

Towards A Computational Assessment of Freewriting Quality

Jennifer L. Weston, Scott A. Crossley* and Danielle S. McNamara

Department of Psychology
The University of Memphis
Memphis TN
Jen.Weston@memphis.edu
dsmcnamr@memphis.edu

*Department of English
Mississippi State University
Starkville MS
sc544@msstate.edu

Abstract

This study examines the linguistic features of freewrites and how those features relate to human scores of freewrite quality. Freewriting is a common prewriting strategy that has received little attention by researchers, particularly in terms of the linguistic features of good and poor freewrites. To address this issue, we developed a scoring rubric to assess the qualities of freewrites and how they are correlated with linguistic features. The results showed that many linguistic features positively correlated with human scores (e.g., referential cohesion, syntactic complexity, lexical difficulty), but the only significant predictors in a regression analysis were, number of words and noun overlap. Better freewrites are longer ones with lexical overlap between sentences. While these results fail to conclusively exclude other potentially important features of higher quality freewrites, this study is a first step toward computationally defining freewrite quality.

Introduction

Writing is possibly the most important skill a person learns during their education. Many students nonetheless exit high school lacking the necessary writing skills to get into or be successful in college and the work place. In fact over 90% of professionals state that writing is essential to their everyday activities on the job (Light, 2001). With increasing class sizes it has become difficult for students to receive the individual attention and help needed to improve their writing proficiency. Increases in class size also potentially limits the amount of feedback available to students during the writing process. It is for this reason that we have begun development on the intelligent tutoring system, Writing-Pal (W-Pal), which teaches writing strategies to high school students. W-Pal teaches students strategies that encompass the entirety of the writing process from prewriting through drafting and revision. Strategies are useful because they can help to diminish demands on working memory and aid in the activation of prior knowledge in long-term working memory (McNamara & Scott, 2001). The use of writing strategies help to keep the writer focused on the steps they need to take to produce a successful written product. This study specifically focuses on the prewriting strategy called freewriting.

When writers are in the first stage of a writing task, they may engage in a brainstorming strategy called freewriting (Renyolds, 1984). Freewriting is an exercise during which a person writes as much as they can and as fast as they can, usually for a set period of time, with little regard to the rules of structure, grammar, and punctuation (Elbow, 1979). Towards the ultimate goal of establishing the impact of freewriting on final writing products (i.e., essays), the purpose of this research is to examine the linguistic features of a freewrite that are predictive of human evaluations of freewriting quality.

While a good deal has been published about freewriting, most of it is anecdotal. Many of these works are based on conjecture with little to no data to support claims that are made. While there is research that has been conducted on freewriting, this research has been limited almost exclusively to anecdotal or qualitative research. The few experimental studies that have been done concerning freewriting have approached freewriting as either a prewriting strategy or a comprehension strategy. Several researchers have investigated the effect of participating in freewriting on final writing products and on comprehension of material (Hinkle & Hinkle, 1990; Knudson, 1989). These studies both found a positive effect for freewriting as a strategy, but neither study examined the features of the freewrites.

Of the studies on freewriting, only one report by Belanoff (1991) examined differences in freewrites as a function of skill level. Belanoff sorted his students into five categories of skilled and unskilled writers based on their previously submitted assignments. He qualitatively examined the freewrites of students in the highest (n=5) and lowest skill group (n=4). Belanoff noted four main qualities of skilled writers' freewrites. First, he noted that skilled freewrites were more chaotic, meaning that they lacked logical connections, and rarely arrived at closure. Second, Belanoff noted that there were discernable passages where the language was eloquent and well formed. A third difference that he noted between freewriting styles regarded the writers' meta-awareness of the task and reference to the task in their freewriting. The fourth and possibly most important of Belanoff's observations was that freewrites that became good essays differed greatly from the original freewrites especially in

structure. These findings suggest that the normal hallmarks of good writing may not be adequate indicators of a successful freewrite.

