

## Number of Words versus Number of Ideas: Finding a Better Predictor of Writing Quality

Jennifer L. Weston, Scott A. Crossley\*, Philip M. McCarthy and Danielle S. McNamara

Institute for Intelligent Systems  
The University of Memphis  
Memphis, TN  
Jlweston; pmmccrth; dsmcnamr @Memphis.edu

\*Department of Applied Linguistics  
Georgia State University  
Atlanta, GA  
Scrossley@gsu.edu

### Abstract

This study examines the relation between the linguistic features of freewrites and human assessments of freewriting quality. This study builds upon the authors' previous studies in which a model was developed based on the linguistic features of freewrites written by 9<sup>th</sup> and 11<sup>th</sup> grade students to predict freewrite quality. The current study reexamines this model using number of propositions as a predictor instead of number of words because the number of propositions was expected to be a better proxy for number of ideas in contrast to simple text length. The results indicated that there were only slight advantages for using a measure for number of propositions, indicating that from an artificial intelligence perspective, the number of words was the better measure.

### Introduction

Writing proficiently is one of the most important skills that a person develops during their education. However, only 25% of students leave high school as proficient writers (NAEP, 2007), which could hinder their ability to be successful in higher education and the work place. Writing in the workplace is also important with over 90% of professionals responding that writing was essential to their jobs (Light, 2001). Considering the importance of writing to future success, it is important that steps be taken to increase writing proficiency for students leaving high school and college. Increasing proficiency can be done in a variety of ways; one way that has been shown to be successful has been the instruction of and use of strategies. Strategies aid writers by activating prior knowledge and by decreasing working memory demands. Additionally, the use of writing strategies focuses the writer on the steps needed to produce a successful written product. In conjunction with strategy instruction and use, students must be given ample opportunities to both practice writing strategies and write complete essays. Current demands on teachers and class sizes prevent teachers from providing enough of these opportunities to students (National Commission on Writing, 2003). To

facilitate opportunities for practice, an automated tutoring system, the Writing-Pal (McNamara et al., in press), has been developed to provide training on strategies to write argumentative essays.

Computational algorithms that predict writing quality are needed to provide feedback to students in the context of automated tutoring. One challenge in the development of these algorithms is the effect that text length plays in human assessments of writing quality. Across multiple corpora, longer writing samples receive higher scores, and the number of words is the most highly correlated index with quality ratings (Crossley & McNamara, in press). Some researchers have tried to circumvent this problem by instituting minimum word counts for automated graders. Such word counts help assure that writing samples are evaluated accurately regardless of length.

Using number of words as a predictor is also problematic because the number of words correlates highly with other linguistic measures that are theoretically important indicators of good writing such as type token ratio and overlap indices (McNamara, Crossley, and McCarthy, 2010). This is particularly problematic because the number of words is a surface feature that can be easily gamed by a writer.

By contrast, a measure of the number of propositions may provide an index that is a better proxy for the number of ideas and more strongly related to writing quality. Models of comprehension (e.g., Kintsch, 1998) assume that the fundamental unit of comprehension is the proposition, which consists of a predicate and argument. The proposition represents the underlying meaning of the explicit information in the text, discourse, or scene. A proposition generally consists of PREDICATE (ARGUMENT, ARGUMENT), where the arguments fill slots determined by the predicate (e.g., agent, object, instrument, goal). For example, the proposition *submit (she, paper, FLAIRS)* includes a predicate (submit) and three arguments including an agent (she), theme (paper) and goal (FLAIRS). Researchers have proposed that the number of propositions, as compared to number of words is a more accurate measure of text complexity (Kintsch and Keenan, 1973). Both number of words and number of propositions can be considered as proxies for the number of ideas in text. Howev-

er, the number of propositions is generally considered to be an index that is more reflective of a deeper level construct.

