Cohesion-Based Prompt Effects in Argumentative Writing

Scott A. Crossley[1], Laura K. Varner[2], and Danielle S. McNamara[2]

[1]Department of Applied Linguistics/ESL, Georgia State University, Atlanta, GA 30302
[2]Department of Psychology, Learning Science Institute, Arizona State University, Tempe, AZ 85287

scrossley@gsu.edu, laura.varner@asu.edu, dsmcnamara1@gmail.com

Abstract

Linguistic properties of writing prompts have been shown to influence the writing patterns contained in student essays. The majority of previous research on these prompt-based effects has focused on the lexical and syntactic properties of writing prompts and essays. The current study expands this research by investigating the effects of prompt cohesion on the cohesive features of student essays. Results indicate that prompt-based cohesion effects were observed for all the measured cohesion variables. Further, these cohesion prompt-effects were stronger than the effects observed for many lexical features and all syntactic features. Implications of these results in light of writing research are discussed.

Introduction

Prompt-based effects in essay writing are a well-known phenomenon (Huot, 1990). Numerous studies have demonstrated that the linguistic features found in a writing prompt can influence the writing patterns found in essays written on that prompt. Such effects are a concern because linguistic features are strong predictors of human ratings of writing quality (McNamara, Crossley, and McCarthy, 2010; McNamara, Crossley, and Roscoe, 2012). Thus, if a prompt promotes writers to produce infrequent words and complex syntax, these features may affect how a rater judges essay quality. Therefore, the rater may not be purely evaluating writer-based quality, but also prompt-based effects.

In this study, we are primarily interested in examining prompt-based effects that occur as a result of cohesion features in the prompt. Our goal is to investigate whether prompts that exhibit high cohesion lead to essays that exhibit high cohesion and vice-versa. Knowing that cohesive features can be significant indicators of writing quality, it is important to understand how writing prompts may affect the cohesive properties of essays.

Cohesion and Writing Quality

Cohesion refers to the presence or absence of explicit textual cues that allow the reader to make connections among the ideas in the text. For example, overlapping words and concepts between sentences indicate that similar ideas are being referred to across sentences, creating cohesive links. Cohesion is contrasted with *coherence*, which refers to the understanding that the reader derives from the text. This coherence may be dependent on a number of factors, such as text cohesion, prior knowledge and reading skill (McNamara et al., 1996).

Text cohesion is generally thought to be related to the coherence of an essay as can be seen in the literature about writing (e.g., DeVillez, 2003), writing textbooks (Golightly and Sanders, 1990), and intelligent tutoring systems that teach cohesion strategies to students (e.g., Writing-Pal, McNamara et al., 2012). However, empirical support for such assumptions has been mixed.

In two studies, Crossley and McNamara (2010, 2011) investigated the degree to which analytical rubric scores of essay quality (e.g., essay coherence, strength of thesis) predicted holistic essays scores. Results of both studies found that human judgments of text coherence were the most informative predictor of human judgments of essay quality. Both studies, however, found that computational indices of cohesion (e.g., indices of causal cohesion, spatial cohesion, temporal cohesion, connectives, and word overlap) computed by the Coh-Metrix tool (McNamara and Graesser, 2012) were not strongly correlated with human judgments of text coherence, indicating that cohesive devices may not underlie the development of coherent textual representations of essay quality.

While cohesive devices may not be strongly linked to coherence, there are some indications that cohesive properties of essays are important in predicting human judgments of essay quality, although such research has been mixed. For instance, some research has demonstrated no differences in cohesive devices between low and high quality essays (McNamara, Crossley, and McCarthy, 2010), and some research has indicated negative

correlations between temporal cohesion indices (Crossley and McNamara, 2012) and lexical overlap (McNamara, Crossley, and Roscoe, 2012) with human scores of essay quality. In contrast, McNamara, Crossley, and Roscoe (2012) reported that a cohesion feature related to givenness (i.e., the amount of given versus new information in text) was positively predictive of essay quality. In addition, Connor (1990) found that more proficient writers produced more cohesive devices than less proficient writers. Between grade levels, Crossley, et al. (2011) found that cohesive devices were important indicators of writing quality, but that more cohesive texts were produced by less skilled writers. As an example, 9th grade writers were more likely to produce texts with a higher incidence of positive logical connectives and more content word overlap than college freshmen. Overall, these studies indicate that cohesive devices are often important predictors of human judgments of essay quality but, in some cases, the presence of cohesive features leads to an essay that is assessed to be of a lower quality.