Belanoff further noted that unskilled writers' freewrites were structured and included full punctuation and connectives. In addition, these freewrites were classified as being constrained in form, consistently arriving at closure, and showing little awareness of the task at hand. Belanoff also noted that the freewrites of less skilled writers often resembled their finished works of writing. Another important difference to note between the skilled and unskilled writers is that while skilled writers tended to ask questions in their writing, the less skilled writers wrote only about what they knew.

While Belanoff's report is the only one that examined the differences in freewrites across writers of different ability, others have looked at freewriting products. These reports have noted some of the same characteristics of freewrites that were noted by Belanoff (Haswell, 1991; Fontaine, 1991), such as their chaotic nature and the presence of occasional eloquent passages, lending credence to Belanoff's assumptions about freewriting. However, these reports were both qualitative and anecdotal in nature and thus lack the control needed to draw broader inferences about freewriting.

Method

Our goal is to better understand the linguistic features that best represent human assessments of quality freewrites. If we can predict how humans evaluate freewriting, then we can provide more accurate feedback to W-Pal users to help them develop and produce better quality freewrites. Assuming a link between quality freewrites and quality essays, we can thus better prepare writers to produce higher quality essays, which is the goal of W-Pal. In order to examine the linguistic features that best represent quality freewrites, we analyzed a corpus of freewrites using linguistic indices taken from the computational tool Coh-Metrix (Graesser et al., 2004). Trained raters scored the freewrite samples using a holistic lexical proficiency rubric. We then divided the scored writing samples into *training* and *test* sets (Whitten & Frank, 2005). To examine which linguistic variables were most predictive of quality freewrites, we conducted correlations and a linear regression model comparing the human ratings and the Coh-Metrix indices using the training set only. The results of this analysis were later extended using the regression model on the independent test set data.

Corpus Collection

Freewriting samples were collected from 105 students across 5 classes (64 9th graders and 41 11th graders) at Webster Schroeder high school (WSHS) in Webster, New York. WSHS is a suburban public school located in upstate New York and is one of two high schools in the district. The two high schools had a combined enrollment of 2961

students for the 2008-2009 school year, with just under 1600 of these students attending WSHS. In spite of its size WSHS maintains an average class size of 23 students and a graduation rate of 99% (Carmody, 2009).

The students who wrote the freewrites were enrolled in either an 11th grade advanced English class or a 9th grade English class and ranged in age from 14 to 18. All of these classes were taught by one teacher who volunteered her classes to participate in this study. The 11th grade students represent the highest level of student in that grade, while the 9th graders were considered more "typical" students. Both levels of students received the same instructions and materials. This range of students ensures a heterogeneous sample of essays, but we do not examine quality as a function of grade in this study.

The data used in the present study is from a larger data set that originally intended students to write two freewrites with paired essays on the same prompts. In addition, they wrote two other freewrites along with completed a set of questionnaires. However, no class moved through the material as quickly as expected and the second freewrite and essay pair were never completed by the students. Students were instructed to skip the pages corresponding to the second freewrite-essay set and to move directly to the first stand alone freewrite. Each student received an experimental packet containing all of the tasks. The order of tasks were as follows: Reading the freewriting instructions (adapted from Elbow, 1973), a five minute freewrite, a 25 minute essay, a five minute freewrite, and a last five minute freewrite (only for the 11th grade students). Each freewrite was completed on a different prompt. Students were read the freewriting and essay instructions and informed when to move onto the next task in their packets. Students completed differing numbers of freewrites depending on the time it required to distribute experimental materials and explain the tasks. The 11th grade classes were more cooperative and were thus able to finish three freewrites, whereas the 9th grade classes completed only two freewrites.