The present study focuses on writing samples using one common writing strategy, freewriting. Freewriting is a timed writing exercise during which the writer produces as many ideas as possible as quickly as possible with little regard for the conventional rules of writing (Elbow, 1979). The freewrites written for this study were *focused* freewrites. *Focused* freewrites are different from regular freewrites because the student writes using a topic or prompt rather than writing on anything that comes to mind. Freewriting is generally completed as a prewriting task and is often part of a larger planning process (Renyolds, 1984). Planning can take many forms including freewriting, outlining, concept mapping and list making and is generally the first step completed in a writing task.

Our goal in this study was to assess an alternative measure to number of words: namely, *number of propositions*. More specifically, this study assesses if a measure for the number of propositions can be used to build models that better predict freewriting quality. Additionally, we aim to assess the association between text length and number of ideas. Longer length is often associated with a greater number of ideas (Weston, Crossley, and McNamara 2010a, 2010b); however, a definitive link between these two factors has not been established. Discerning which linguistic feature provides the most predictive power for assessing freewriting quality will potentially improve automated assessments of writing quality. These tools can then allow educators and designers of intelligent tutoring systems to provide targeted feedback to writers engaging in freewriting and other essay practice tasks.

Recent studies by Weston et al. (2010a, 2010b) investigated linguistic aspects of freewrites indicative of freewrite quality. Freewrites were assessed by expert raters using a holistic scale similar to the SAT holistic essay scale and analyzed using the computational tool Coh-Metrix (McNamara and Graesser, in press). A quality freewrite developed ideas relevant to the prompt, used appropriate examples, and used 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. Weston et al. selected linguistic indices from Coh-Metrix based on the strength of the correlations between the indices and the holistic freewrite score and the absence of multicollinearity among the indices. Significant predictors for freewrite quality in these studies were *number of words* and *lexical overlap* (noun or stem). These two predictors explained 21% of the variance of the humans' ratings of freewrite quality in a test set. The Weston et al. results indicated that higher quality freewrites are longer and contain overlap in ideas between consecutive sentences. While these characteristics have been equated with more ideas, this has not been concretely established. The present study aims to directly relate number of ideas to quality in freewrites as well as determine if number of ideas is a stronger predictor of human freewrite quality ratings.

## Method

### Corpus Collection

Prompt-based freewrites were collected from high school students at a suburban public high school in upstate New York. The 105 students who participated were enrolled in either an 11<sup>th</sup> grade advanced placement English class or in a 9<sup>th</sup> grade Regents level (basic state level) English class (64 9<sup>th</sup> graders and 41 11<sup>th</sup> graders). These students ranged from 14 to 18 years of age. All students were taught by the same instructor (5 classes) who volunteered her classes to participate in this study. All students received the same instructions and materials.

The data used in the present study is part of a larger data set that also includes essays and questionnaires pertaining to student writing habits. The writing tasks were completed in a preselected order contained in a packet given to each student at the beginning of class. The experimental packets contained the following tasks: the freewriting instructions (adapted from Elbow, 1973), a 5-minute freewrite, a 25-minute essay, a 5-minute freewrite, and a final 5-minute freewrite (only for the 11<sup>th</sup> grade students). Each task was completed on a unique prompt with the exception that the prompt for the first freewrite was matched with an essay prompt. The experimenter read aloud the freewriting and essay instructions, timed the tasks, and informed students when to move onto the next task in their packets. A different number of freewrites was completed by the 11<sup>th</sup> grade participants due to time constraints stemming from the time required to distribute and explain materials.

The paired essay and freewrite were completed on one of two SAT style prompts, counter-balanced with the freewriting always being completed prior to the essay. Additional freewrites were completed on additional prompts selected from a pool of four prompts. The prompts used for this study were adapted from past SAT prompts obtained from [www.onlinemathlearning.com/sat-test-prep.html](http://www.onlinemathlearning.com/sat-test-prep.html). The essay instructions presented to students were adapted from the SAT writing section instructions (The College Board, 2009). The students' freewrites were transcribed as written, with the spelling and grammar errors retained. The 105 students produced 247 freewrites each of which was transcribed and analyzed for this study. The distribution of freewrites across the sessions was 104 1<sup>st</sup> freewrites, 104 2<sup>nd</sup> freewrites, and 39 3<sup>rd</sup> freewrites. One student's data was not included because the assignment was completed in Spanish instead of English.