Prompt-based Writing Effects

Writing prompts are commonly used in independent and integrated writing assignments to provide the writer with a discourse mode and a topic, both of which can influence writing quality (Brown, Hilgers, and Marsella, 1991). However, prompts also provide writers with key words on which to focus and syntactic structures to emulate. These lexical and structural samples can prime the writer to produce specific words and syntactic patterns, both of which can affect human judgments of writing quality. Thus, equivalence of topics has been one of the main goals in writing assessment, since judgments of writing ability can become biased and problematic if some writers write on easier topics and others write on the harder ones (Crossley, in press).

The majority of the research looking at prompt-based linguistic effects has focused on syntactic properties. For instance, Crowhurst and Piche (1979), Tedick (1990), and Hinkel (2002) found that prompt had a significant effect on the syntactic complexity of essays. Crowhurst and Piche (1979) reported significant differences in the mean length of T-units (defined as a dominant clause and its dependents) and mean number of clauses per T-unit for essays written in response to different writing prompts. Tedrick (1990) found that more specific prompts produced higher mean length of T-units and mean length of error-free T-units as compared to more general prompts. Hinkel (2002) reported prompt-based differences in present tense verbs, number of infinitives, *be* copulas, and phrase level conjunctions. The findings from these studies indicate that different prompts can lead to greater syntactic complexity on the part of writers.

Fewer studies have examined prompt-based differences in relation to lexical output. Hinkel (2002) found that certain prompts led to the production of greater number of nominalizations. Crossley et al. (2011) found that between two prompts, significant differences were reported in word specificity (i.e, hypernymy) and word familiarity. These studies demonstrate that the words in a prompt can lead to greater or less lexical sophistication.

In general, such findings indicate the structure and wording of writing prompts can have important consequences for writing production and writing quality. Such consequences exist not only for particular writing prompts, but also for writing contexts and writing populations (Hout, 1990). However, little to no work has assessed cohesion-based prompt differences and their potential effects on human judgments of quality.

Method

In this study we are primarily interested in investigating possible effects of prompt-based cohesion on writing samples. Our hypothesis is that prompts that are high in cohesion may lead to written essays that are also high in cohesion. We collected argumentative essays written on seven different prompts. To assess the strength of prompt-based cohesion effects in comparison to lexical and syntactic effects, we used cohesion, lexical, and syntactic indices taken from the computational tool Coh-Metrix. This analysis was followed by a cohesion-specific analysis in which we calculated a composite score for the cohesion features in the seven prompts. This cohesion composite score was then used to classify prompts as being either high or low cohesion prompts. We then calculated the cohesive features found in the argumentative essays for each prompt using Coh-Metrix. The essay specific cohesion features were then used to classify the essays as being written on either high or low cohesion prompts.

Table 1
Descriptive statistics for corpus

Short prompt	n	Grade level	Region
Competition	126	10th	Washington DC
Fitting In	35	13th	Tennessee
Heroes	158	13th	Mississippi
Images	126	10th	Washington DC
Memories	45	9th and 11th	New York
Optimism	56	9th and 11th	New York
Uniqueness	155	13th	Mississippi

Corpus

We collected 701 argumentative essays written by students at four different grade levels: 9th grade, 10th grade, 11th grade, and college freshman (i.e., 13th grade) from four different geographic areas. The essays were written in response to seven different prompts used in the Scholastic

Achievement Test (SAT) writing section. The prompts were independent writing prompts that did not require domain knowledge. In all cases, students were allowed 25 minutes to write the essay. Descriptive statistics for the corpus are presented in Table 1.

Coh-Metrix

We used the computational tool Coh-Metrix (McNamara and Graesser, 2012) to analyze the linguistic features of the essays. We selected indices from Coh-Metrix that measure properties at the word (lexical sophistication), sentence (syntactic complexity), and cohesion levels. The selected indices are discussed below.

Lexical indices. We used Coh-Metrix to calculate lexical scores for word concreteness, word frequency, and lexical overlap. Word concreteness refers to here-and-now concepts, ideas, and things (Toglia and Battig, 1978). Coh-Metrix calculates word concreteness using human word judgments taken from the MRC Psycholinguistic Database (Wilson, 1988). Word frequency indices measure how often particular words occur in the English language. Coh-Metrix reports frequency counts taken from the CELEX database (Baayen, Piepenbrock, and Gulikers, 1995), which consists of frequencies taken from the early 1991 version of the COBUILD corpus. Lexical diversity indices reflect type-token ratios (TTR; Templin, 1957) with higher TTR indicating more lexical variety. Coh-Metrix calculates lexical diversity through a number of sophisticated algorithms that control for text length effects including the Measure of Text Length and Diversity (MTLD; McCarthy and Jarvis, 2010).