Students wrote on one of two essays prompts. These prompts were counter-balanced and each prompt had a matched freewriting prompt as a prewriting task. In addition, students wrote on between one and two other prompts for freewriting. The prompts were assigned in groups of two from four possible prompts. These prompts were adapted from past SAT prompts obtained from www.onlinemathlearning.com/sat-test-prep.html. We presented students with slightly revised SAT writing section instructions (The College Board, 2009). Instructions were removed regarding the use of paper because we provided the students with as much paper as they needed. In addition, a reminder that the experimenter could not clarify the prompt was added. The freewrites were transcribed as written (i.e., spelling errors were retained). In total, we collected 247 freewrites, of which 97 were randomly selected to be scored and used in the current analysis. Only 97 were used because of time constraints on the raters. The 97 freewrites represented 75

different students (40 9th graders and 35 11th graders) and were written on one of 3 different prompts (memories, $n=26$; choices, $n=37$; and loyalty, $n=34$). The distribution of freewrites across the experimental session was 24 1st freewrites, 65 2nd freewrites, and 8 3rd freewrites.

Survey Instrument

The survey instrument used in this study was a holistic experimenter-created grading rubric. This instrument was based on standardized grading rubrics, such as those used by the College Board to score SAT essays. The report by Belanoff (1991) combined with other anecdotal reports and our own expectations were used to guide particular features included within the rubric. The rubric focused on the development of ideas, the use of appropriate examples, freewrite organization, coherence, and lexical and syntactic variety. We defined a quality freewrite within the rubric as developing a variety of ideas with appropriate examples using a variety of lexical and syntactic structures. It was specified that the freewrite did not need to be well organized, coherent, or grammatical to be of high quality.

Human Ratings

To assess the 97 writing samples that comprise our written corpus, two native speakers of English who were experienced composition instructors were trained as expert raters. The raters were trained on an initial selection of freewrites taken from a training corpus not included in the written corpus used in the study. The raters assigned each freewrite a score between 1 (minimum) and 6 (maximum). To assess inter-rater reliability, Pearson correlations calculated between the raters' responses on the training set exceeded .70 ($p < .001$). The average correlation between the three raters on the target freewrites was .77 ($p < .001$), showing an acceptable level of agreement.

Variable selection

The linguistic indices used in this study were provided by Coh-Metrix (Graesser et al., 2004). Coh-Metrix is a computational tool that reports on over 600 linguistic indices related to conceptual knowledge, cohesion, lexical difficulty, syntactic complexity, and simple incidence scores. Because of the text size of many of the freewrites, not all indices could be investigated. For instance, many of the freewrites were only one paragraph in length, making paragraph to paragraph comparisons impossible. Also, many freewrites comprised fewer than 100 words, the minimum threshold needed for lexical diversity measures.

We used a training set to identify which of the Coh-Metrix variables best correlated to the human scores assigned to each freewrite. Following Whitten and Frank (2005), we divided the corpus into two sets: a training set ($n = 64$) and a testing set ($n = 33$) based on a 67/33 split. These variables selected from the training set were then used to predict the human scores in the training set, using a linear regression model. The freewrites in the test set were analyzed using the regression model from the training set

to calculate the predictability of the variables in an independent corpus (Witten & Frank, 2005).

We ensured that there were at least 15 times more cases (texts) than variables (the lexical indices) to allow for a more reliable interpretation of the multiple regression. A 15 to 1 ratio allows for the interpretation of each variable's individual contribution to the regression model (Field, 2005). With 64 ratings in the training set, this allowed us to choose four linguistic variables. We used Pearson correlations to select a variable from each bank of measures to be used in the multiple regression. Only those variables that demonstrated significant correlations with the human ratings and did not exhibit multicollinearity ($r > .70$) were then used in the multiple regression analysis.

The measures and their respective indices are discussed below in reference to their importance in lexical proficiency.

Measures

Number of Words. The presence of more words in the freewrite indicates that the writer was more prolific and was able to write more on the topic. Number of words may also be correlated with the number of ideas (Kintsch & Keenan, 1973).