Two composition instructors from Mississippi State University were trained as expert freewrite raters. Both raters had Master's degrees in English and at least three years of experience teaching English. Inter-rater reliability was assessed using Pearson correlations. The Pearson correlations on the training set exceeded .70 ( $p < .001$ ) and the average correlation between the two raters on the freewrites was .77 ( $p < .001$ ) with a weighted Kappa of .56, suggesting an acceptable level of agreement. The raters agreed on score for more than 56% of the freewrites and disagreed by 1 on 40% of the freewrites with a difference

of 2 points being seen on only 3% of the freewrites. When scores varied, a final score was computed using an average of the two given scores.

### Variable Selection

Coh-Metrix (McNamara and Graesser, 2010) and The Computerized Propositional Idea Density Rater (CPIDR version 3, Brown, et al., 2008) were used to examine the linguistic features of each freewrite in the corpus. Coh-Metrix is a computational tool used to assess text on over 600 linguistic and lexical indices. These indices are related to conceptual knowledge, cohesion, lexical difficulty, syntactic complexity, and simple incidence scores. Not all indices could be investigated because of the nature of the freewrites in the corpus. For example, many of the freewrites consisted of a single paragraph, making paragraph to paragraph comparisons impossible. In addition, many freewrites were fewer than 100 words, which is the minimum threshold recommended for lexical diversity measures (McCarthy and Jarvis, 2007).

This study uses the same training and test set split used in Weston et al. (2010b) to compare competing models. The training ( $n = 164$ ) and test set ( $n = 83$ ) contained exactly the same freewrites as used in the Weston et al. (2010b) study. Pearson correlations were used to assess which variables were predictive of freewrite quality in the corpus. The training set was used to identify which of the Coh-Metrix variables correlated highly to the expert ratings assigned to each freewrite. The variables identified in the training set were selected and used to predict the expert ratings in the training set by using a linear regression model. The same regression equation was used to evaluate the freewrites in the test set in order to test the accuracy of this model (Whitten and Frank, 2005).

Because the freewrites were transcribed without modifications (e.g., maintaining spelling errors), susceptibility of the linguistic and lexical measure to spelling mistakes was assessed. Susceptibility to spelling errors was assessed by comparing Coh-Metrix indices of lemmatized and original freewrites. The lemmatization process corrected for spelling mistakes and transformed each word into its root. If the target index from the normal and the lemmatized freewrite did not correlate at least at a .85 level, the index was considered to be susceptible to spelling errors. The measures and their respective indices are discussed below in reference to their importance to writing quality.

### Measures

**Number of Propositions.** The CPIDR program (Brown et al., 2008) measures the number of ideas in text by using a part-of-speech tagger to count parts of speech. Auxiliary verbs and appositive and modifying nouns are not included when counting propositions. The CPIDR reports the idea density of a text. A related ratio measure is calculated by dividing the number of propositions by the number of words.

**Syntactic Complexity.** Coh-Metrix measures syntactic complexity in three principal ways. The first measure calculates the mean number of words before the main verb. The second and third metrics assess 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 structural density, syntactic ambiguity, or ungrammaticality (Graesser et al., 2004). Consequently, more complex structures are more difficult to process and comprehend (Perfetti et al., 2005).

**Connectives and Logical Operators.** The density of connectives is measured in Coh-Metrix 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 Louwerse (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 (e.g., Longo, 1994).

**Causality.** Coh-Metrix measures causal cohesion by calculating the ratio of causal verbs to causal particles (Graesser et al., 2004). The causal verb count is based on the number of main causal verbs identified through WordNet (e.g., Fellbaum, 1998). Causal verbs and particles help the reader infer the causal relations in the text (Kintsch, 1998). A measure of causal verbs is investigated here to assess causal cohesion in freewrites.