Syntactic indices. We used Coh-Metrix to calculate indices of syntactic complexity. The indices we selected were mean number of words before the main verb and syntactic similarity. Higher mean number of words before the main verb indicates greater syntactic complexity. Syntactic similarity is measured in Coh-Metrix by calculating the consistency and uniformity of the clausal, phrasal, and part of speech constructions located in the text (i.e., a text's sentence variety).

Cohesion indices. We used Coh-Metrix to calculate indices of cohesion. The indices we selected all related to structural overlap or similarity measures. These indices included lexical overlap, semantic overlap, and sentential positioning. Lexical overlap refers to the extent to which words and phrases overlap across sentences and text. Greater overlap results in greater text cohesion (Kintsch and van Dijk, 1978). Coh-Metrix considers four forms of lexical overlap between sentences: noun overlap, argument overlap, stem overlap, and content word overlap. Semantic overlap refers to the extent to which words, phrases, and sentences overlap semantically across text. Coh-Metrix measures semantic overlap using Latent Semantic Analysis (LSA), a mathematical and statistical technique for representing deeper world knowledge based on large corpora of texts (Landauer et al., 2007). Coh-Metrix also uses LSA to calculate text givenness, which is information that is recoverable from the preceding discourse (Halliday 1976). For sentential positioning, Coh-Metrix computes the Minimal Edit Distance (MED) for a text sample by measuring differences in the sentential positioning of content words, lemmas, and phrase structures. A high MED value indicates that words and phrases are located in different places within sentences across the text, suggesting lower structural cohesion.

Analyses

We used Coh-Metrix to calculate cohesion scores for the prompts associated with the essays. We calculated a composite score for the cohesion features in each prompt by averaging the scores for cohesion indices in the following groups: lexical overlap, semantic overlap, and MED. Each prompt was then categorized as low, medium, or high for the specific cohesion feature. We also used Coh-Metrix to calculate lexical, syntactic, and cohesion scores for the essays in the corpus. We used these scores to first examine the strength of cohesion prompt-based differences as compared to lexical and syntactic prompt-based difference using a Multivariate Analysis of Variance (MANOVA) that included grade level as a between subjects covariate. We then used the cohesion scores for the individual essays to predict whether an essay was written on a low or high cohesion prompt. To accomplish this, we first conducted a MANOVA analysis for each group of cohesion features (i.e., lexical overlap, semantic overlap, and MED) that factored grade level as a between subjects covariate. We then conducted a Discriminant Function Analysis (DFA) using leave-one-out cross-validation techniques to classify essays as written on high or low cohesion prompts.

Results

Comparison Prompt-Based Effects Analysis

To compare cohesion-based prompt effects to lexical and syntactic, we used the selected lexical and syntactic indices discussed earlier and one prototypical index from each of cohesion groupings in the analysis. The selected cohesion indices were argument overlap (for our index of lexical overlap), LSA givenness (for our index of semantic overlap), and MED all lemmas (for structural overlap).

The MANOVA from this analysis examined differences in the linguistic features for the essays written on the seven prompts (i.e., prompts were the independent variables and the linguistic features were the dependent variables) and included grade level as a between subjects covariate. The analysis showed that all linguistic features demonstrated significant differences among the prompts. The strongest prompt-based effects were reported for lexical and cohesion indices. The effect sizes reported in Table 2 indicates that prompt-based cohesion effects are relatively

strong in comparison to lexical and syntactic prompt-based effects.

Prompt-Based Cohesion Effects Analyses

We next analyzed the cohesion features of the essays to examine if cohesion indices could be used to classify the essays as being written on prompts that exhibited low or high cohesion.

