were also part of the academic community. Individuals from other communities or cultures may have different perspectives on what constitutes good writing. An additional concern is that the writers in this study were responding to the particular demands of the task at hand—that is, to answer a prompt-based essay given a limited time. Thus, although we are making generalizations concerning the quality of writing, these judgments may differ across writing tasks, communities, and cultures.

Overall such potential limitations do not negate the findings of this study. Such potential limitations do highlight the fact that much work lies ahead.

Nonetheless, this study offers an important step toward identifying and evaluating the characteristics of quality writing.

Appendix A

*SAT Scoring Guide*⁴

SCORE OF 6: Demonstrates clear and consistent mastery, although it may have a few minor errors. Effectively and insightfully develops a point of view on the issue and demonstrates outstanding critical thinking, using clearly appropriate examples, reasons, and other evidence to support its position is well organized and clearly focused, demonstrating clear coherence and smooth progression of ideas exhibits skillful use of language, using a varied, accurate, and apt vocabulary, meaningful variety in sentence structure, free of most errors in grammar, usage, and mechanics.

SCORE OF 5: Demonstrates reasonably consistent mastery, although it will have occasional errors or lapses in quality. A typical essay effectively develops a point of view on the issue and demonstrates strong critical thinking, generally using appropriate examples, reasons, and other evidence to support its position is well organized and focused, demonstrating coherence and progression of ideas exhibits facility in the use of language, using appropriate vocabulary, variety in sentence structure, generally free of most errors in grammar, usage, and mechanics.

SCORE OF 4: Demonstrates adequate mastery, although it will have lapses in quality. A typical essay develops a point of view on the issue and demonstrates competent critical thinking, using adequate examples, reasons, and other evidence to support its position is generally organized and focused, demonstrating some coherence and progression of ideas exhibits adequate but inconsistent facility in the use of language, using generally appropriate vocabulary demonstrates some variety in sentence structure has some errors in grammar, usage, and mechanics.

SCORE OF 3: Demonstrates developing mastery and is marked by ONE OR MORE of the following weaknesses: develops a point of view on the issue, demonstrating some critical thinking, but may do so inconsistently or use inadequate examples, reasons, or other evidence to support its position is limited in its organization or focus, or may demonstrate some lapses in coherence or progression of ideas displays developing facility in the use of

(continued)

Appendix A (continued)

language but sometimes uses weak vocabulary or inappropriate word choice lacks variety or demonstrates problems in sentence structure contains an accumulation of errors in grammar, usage, and mechanics.

SCORE OF 2: Demonstrates little mastery, and is flawed by ONE OR MORE of the following weaknesses: develops a point of view on the issue that is vague or seriously limited, and weak critical thinking, providing inappropriate or insufficient examples, reasons, or other evidence to support its position is poorly organized and/or focused or demonstrates serious problems with coherence or progression of ideas displays very little facility in the use of language, using very limited vocabulary or incorrect word choice demonstrates frequent problems in sentence structure contains errors in grammar, usage, and mechanics so serious that meaning is somewhat obscured.

SCORE OF 1: Demonstrates very little or no mastery and is severely flawed by ONE OR MORE of the following weaknesses: develops no viable point of view on the issue, or provides little or no evidence to support its position is disorganized or unfocused, resulting in a disjointed or incoherent essay displays fundamental errors in vocabulary demonstrates severe flaws in sentence structure contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning.

Essays not written on the essay assignment will receive a score of zero.

Appendix B

Essays Examples in Response to the Prompt

Marx once said that religion was the opium of the masses. If he was alive at the end of the 20th century, he would replace religion with television.

Example 1. High Proficiency. Under Marx's observations, replacing religion with television would be like dying a person's hair from brown to brown-but-not-quite-as-brown-as-mentioned. When people sit down to watch TV or to listen to a sermon, they all have to focus on one object for a certain period of time—usually longer than their brains allow them to focus on any given topic. The parallels between the boredom in the routines that are presented to people by religion and TV are undeniable and sometimes inseparable. Flipping through channels nowadays is like having faith “on demand” from your local cable company. Marx would have seen television as a

(continued)

Appendix B (continued)

conductor for mass apathy and would have strongly disagreed with having it as mainstream as it is today. Marx's view of religion's use as a tool to sedate the masses can be applied to modern television programming's strategy of consuming the viewer's attention and leaving them in a near drunken stupor all for the sake of keeping the programs running thus completing the vicious cycle of comatose fixation.