**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 proportional intervals (including pronouns). Lexical overlap has been shown to aid in text comprehension and reading speed (e.g., Rashotte and Torgesen, 1985).

**Semantic Coreferentiality.** Coh-Metrix measures semantic coreferentiality using Latent Semantic Analysis (LSA; Landauer et al., 2007), a mathematical technique for representing deeper world knowledge based on large corpora of texts. 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 et al., 2007). Coh-Metrix also assesses given/newness through LSA by measuring the proportion of new information each sen-



tence provides (Hempelman et al., 2005). The given information is assumed to be recoverable from the preceding discourse (Halliday, 1967) and does not require reactivation (Chafe, 1975).

**Spatiality.** Coh-Metrix measures spatial cohesion using motion verbs and location nouns (Dufty, Graesser, Lightman et al., 2000). Classifications for both motion verbs and location nouns are taken from WordNet (Fellbaum, 1998). Spatial cohesion helps to construct a text and ensures that the situational model of the text (Kintsch & Van Dijk, 1978) is well structured and clearly conveys the text meaning to the reader.

**Word Characteristics.** Coh-Metrix reports on a variety of lexical indices taken from WordNet (e.g., Fellbaum, 1998) 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*). 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 imageability. All of these indices relate to the accessibility of core lexical items.

**Word Frequency.** Word frequency indicates how often particular words occur in the English language. Coh-Metrix utilizes frequency counts from CELEX (Baayen, Piepenbrock, and Gulikers, 1995) CELEX uses frequency counts based on the words in representative text corpora (the 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 14 measures correlated significantly with the expert ratings. The five variables that were selected, after assessing multicollinearity and susceptibility to spelling errors, along with their  $r$  and  $p$  values, are presented in Table 1 sorted by the strength of the correlation. Multicollinearity between variables was considered problematic when two variables correlated with each other above .70.

We also assessed multicollinearity between number of words and number of propositions. A Pearson Correlation between the two indices for the training set data yielded a correlation of  $r = .957$ .

Table 1

Selected Variables Based on Person Correlations	
Variable	$r$ value
Number of Ideas	0.682**
LSA Given Information	0.330**

Spatial Cohesion	0.300**
Stem Overlap	0.250*
Word Familiarity	-0.180*

\* $p < .05$ ; \*\*  $p < .001$

### Multiple Regression Training Set

A stepwise linear regression was conducted that regressed the five variables (number of ideas, LSA given information, spatial cohesion, stem overlap, word familiarity) onto raters' score for the 164 freewrites in the training set. The stepwise method was used to determine which indices were most predictive of expert ratings of freewrite quality.

The stepwise linear regression using the five variables yielded a significant model,  $F(2, 161) = 80.622$ ,  $p < .001$ ; adj.  $r^2 = .494$ . However, the only significant predictors were number of ideas ( $B = .044$ ,  $t(161) = 11.877$ ,  $p < .001$ ) and stem overlap ( $B = .592$ ,  $t(161) = 3.351$ ,  $p = .001$ ). The results from the stepwise linear regression demonstrate that these two variables account for 49% of the variance in the expert evaluations of freewriting quality for the training set. These results provide support for the notion that better freewrites contain more ideas and that these ideas are related to each other.

### Test Set Model

To further evaluate the regression model, the equation generated by the training set was applied to the test set to generate predicted scores (Predicted score =  $1.027 + (.044 \times \text{number of ideas}) + (.592 \times \text{stem overlap})$ ). A Pearson correlation between the predicted score and the actual score was conducted to assess the model. In addition, the adjusted  $r^2$  obtained from running the linear regression on the test data was used to demonstrate the strength of the model on an independent data set. Predicted scores for the test set significantly correlated with the actual scores,  $r = .455$ ,  $p < .001$ . The model for the test set yielded an adj.  $r^2 = .245$ ,  $p < .001$ . The results from the test set model demonstrate that the combination of these variables accounted for 25% of the variance in the evaluation of the 83 freewriting samples comprising the test set.