Syntactic Complexity. Syntactic complexity is measured by Coh-Metrix in three principal ways. The first is a measure that calculates the mean number of words before the main verb. The second and third metrics used by *Coh-Metrix* measure the mean number of high level constituents (sentences and embedded sentence constituents) per word and per noun phrase. Sentences with difficult syntactic constructions include the use of embedded constituents and are often structurally dense, syntactically ambiguous, or ungrammatical (Graesser et al., 2004). Consequently, they are more difficult to process and comprehend (Perfetti et al., 2005).

Connectives and Logical Operators. Coh-Metrix measures the density of connectives using two dimensions. The first dimension contrasts positive versus negative connectives, whereas the second dimension is associated with particular classes of cohesion identified by Halliday and Hasan (1976) and Louwse (2001). These connectives are associated with positive additive (*also, moreover*), negative additive (*however, but*), positive temporal (*after, before*), negative temporal (*until*), and causal (*because, so*) measures. The logical operators measured in *Coh-Metrix* include variants of *or, and, not, and if-then* combinations. Connectives and logical operators play an important role in the creation of cohesive links between ideas and clauses (Crismore, Markkanen, & Steffensen, 1993; Longo, 1994).

Causality. Causal cohesion is measured in Coh-Metrix by calculating the ratio of causal verbs to causal particles (Graesser et al., 2004). The incidence of causal verbs and causal particles in a text relates to the conveyance of causal content and causal cohesion. The causal verb count is based on the number of main causal verbs identified

through WordNet (Fellbaum, 1998; Miller et al., 1990). Causal verbs and particles help the reader infer the causal relations in the text (Kintsch & van Dijk, 1978).

Lexical Overlap. Coh-Metrix considers four forms of lexical overlap between sentences: noun overlap, argument overlap, stem overlap, and content word overlap. Noun overlap measures how often a common noun of the same form is shared between two sentences. Argument overlap measures how often two sentences share nouns with common stems (including pronouns), while stem overlap measures how often a noun in one sentence shares a common stem with other word types in another sentence (not including pronouns). Content word overlap refers to how often content words are shared between sentences at binary and proportional intervals (including pronouns). Lexical overlap has been shown to aid in text comprehension and reading speed (Douglas, 1981; Kintsch & van Dijk, 1978; Rashotte & Torgesen, 1985).

Semantic Co-referentiality. Coh-Metrix measures semantic co-referentiality using Latent Semantic Analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007), a mathematical technique for representing deeper world knowledge based on large corpora of texts. LSA uses a general form of factor analysis to condense a very large corpus of texts to approximately 300 dimensions. These dimensions represent how often a word occurs within a document (defined at the sentence level, the paragraph level, or in larger sections of texts) and each word, sentence, or text becomes a weighted vector (Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). Unlike lexical overlap indices of co-referentiality, LSA measures associations between words based on semantic similarity, which can be used to assess the amount of semantic coreferentiality in a text (Crossley, Louwrese, McCarthy, & McNamara, 2007). Coh-Metrix also assesses given/newness through LSA by measuring the proportion of new information each sentence provides. The given information is thought to be recoverable from the preceding discourse (Halliday, 1967) and does not require activation (Chafe, 1975).

Word Characteristics. Coh-Metrix reports on a variety of lexical indices taken from WordNet (Fellbaum, 1998; Miller, G., Beckwith, Fellbaum, Gross & Miller, K., 1990) and MRC Psycholinguistic Database (Wilson, 1988). Coh-Metrix derives hypernymy and polysemy indices from WordNet. Hypernymy indices relate to the specificity of words (*cat vs animal*). A lower hypernymy score equates to less specific word choices. Polysemy indices relate to how many senses a word contains. Some words have more senses (e.g., *class*) while others have fewer (e.g., *apricot*). The more senses a word has, the more ambiguous it is. From the MRC Psycholinguistic Database, Coh-Metrix derives indices of word familiarity, word concreteness, and word imagability. All of these indices relate to the accessibility of core lexical items. Core items are closer to prototypical items so higher scores equate to words that are more concrete or more familiar and imagable. The MRC indices are based on the works of Paivio (1965), Toglia

and Battig (1978) and Gilhooly and Logie (1980), who used human subjects to rate large collections of words for psychological properties.