Table 2

ANOVA results for linguistic indices among prompts: f value, p value, and hp2

Index (category)	*f* value	*p* value	*hp2*
Word concreteness (L)	32.528	< .001	0.219
LSA givenness (C)	22.640	< .001	0.164
Word frequency (L)	18.774	< .001	0.140
Argument overlap (C)_	17.240	< .001	0.130
Lexical diversity: D (L)	17.350	< .001	0.130
Minimal edit distance (C)	8.705	< .001	0.070
Syntactic similarity (S)	8.061	< .001	0.065
Causality (C)	4.206	< .001	0.035
Syntactic complexity (S)	3.372	< .010	0.028
Incidence of connectives (C)	2.121	< .050	0.018
L = lexical index; S = syntactic index; C = cohesion index			

Lexical overlap indices. Of the seven prompts, two were rated as containing low lexical overlap (*images* and *uniqueness*) and two were rated as containing high lexical overlap (*competition* and *fitting in*). This provided us with 281 essays written on low cohesion prompts and 161 essays written on high cohesion prompts. We used four lexical overlap indices to classify the essays as being written on prompts low or high in lexical overlap. These indices were content word overlap, argument overlap, stem overlap, and noun overlap.

Table 3

Descriptive statistics essay overlap indices

Index	Low prompts	High prompts
Content word overlap	0.086 (0.036)	0.105 (0.039)
Argument overlap	0.415 (0.158)	0.470 (0.164)
Stem overlap	0.365 (0.162)	0.402 (0.176)
Noun Overlap	0.259 (0.145)	0.310 (0.159)

The MANOVA analysis with grade level as a covariate indicated that all indices except stem overlap demonstrated significant differences between the low and high essays (see Table 3 for descriptive statistics for these indices and Table 4 for MANOVA results). The three indices that demonstrated significant differences were used as predictor variables in a DFA. The results demonstrate that the DFA correctly allocated 282 of the 442 essays as belonging to either low or high lexical overlap prompts, $\chi 2$ (df=1, n=442) = 29.968, p < .001, for an accuracy of 63.8%. The measure of agreement between the actual and predicted category assigned by the model produced a Cohen's Kappa of 0.260, demonstrating a fair agreement. The results

indicate that prompts higher in lexical overlap produced essays higher in lexical overlap.

Table 4

ANOVA results for overlap indices among prompts: f value, p value, and hp2

Index	*f* value	*p* value	*hp2*
Content word overlap	29.436	< .001	0.063
Argument overlap	5.702	< .050	0.013
Stem overlap	2.645	> .050	0.006
Noun Overlap	4.049	< .050	0.009

Semantic overlap indices. Of the seven prompts, four were rated as containing low semantic overlap (*fitting in, images, memories,* and *uniqueness*) and three were rated as containing high semantic overlap (*competition, heroes* and *optimism*). This provided us with 361 essays written on low cohesion prompts and 340 essays written on high cohesion prompts. We used three semantic overlap indices to classify the essays as being written on prompts low or high in semantic overlap. These indices were LSA sentence to sentence, LSA sentence to paragraph, and LSA givenness.

Table 5

Descriptive statistics essay LSA indices: Mean (SD)

Index	Low prompts	High prompts
LSA sentence to sentence	0.197 (0.067)	0.235 (0.078)
LSA sentence to paragraph	0.177 (0.065)	0.217 (0.078)
LSA givenness	0.306 (0.039)	0.335 (0.046)

Table 6

ANOVA results for LSA indices among prompts: f value, p value, and hp2

Index	*f* value	*p* value	*hp2*
LSA sentence to sentence	46.111	< .001	0.062
LSA sentence to paragraph	51.217	< .001	0.068
LSA givenness	88.717	< .001	0.113

The MANOVA analysis with grade level as a covariate indicated that all indices demonstrated significant differences between the low and high essays (see Table 5 for descriptive statistics for these indices and Table 6 for MANOVA results). The three LSA indices were used as predictor variables in a DFA. The results demonstrate that the DFA using the three semantic overlap variables correctly allocated 446 of the 701 essays as belonging to either low or high lexical overlap prompts, $\chi 2$ (df=1, n=701) = 51.725, p < .001, for an accuracy of 63.5%. The measure of agreement between the actual and predicted category assigned by the model produced a Cohen's Kappa of 0.272, demonstrating a fair agreement. The results indicate that prompts higher in semantic overlap produced essays higher in semantic overlap.

Minimal edit indices. Of the seven prompts, three were rated as containing low sentential positioning (*competition, fitting in,* and *images*) and two were rated as containing high sentential positioning (*heroes* and *memories*). This provided us with 287 essays written on prompts low in sentential positioning and 203 essays written on prompts high sentential positioning. We used three MED indices to classify the essays as being written on prompts low or high in sentential positioning. These indices were MED all words, MED all words tagged, and MED lemmas.