### Model Comparison

The Weston et al. (2010b) model using the same training and test set yielded a correlation with actual score of  $r = .469$ ,  $p < .001$ ; and an adj.  $r^2 = .210$ . The current model was able to account for 3.5% more variance in the test set than the model using number of words as a predictor. This difference demonstrates some statistical advantage, but is otherwise small.

However, another way of comparing the models is in a head-to-head comparison using the full corpus. The regression model for the entire corpus reported in Weston et al. (2010b) including number of words and stem overlap yielded,  $F(1,244) = 87.616$ ,  $p < .001$  adj.  $r^2 = .413$ . By contrast, the regression model including number of ideas and

stem overlap yielded,  $F(1,244) = 74.069, p < .001, \text{adj. } r^2 = .373$ . Thus, using this approach there was little difference in outcome, with some advantage statistically for using number of words.

## Discussion

This study compares the number of words and the number of ideas in freewrites as predictors of freewrite quality. Our hypothesis was that number of ideas would be more reflective of the underlying construct, and thus more predictive of quality.

A regression of linguistic features onto expert ratings of freewrite score indicated that among linguistic features, the only significant predictors of freewriting quality were the number of ideas and lexical overlap (i.e., stem overlap). The previous study by Weston et al. (2010b) had similarly included lexical overlap, but had used number of words rather than number of ideas. The Weston et al. (2010b) model using number of words showed slightly less predictive power in the test set than found in the current study. A gain of 3.5% in predictive power was observed with the use of number of ideas in comparison to number of words. However, a regression including the entire corpus indicated a slight advantage for number of words. Importantly, the two measures (number of words and ideas) are highly correlated in these freewrites and thus we might expect the results to fluctuate as demonstrated in our various analyses.

There is an important artificial intelligence question addressed in this study: In the context of W-Pal, when automatically assessing the quality of freewrites, should we use the number of words as a predictor, or should we use a more psychologically grounded measure for the number of ideas: the number of propositions? Using the more methodologically accepted method of training/test set, and relying on the test set as our benchmark, we might come to the conclusion that, indeed, the number of propositions is more advantageous. However, when we examine the full set, we see that number of words remains a contender. From an artificial intelligence perspective, the number of words is by far easier to implement (in terms of computational expense) and thus, the decision is relatively facile – one would use the number of words. Or, in the case of W-Pal, we will not choose to invest in implementing a more expensive alternative, number of propositions, at least not based on the current data.

This study also allows us to draw theoretical inferences about freewriting and the use of number of ideas as a predictor of freewriting quality. Focused freewriting is an activity completed for the purpose of generating a large number of ideas in a short period of time. Given this objective, it makes intuitive sense that number of ideas would be highly related to quality judgments of freewrites. However, the lack of substantial gains over the use of the measure of number of words along with the high correlations between number of words and ideas suggests that the number of words is roughly equivalent to the number of ideas in our corpus. Freewriting is designed to generate a large number of ideas; keeping with that goal, the rubric used by the

raters to score the freewrites focused on the number of ideas generated, suggesting that the ideas should be more predictive than overall length. The discrepancy between how number of ideas and number of words should intuitively differ in predictive power and how they do (or do not in some cases) may be explained by the nature of freewriting and the way that number of ideas is calculated.

Freewriting is completed without regard to the conventional rules of writing. In contrast, number of ideas is computed based on propositional phrases, which require proper syntactic subcategorization. If a student were not using prescriptive grammar or appropriate syntax, this may influence the number of ideas identified by CPIDR. Perhaps identifying the number of ideas in freewrites requires a different method than in conventional writing samples. For instance, if a student made a list, the tagger may not have been able to successfully identify all of the ideas included in that list.

Further work needs to be completed to ascertain the utility of this measure on other types of writing. If the measure is indeed limited by sentence structure then it may prove to be more predictive of quality in formal writing. It is important that as researchers automatically evaluate student writing samples, those evaluations are based on the properties theoretically tied to quality. Assessing the number of propositions, being more central to number of ideas and writing quality than mere length, might be a step towards achieving that objective.