Television in its essence is supposed to bring information from one place to a millions of other places at the speed of light. This luxury has created a pampered society of couch potatoes and TV zombies that don't have to move to get instant pleasure. All they need is their handy-dandy remote control or a passing sibling, child, friend, or spouse to click to the next channel. This lethargic behavior sometimes interferes with the normal activities that people should be accomplishing on a daily basis; furthermore, it can, at times, hinder it completely when someone's favorite show comes on. Some people even find that they don't have to participate in some normal activities because the television can yield the exact same outcome as actually doing the activity. The convenience of the television is outweighed by fad shows, infomercials, and the infamous wacky game shows that spam almost every channel. Most family shows that are supposed to bring a household closer together only bring them to a common room in a house where they sit in near silence to enjoy each others company. The area around the television has turned into the epicenter of many households completely changing the habits of every member. Marx would have realized this in the early stages of the evolution of the television and certainly would have disposed of this digital disease.

The relationship between television and religion has also become more and more intertwined over the years with every voice-cracking televangelist ever to shout in the name of the lord. This trend has lead to even lazier religion seekers. Some have found that staying in on a Sunday morning and watching a sermon on TV is easier than driving down to the local church for some personal meeting with God. In some cases, the sermons they are watching are from those local channels that are airing it so all the churchgoers that aren't present that day can still have the same experiences. Little do they know that very broadcast is the reasoning behind so many vacant seats! Religiously influenced television shows and TV movies that show on Religious TV stations show the viewers profound religious experiences instead of having them go out on their own and discover the religion for themselves. The use

(continued)

Appendix B (continued)

of television has changed the way most people are introduced to the practices of religion thusly tainting any further study they may have to further their beliefs.

With all of this laziness and moral stagnation, Marx's opinion on religion would be one that should be recognized when conversing about the programming on television if Marx was still alive. With a television, not only will there be pseudoreligious overtones to a hefty portion of the channel listings but also will there be an array of mindless chat shows, sitcoms, and reality TV all of which can be set to Marx's standards about the way religion negatively affects humanity. Television was truly a revolution in communication but has slowly decayed into a rotting mass of media that plagues almost every household. It's invention was certainly a major turning point in society and will have lasting affects as it continues changing into forms that are even more convenient for people to rely on instead of creating a reality in which they actually have lives outside of the TV room. Marx's views may be that on religion, but the theme of a society dulled by the pounding effect of too much television is one that is common in both topics.

Example 2. Low Proficiency. Karl Marx may have had some questionable beliefs, but there is no denying his genius. Marx could demand people's attention on a worldwide scale and make the idea of communism sound like a relative utopia compared to the cesspool the world had become. One of the foremost principles of communism was the banning of any and all remnants of religion. Why was it important to separate people from religion when every other society on the planet encouraged religion as a basis to live your life? Marx believed that "religion is the opium of the people." Now the first reaction to this statement in most cultures would be sheer and utter disgust, but on further investigations into this theory shows some promise. Too many people during that time had started to use religion as a shield from the true nature of things, as if the only reason to live was to satisfy some ominous being who would take offense if his creations did not pledge undying loyalty to him. By this time in history, religion had claimed more lives than any other single force. Yet somehow people threw away their own ideals and inhibitions in the name of the church.

Now I have a hard time believing Marx would feel as strongly about the poison of television. I will not stand here and say television is not one of the largest contributions to worldwide obesity, especially in the United States. However, no one has slain an entire race of people because of the religious heritage. In its most basic and primitive form, television is an incredible communication that has done more for the spread of knowledge and information

(continued)

Appendix B (continued)

that the telephone, the telegraph, and the organized mailing system combined. Television has allowed mankind to broadcast to more than 75% of the world's population at one single time. Take for instance the attack on the world trade center, the news of this attack took less than 5 minutes to become global knowledge. A man of Marx's intelligence cannot deny the unrivaled promise the television provides.

No matter how much promise something has, human usually find a way to foul things up. Television went from one of the most promising inventions of history to the scourge of the modern world. Now I do not believe Marx would ban television, but I believe he would limit its use from entertainment into an information transmission device. Now back to the question if television is the new opium of the people, I disagree. Some television shows have absolutely no relevance to the world, but most television has some relevance to the audience in which the show was targeted at—whether it teaches morals for infants or allows housewives to shop without leaving home. Now this is not necessarily allowed in the eyes of a Marxist, but it is not numbing people to the truth about the fleeting existence of a human life. People do not use television to explain why they exist.