The MANOVA analysis with grade level as a covariate indicated that two of the three indices demonstrated significant differences between the low and high essays (see Table 6 for descriptive statistics for these indices and Table 7 for MANOVA results). The two significant MED indices were used as predictor variables in a DFA. The results demonstrate that the DFA using the two MED indices correctly allocated 300 of the 490 essays as belonging to either low or high cohesion prompts, χ^2 (df=1, *n*=490) = 22.251, p < .001, for an accuracy of 61.2%. The measure of agreement between the actual and predicted category assigned by the model produced a Cohen's Kappa of 0.212, demonstrating a fair agreement. Results indicate that prompts higher in sentential positioning produced essays higher in sentential positioning.

Table 6
Descriptive statistics essay MED indices: Mean (SD)

Index	Low prompts	High prompts
MED all words	0.658 (0.040)	0.661 (0.036)
MED all words tagged	0.889 (0.035)	0.881 (0.035)
MED all lemmas	0.865 (0.035)	0.849 (0.037)

Table 7
ANOVA results for MED indices among prompts: f value, p value, and hp2

Index	*f* value	*p* value	*hp2*
MED all words	0.025	> .050	0.000
MED all words tagged	5.436	< .050	0.011
MED all lemmas	11.867	< .001	0.024

Discussion and Conclusion

Given that writing assessments are widely used to evaluate language and writing ability, it is critical to examine how elements of these tests may impact student performance. In this study, we investigate the degree to which cohesion features of writing prompts affected the overall cohesion in students' essays. Results indicated that there are strong cohesion-based prompt effects. These effects hold for all cohesion measures that we analyzed (i.e., lexical overlap, semantic overlap, and sentential overlap). Notably, many of the cohesion indices also showed stronger prompt-based effects than our selected lexical and syntactic indices. These findings indicate that cohesion-based prompt effects

are as prevalent as other linguistic-based prompt effects and that such effects may cause writers to emulate the cohesion features in given prompts. Because cohesion features can be important predictors of writing quality (Crossley and McNamara, 2012; McNamara et al., 2012), controlling for such prompt-based effects becomes an important element of writing assessment.

A comparison between the cohesion features and the lexical and syntactic indices indicates that only one lexical feature (word concreteness) demonstrated a stronger effect size across prompts than any of the cohesion features. All syntactic features demonstrated lower effect sizes than our selected cohesion features. Thus, we can say with some confidence that prompt-based effects are strongest for lexical and cohesion features. In contrast to previous research that has primarily focused on syntactic prompt effects, this finding indicates a large research gap in the prompt-based writing literature. Additionally, it calls into question the findings from a number of writing quality studies that may have controlled for lexical and syntactic prompt-based differences, but not cohesive differences.

The strongest findings reported in this study were for lexical overlap indices. Three of these indices, related to content, stem, and argument overlap, were able to predict whether an essay was written on a low or high cohesion prompt with an accuracy of 64%. Similar, but lower findings, were reported for our semantic overlap and minimal edit distance indices. In all cases, the findings confirmed that prompts higher in cohesive properties lead to essays being written that contained higher cohesive values. Such a finding indicates that some of the linguistic properties found in prompt-based argumentative essays are not the result of writer choices, but rather primed by the prompt. In this sense, some of the linguistic features found in an essay may not accurately reflect a writer's proficiency level, but rather reflect the properties of the prompt. Ratings of students' writing proficiency are often used to make important educational decisions, such as university acceptance in the case of SAT writing samples or course grades and graduation in the case of classroom assignments. In the absence of controlling for cohesion-based prompt effects, such decisions may be unsupported.

For the most part, the findings from this analysis seem to hold across grade-levels. Therefore, the cohesion-based prompt effects we report in this study seem to be generalizable to a wide population of writers that range from young adolescents to college freshman. This indicates that writers at all levels may use the prompt to provide them with not only discourse modes and topics on which to write, but also with linguistic cues for producing cohesive writing samples (not to mention lexically and syntactically similar samples). Although it is possible, it is highly unlikely that writers are intentionally copying the writing style found in the prompts; rather, it is more probable that in reading and referring back to the prompt, the writer is primed to subconsciously produce language that matches the style of the prompt language. More experiments are necessary to support such an assertion, but there is strong

evidence that the linguistic features in the prompt influence the words and structures that writers place in an essay. This raises questions about the reliability of prompt-based writing and the validity of human judgments of writing quality that may be influenced by prompt differences.