## 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.
- Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., and Covington, M. 2008. Automatic Measurement of Propositional Idea Density From Part-of-speech Tagging. *Behavior Research Methods*. 40: 540-545.
- 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>.
- Crossley, S. A., Louwerse, 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.
- Crossley, S. A., and McNamara, D. S. . in press. Predicting Second Language Writing Proficiency: The Role of Cohesion, Readability, and Lexical Difficulty. *Journal of Research in Reading*. Elbow, P. 1973. *Writing without Teachers*. New York: Oxford University Press.
- Duffy, D. F., Graesser, A. C., Lightman, E., Crossley, S. A., & McNamara, D. S. 2006. An algorithm for detecting spatial cohesion in text. Paper presented at the 16th annual meeting of the Society for Text and Discourse, Minneapolis, MN.
- Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- 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.
- Hempelmann, C. F., Duffy, D., McCarthy, P. M., Graesser, A. C., Cai, Z. and McNamara, D. S. 2005. Using LSA to Automatically Identify Givenness and Newness of Noun Phrases in Written Discourse. In B. G. Bara, L. Barsalou, and M. Bucciarelli (Eds.), *Proceedings of the 27<sup>th</sup> Annual Conference of the Cognitive Science Society* (pp.941 - 946). Mahwah, NJ: Erlbaum.
- Kintsch, W. 1998. *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kintsch, W. and Keenan, J. M. 1973. Reading Rate and Retention as a Function of the Number of Propositions in the Base Structure of Sentences. *Cognitive Psychology*, 5: 257 - 274.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M. and Cai, Z. 2004. Coh-Metrix: Analysis of Text on Cohesion and Language. *Behavioral Research Methods, Instruments and Computers*. 36: 193 - 202.
- Kintsch, W., & Van Dijk, T. A. 1978. Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- 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.
- Louwerse, M. M. 2001. An Analytic and Cognitive Parameterization of Coherence Relations. *Cognitive Linguistics*. 12: 291-315.
- McCarthy, P. M., and Jarvis, S. 2007. A Theoretical and Empirical Evaluation of *VOCD*. *Language and Testing*, 24: 459-488.
- McNamara, D. S., Crossley, S. A., and McCarthy, P. M. 2010. Linguistic Features of Writing Quality. *Written Communication*, 27, 57-86.
- McNamara, D. S., & Graesser, A. C. 2010. Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution*. Hershey, PA: IGI Global.
- McNamara, D. S., Raine, R., Roscoe, R., Crossley, S., Jackson, G. T., Dai, J., Cai, Z., Renner, A., Brandon, R., Weston, J., Dempsey, K., Lam, D., Sullivan, S., Kim, L., Rus, V., Floyd, R., McCarthy, P. M., & Graesser, A. C. in press. The Writing-Pal: Natural language algorithms to support intelligent tutoring on writing strategies. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution*. Hershey, PA: IGI Global.
- 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.
- National Assessment of Educational Progress. 2007. *The Nation's Report Card: Writing 2007*. Retrieved Nov. 5 2010, [nces.ed.gov/nationsreportcard/writing/](http://nces.ed.gov/nationsreportcard/writing/)
- National Commission on Writing. 2003. *The Neglected "R."* NY: College Entrance Examination Board.
- 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.
- Weston, J. L., Crossley, S.A., and McNamara, D.S. 2010. Towards a Computational Assessment of Freewrite Quality. in H.W. Guesgen and C. Murray (Eds.), *Proceedings of the 23<sup>rd</sup> International Florida Artificial Intelligence Research Society (FLAIRS) Conference* (pp. 283-288). Menlo Park, CA: The AAAI Press.
- Weston, J. L., Crossley, S.A., and McNamara, D.S. 2010b. Computationally Assessing Expert Judgments of Freewrite Quality. In P.M. McCarthy and C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution*. Hershey, PA: IGI Global.
- 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.