WordFrequency. Word frequency indices measure how often particular words occur in the English language. The Coh-Metrix frequency indices derive their frequency counts from CELEX (Baayen, Piepenbrock, & Gulikers, 1995), which uses frequency counts based on the majority of the words in the text. CELEX is a database from the Centre for Lexical Information, which consists of frequencies taken from the early 1991 version of the COBUILD corpus, a 17.9 million word corpus.

Results

Pearson Correlations Training Set

Pearson Correlations from the training set demonstrated that indices from seven measures reported significant correlations with the human ratings. The seven variables along with their *r* values and *p* values are presented in Table 1, sorted by the strength of the correlation. Because we were limited to four variables to protect the model from overfitting, we selected the four variables that demonstrated the highest Pearson correlation when compared to the human ratings of the freewriting samples and that were not conceptually related.

Table 1

Selected Variables Based on Person Correlations	
Variable	<i>r</i> value
Number of Words	0.664**
Noun Overlap	0.360*
LSA Given Information	0.349*
Number of Words before Main Verb	0.300*
Hypernymy	0.268*
Word Familiarity	0.277*
Word Meaningfulness	0.264*

p* < .05; *p* < .001

Multiple Regression Training Set

A stepwise linear regression analysis was conducted that regressed the four variables (number of words, noun overlap, LSA given information and mean number of words before the main verb) onto raters' score for the 64 freewrites in the training set. The variables were checked for outliers and multicollinearity. Coefficients were checked for both variance inflation factors (VIF) values and tolerance. All VIF values were at about 1 and all tolerance levels were well beyond the .2 threshold, indicating that the model data did not suffer from multicollinearity (Field, 2005).

The stepwise linear regression using the four variables yielded a significant model, $F(2, 61) = 30.732, p < .000$;

adj. $r^2 = .486$. Significant predictors were number of words ($\beta = .620$, $p < .000$) and noun overlap ($\beta = .250$, $p < .05$). The results from the stepwise linear regression demonstrate that these two variables account for 49% of the variance in the human evaluations of freewriting quality for the 64 essays examined in the training set.

Test Set Model

To further support the results from the multiple regression conducted on the training set, we used the B weights and the constant from the training set multiple regression analysis to estimate how well the model would function on an independent data set (the 33 scored freewritings samples in the independent test set). The prediction equation produced an estimated value for each writing sample in the test set. We then conducted a Pearson Correlation between the estimated score and the actual score. We used this correlation along with the adjusted r^2 from running a linear regression on the test data with the training model to demonstrate the strength of the model on an independent data set. The prediction equation was $Y = .924 + .022(\text{number of words}) + .855(\text{noun overlap})$. Predicted scores for the test set significantly correlated with the actual scores, $r = .715$, $p < .000$. The model for the test set yielded an adj. $r^2 = .581$, $p < .000$. The results from the test set model demonstrate that the combination of these variables accounted for 58% of the variance in the evaluation of the 33 freewriting samples comprising the test set.

Discussion

Based on the prior qualitative research combined with intuitive notions of the function of freewriting, we developed a rubric that emphasized the number of ideas generated, the appropriateness of examples, organization, coherence, and syntactic and lexical variety. The regression analysis of linguistic features on score showed that the only significant predictors of freewriting quality were the mean number of words and noun overlap. When looking at these findings in conjunction with the holistic freewriting rubric, we can begin to draw some interesting inferences on what drives freewriting quality.

The purpose of freewriting is to produce a large number of ideas quickly. This being said, it makes sense that number of words is a significant predictor of human scores of freewrite quality. Those students who wrote more likely produced more ideas and thus likely had higher rated freewrites because the number of ideas was a major criteria on which the freewrites were rated. Furthermore the significance of noun overlap is consistent with prior claims that a better freewrite has a larger number of overlapping ideas. These findings suggest that not only are the writers generating ideas but that these ideas are related to each other. The results further show that some of the features we identified as markers of quality freewrites can be computationally identified using Coh-Metrix.