Marx did have a weird way of looking at things. In some ways, religion did the same thing Marx was trying to do in a different way. It discouraged people from free thinking and standing out from the group. Religion had somehow convinced people to accept their station in life. So this brings up the question, did Marx believe his own statement or was it just a way of making his ideas sound more profound. There is some truth in his words, but was he seeking truth or just trying to convince the people that they do not need religion. This also connects back to the idea of television. Would Marx go against the obvious potential to use television just to keep his people from enjoying some aspect of their lives? It is definitely a possibility that he would use his incredible talent of persuasion to discourage the use of television. Or would Marx allow television in homes under extensive scrutiny? It is hard to say what Marx would do in this situation, but I do not believe he would feel as strongly about television as he did about religion. This brings up a fundamental difference between religion and television. Religion complicates a person's existence and television only seeks to relax people. However, they do have somewhat the same effect on people's everyday lives. Ultimately though, television and religion do not have the same effect on the grand scale. Religion gives the mass population a false sense of belonging, and television only makes a person's life seem meaningless in a grand scale. So, I guess my answer to the question is NO.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

This research was supported in part by the Institute for Education Sciences (IES R305A080589).

Authors' Note

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES.

Acknowledgements

The authors would also like to express their sincere appreciation to Nicole Miller at Mississippi State University for her help in data collection. The authors also thank Daniel Austin, David Johnson, James Maroney, Justin McElroy, and Daniel White, also from Mississippi State University, for their assistance in scoring the essays used in this study.

Notes

1. Stylistician David Crsytal also discusses this issue, and many others, regarding Obama's language mix of sophistication and complexity (see for example <http://david-crystal.blogspot.com>).
2. A potential concern was that these results may be influenced by the essays with midrange scores. However, a discriminant analysis that excluded essays with scores between 3 and 3.8 yielded highly similar results, with an overall accuracy for the discriminant analysis of 68.2%.
3. A potential concern was that these results may be influenced by the essays with midrange scores. However, a regression analysis that excluded essays with scores between 3 and 3.8 yielded highly similar results, with 18% of the variance accounted for by the three indices.
4. Excerpted from http://www.collegeboard.com/student/testing/sat/about/sat/essay_scoring.html

References

- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database*. Philadelphia: University of Pennsylvania.
- Bamberg, M. (1983). Hans Horman to mean, to understand: Problems of psychological semantics. *Studies in Language*, 7, 431-438.

- Biber, D. (2003). Variation among university spoken and written registers: A new multi-dimensional analysis. In C. Meyer & P. Leistyna (Eds.), *Corpus analysis: Language structure and language use* (pp. 47-70). Amsterdam, Netherlands: Rodopi.
- Bormuth, J. R. (1969). *Development of readability analyses*. Washington, DC: U.S. Office of Education.
- Brace, N., Kemp, R., & Snelgar, R. (2006). *SPSS for psychologists: A guide to data analysis using SPSS for Windows*. London: Palgrave.
- Chall, J., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge, MA: Brookline Books.
- Charniak, E. (2000). A maximum-entropy-inspired parser. *Proceedings of the First Conference on North American Chapter of the Association for Computational Linguistics* (pp. 132-139). San Francisco: Morgan Kaufmann.
- Collins, J. L. (1998). *Strategies for struggling writers*. New York: Guilford.
- Coltheart, M. (1981). MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33, 497-505.
- Connor, U. (1984). A study of cohesion and coherence in ESL students' writing. *Papers in Linguistic: International Journal of Human Communication*, 17, 301-316.
- Crossley, S. A., Louwerse, M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal*, 91, 15-30.
- Crowhurst, M. (1990). Reading/writing relationships: An intervention study. *Canadian Journal of Education*, 15, 155-172.
- Davison, A., & Kantor, R. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly*, 17, 187-209.
- DeVillez, R. (2003). *Writing: Step by step*. Dubuque, IA: Kendall Hunt.
- Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683-703). New York: Cambridge University Press.
- Field, J. (2004). *Psycholinguistics: The key concepts*. New York: Routledge.
- Field, A. (2005). *Discovering statistics using SPSS*. London: Sage.
- Geiser, S., & Studley, R. (2001). *UC and SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California*. Oakland: University of California.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Earlbaum.
- Givón, T. (1995). *Functionalism and grammar*. Philadelphia: John Benjamins.
- Graesser, A. C., Jeon, M., Yang, Y., & Cai, Z. (2007). Discourse cohesion in text and tutorial dialogue. *Information Design Journal*, 15, 199-213.

- Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82-98). New York: Guilford.
- 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.
- Haberlandt, K., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 503-515.
- Hall, C., Lewis, G. A., McCarthy, P. M., Lee, D. S., & McNamara, D. S. (2007). A Coh-Metrix assessment of American and English/Welsh Legal English. *Coyote Papers: Psycholinguistic and Computational Perspectives. University of Arizona Working Papers in Linguistics*, 15, 40-54.
- Halliday, M. A. K., & Hassan, R. (1976). *Cohesion in English*. London: Longman.
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1, 1-26.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Leki, I. (1993). Reciprocal themes in ESL reading and writing. In J. Carson & I. Leki (Eds.), *Reading in the Composition Classroom: Second Language Perspectives* (pp. 9-32). Boston: Heinle and Heinle.
- Light, R. (2001). *Making the most of college*. Cambridge, MA: Harvard University Press.
- Lightman, E. J., McCarthy, P. M., Dufty, D. F., & McNamara, D. S. (2007). The structural organization of high school educational texts. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the 20th International Florida Artificial Intelligence Research Society (FLAIRS) Conference* (pp. 235-240). Menlo Park, CA: AAAI Press.
- Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, 33, 623-636.
- Louwerse, M. M. (2001). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics*, 12, 291-315.
- 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 Cognitive Science Society* (pp. 843-848). Mahwah, NJ: Erlbaum.

- Magliano, J. P., Todaro, S., Millis, K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S. (2005). Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research*, 32, 185-208.
- McCarthy, P. M. (2005). An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD). *Dissertation Abstracts International*, 66, 12. (UMI No. 3199485)
- McCarthy, P. M., Briner, S. W., Rus, V., & McNamara, D. S. (2007). Textual signatures: Identifying text-types using latent semantic analysis to measure the cohesion of text structures. In A. Kao & S. Poteet (Eds.), *Natural language processing and text mining* (pp. 107-122). London: Springer-Verlag.
- McCarthy, P. M., & Jarvis, S. (2007). A theoretical and empirical evaluation of vocd. *Language Testing*, 24, 459-488.
- McCarthy, P. M., Lewis, G. A., Dufty, D. F., & McNamara, D. S. (2006). Analyzing writing styles with Coh-Metrix. In G. C. J. Sutcliffe & R. G. Goebel (Eds.), *Proceedings of the 19th Annual Florida Artificial Intelligence Research Society International Conference (FLAIRS)*; pp. 764-770). Melbourne Beach, FL: AAAI Press.
- McCutchen, D. (2000). Knowledge, processing, and working memory: Implication for a theory of writing. *Educational Psychologist*, 35, 13-23.
- McNamara, D. S. (2001). Reading both high and low coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55, 51-62.
- McNamara, D. S. (Ed.). (2007). *Reading comprehension strategies: Theory, interventions, and technologies*. Mahwah, NJ: Erlbaum.
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (in press). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*.
- McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, 29, 10-17.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K.J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3, 235-244.
- National Center for Educational Statistics. (2003). *2002 NAEP Writing 2002 State Snapshot Records* (NCES 2003-532). Washington, DC: Author. Retrieved on October 18, 2009, from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2003532>
- O'Reilly, T., & McNamara, D. S. (2006). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, 43, 121-152.
- O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures

- of high school students' science achievement. *American Educational Research Journal*, 44, 161-196.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension: Fostering and monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31, 1281-1292.
- Raimes, A. (2001). What unskilled ESL students do as they write: A classroom study of composing. In T. Silva & P. K. Matsuda (Eds.), *Landmark essays: On ESL writing* (pp. 37-62). Mahwah, NJ: Hermagoras Press.
- Ransdell, S., & Wengelin, A. (2003). Socioeconomic and sociolinguistic predictors of children's L2 and L1 writing quality. *Arob@se*, 1-2, 22-29.
- Raynor, A., & Pollatsek, K. (1994). *The psychology of reading*. Cliffs, NJ: Prentice Hall.
- Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology*, 63, 358-385.
- Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics*. New York: Harper.
- Todd, R. W., Khongputb, S., & Darasawanga, P. (2007). Coherence, cohesion and comments on students' academic essays. *Assessing Writing*, 12, 10-25.
- Witte, S. P., & Faigley, L. (1981). Coherence, cohesion and writing quality. *College Composition and Communication*, 22, 189-204.

Bios

Danielle S. McNamara is a professor at the University of Memphis and director of the Institute for Intelligent Systems. Her work involves the theoretical study of cognitive processes as well as the application of cognitive principles to educational practice. Her current research ranges a variety of topics including text comprehension, writing strategies, building tutoring technologies, and developing natural language algorithms.

Scott A. Crossley, PhD, is an assistant professor and director of the TESOL program at Mississippi State University. His interests include computational linguistics, corpus linguistics, and second language acquisition. He has published articles in second language lexical acquisition, multidimensional analysis, discourse processing, speech act classification, cognitive science, and text linguistics.

Philip M. McCarthy is an assistant professor at the University of Memphis. He is a computational linguist, primarily interested in devising, writing, and testing algorithms for text disambiguation. He has numerous publications in several fields including linguistics, artificial intelligence, and cognitive psychology.