The strong correlation between the predicted scores and actual scores in the test set demonstrate that the salient linguistic features identified in the training set are not unique to that data set. The regression analysis conducted on the test set using forced entry of the significant variables from the training set explained more variance in that set than in the training set, approximately 10% more.

In this study, we were able to identify two significant linguistic features of quality freewrites, which explained between 49 and 58% of the variance associated with the human scores. Nonetheless, a larger sample of freewrites may provide the ability to distinguish other linguistic features that are predictive of human ratings of freewrite quality. The sample size in this study limited the number of predictors that could be used in the analysis. A larger sample size would provide the opportunity to more thoroughly analyze other features thought to be important to freewrite quality (word characteristics, causality, connectives and logical operators etc) and determine what, if any, predictive effects they have on human ratings of freewrite quality.

While we have established the presence of two linguistic features in freewrites that are predictive of higher graded freewrites, without the corresponding essays, we lack the means to determine whether freewrite quality; as we have defined it leads to better essay results. If higher quality freewrites (as rated by humans) lead to lower quality essays, or vice versa, then our judgment of which characteristics should be present in a high quality freewrite will need to be reevaluated. Our future research agenda entails examining how these freewrites transform into essays.

It is our hope that with further research we will be able to assess freewriting quality in comparison with essay quality and be able to develop a quantitative assessment for freewriting using computational tools. Through the use of computational tools such as Coh-Metrix, we will be able to provide real time feedback to students who are learning to freewrite more effectively. Through the freewriting and other strategy modules in W-Pal, we hope to scaffold students toward building better writing skills.

Acknowledgements

This research was supported by the Institute for Education Sciences (IES R305G020018-02; IES R305A080589). 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. We would also like to thank our expert raters, Kristen Dechert and Joshua Thompson, for their help in evaluating the quality of the freewrites and Karen Savella for allowing us to come into her classroom.

References

- Baayen, R. H., Piepenbrock, R. and Gulikers, L. 1995. The CELEX Lexical Database (Release 2) [CDROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Belanoff, P. 1991. Freewriting: an Aid to Rereading Theorists. In P. Belanoff, P. Elbow and S. I. Fontaine eds. *Nothing begins with N*. 16-32. Carbondale and Edwardsville: Southern Illinois University Press.
- Carmody, A. 2009. *Vote May19, 2009-2010 Budget Proposal*. Retrieved from <http://www.websterschools.org/files/filesystem/200910%20Budget%20Newsletter.pdf>
- Chafe, W. L. 1975. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In: C. N. Li eds. *Subject and Topic* 26-55. New York: Academic.
- The College Board. 2005-2008. *SAT Essay Prompts*. Retrieved from: <http://www.onlinemathlearning.com/sat-test-prep.html>.
- Crismore, A., Markkanen, R., Steffensen, M. S. 1993. Metadiscourse in Persuasive Writing: A Study of Texts Written by American and Finnish University Students. *Written Communication*. 10: 39-71
- Crossley, S. A., Louwse, M. M., McCarthy, P. M. and McNamara, D. S. 2007. A Linguistic Analysis of Simplified and Authentic texts. *Modern Language Journal*. 91(2): 15-30.
- Douglas, D. 1981. An Exploratory Study of Bilingual Reading Proficiency. In S. Hudelson ed., *Learning to Read in Different Languages* 33-102. Washington, DC: Center for Applied Linguistics.
- Elbow, P. 1973. *Writing without Teachers*. New York: Oxford University Press.
- Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Graesser, A.C., McNamara, D.S., Louwse, M.M., and Cai, Z. 2004. Coh-Matrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*. 36: 193-202.
- Halliday, M. A. K. 1967. Notes on Transitivity and Theme in English. *Journal of Linguistics*. 3: 199-244.
- Halliday, M. A. K., and Hasan, R. 1976. *Cohesion in English*. London: Longman.
- Haswell, R. H. 1991. Bound Forms in Freewriting: the Issue of Organization. In P. Belanoff, P. Elbow and S. I. Fontaine eds. *Nothing begins with N*. 32-71. Carbondale and Edwardsville: Southern Illinois University Press.
- Hinkle, S., and Hinkle, A. 1990. An Experimental Comparison of the Effects of Focused Freewriting and other Study Strategies on Lecture Comprehension. *Teaching of Psychology*. 17(1): 31-35.
- Kintsch, W., and van Dijk, T. A. (1978). Toward a model of Text Comprehension and Production. *Psychological Review*. 85: 363-394.
- Knudson, R. E. 1989. Effect of Instructional Strategies on Children's Informational Writing. *The Journal of Educational Research*. 83(2): 91-96.
- Field, A. 2005. *Discovering Statistics Using SPSS*. London: Sage Publications, Ltd.
- Fontaine, S. I. 1991. Recording and Transforming: the Mystery of the Ten-minute Freewrite. In P. Belanoff, P. Elbow and S. I. Fontaine eds. *Nothing begins with N*. 3-16. Carbondale and Edwardsville: Southern Illinois University Press.
- Gilhooly, K. J. and Logie R. H. 1980. Age of Acquisition, Imagery, Concreteness, Familiarity and Ambiguity Measures for 1944 Words. *Behaviour Research Methods and Instrumentation*. 12: 395-427.
- Landauer, T. K. and Dumais, S. T. 1997. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of the Acquisition, Induction, and Representation of Knowledge. *Psychological Review*. 104: 211-240.
- Landauer, T. K., Foltz, P. W., and Laham D. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes*. 25: 259-284.
- Landauer, T., McNamara, D. S., Dennis, S., and Kintsch, W. eds. 2007. *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Erlbaum.
- Light, R. J. 2001. *Making the Most of College: Students Speaking Their Minds*. Cambridge: Harvard University Press.
- Longo, B. 1994. Current Research in Technical Communication: The Role of Metadiscourse in Persuasion. *Technical Communication*. 41 : 348-352.
- Louwse, M. M. 2001. An Analytic and Cognitive Parameterization of Coherence Relations. *Cognitive Linguistics*. 12: 291-315.
- McNamara, D. S., Crossley, S. A., and McCarthy, P. M. Linguistic Features of Writing Quality. *Written Communication*. Forthcoming.
- McNamara, D. S., and Scott, J. L. 2001. Working Memory Capacity and Strategy Use. *Memory and Cognition*. 29: 10-17.
- Miller, G. A, Beckwith, R., Fellbaum, C., Gross, D. and Miller, K. 1990. Five Papers on WordNet. *Cognitive Science Laboratory*. Princeton University, No. 43.
- Paivio, A. (1965). Abstractness, Imagery, and Meaningfulness in Paired-Associate Learning. *Journal of Verbal Learning and Verbal Behavior*. 4: 32-38.
- Perfetti, C. A., Landi, N., and Oakhill, J. 2005. The Acquisition of Reading Comprehension Skill. In M. J. Snowling and C. Hulme eds. *The Science of Reading: A Handbook*. 227-247. Oxford: Blackwell.
- Rashotte, C. A., and Torgesen, J. K (1985). Repeated Reading and Reading Fluency in Learning Disabled Children. *Reading Research Quarterly*. 20: 180-188.
- Renyolds, M. 1984. Freewritings Origin. *The English Journal*. 73(3): 81-82.
- Toglia, M. P. and W. R. Battig (1978). *Handbook of Semantic Word Norms*. New York: Lawrence Erlbaum Associates.
- Whitten, I. A., and Frank, E. 2005. *Data Mining*. San Francisco: Elsevier.
- Wilson, M.D. 1988. The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*. 201: 6-11.