Acknowledgments

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, J. D., Hilgers, T., and Marsella, J. 1991. Essay Prompts and Topics. *Written Communication* 8: 533-556.

Connor, U. 1990. Linguistic/Rhetorical Measures for International Persuasive Student Writing. *Research in the Teaching of English* 24: 67-87.

Crossley, S. A. in press. Advancing Research in Second Language Writing through Computational Tools and Machine Learning Techniques: A Research Agenda. *Language Teaching.*

Crossley, S. A., and McNamara, D. S. 2010. Cohesion, Coherence, and Expert Evaluations of Writing Proficiency. In Proceedings of the 32nd Annual Conference of the Cognitive Science Society, 984-989. Austin, TX: Cognitive Science Society.

Crossley, S. A., and McNamara, D. S. 2012. Predicting Second Language Writing Proficiency: The Role of Cohesion, Readability, and Lexical Difficulty. *Journal of Research in Reading* 35: 115-135.

Crossley, S. A., and McNamara, D. S. 2011. Text Coherence and Judgments of Essay Quality: Models of Quality and Coherence. In Proceedings of the 29th Annual Conference of the Cognitive Science Society, 1236-1241. Austin, TX: Cognitive Science Society.

Crossley, S. A., Weston, J., Sullivan, S. T., and McNamara, D. S. 2011. The Development of Writing Proficiency as a Function of Grade Level: A Linguistic Analysis. *Written Communication* 28: 282-311.

Crowhurst, M. C., and Piche, G. L. 1979. Audience and Mode of Discourse Effects on Syntactic Complexity in Writing on Two Grade Levels. *Research in the Teaching of English* 13: 101-109.

DeVillez, R. 2003. *Writing: Step by step*. Dubuque, IO: Kendall Hunt.

Golightly, K. B., and Sanders, G. 2000. *Writing and Reading in the Disciplines* (2nd Ed.). New Jersey: Pearson Custom Publishing.

Halliday, M. A. K., and Hasan, R. 1976. *Cohesion in English.* London, England: Longman.

Hinkel, E. 2002. *Second Language Writers' Text*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hinkel, E. 2009. The Effects of Essay Topics on Modal Verb Uses in L1 and L2 Academic Writing. *Journal of Pragmatics* 41: 667-683.

Huot, B. 1990. The Literature of Direct Writing Assessment: Major Concerns and Prevailing Trends. *Review of Educational Research* 60: 237-263.

Landauer, T., McNamara, D. S., Dennis, S., and Kintsch, W. eds. 2007. *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Erlbaum.

Kintsch, W., and van dijk, T. A. 1978. Toward a Model of Text Comprehension and Production. *Psychological Review* 85: 363-394.

McCarthy, P. M., and Jarvis, S. 2007. *vocd*: *A theoretical and empirical evaluation. Language Testing* 24: 459-488.

McNamara, D. S., Crossley, S. A., and McCarthy, P. M. 2010. The Linguistic Features of Quality Writing. *Written Communication* 27: 57-86.

McNamara, D. S., Raine, R., Roscoe, R., Crossley, S., Jackson, G. T., Dai, J., Cai, Z., Renner, A., Brandon, R., Weston, J., Dempsey, K., Carney, D., Sullivan, S., Kim, L., Rus, V., Floyd, R., McCarthy, P. M., and Graesser, A. C. 2012. The Writing-Pal: Natural Language Algorithms to Support Intelligent Tutoring on Writing Strategies. In *Applied Natural Language Processing and Content Analysis: Identification, Investigation, and Resolution*, 298-311. Hershey, PA: IGI Global.

McNamara, D., and Graesser, A. 2012. Coh-Metrix: An Automated Tool for Theoretical and Applied Natural Language Processing. In *Applied natural language processing and content analysis: Identification, investigation, and resolution*, 188-205. Hershey, PA: IGI Global.

McNamara, D. S., Kintsch, E., Butler-Songer, N., and Kintsch, W. 1996. Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning from Text. *Cognition and Instruction* 14: 1-43.

Tedick, D. J. 1990. ESL Writing Assessment: Subject-Matter Knowledge and its Impact on Performance. *English for Specific Purposes* 9: 123–143.

Templin, M. 1957. *Certain Language Skills in Children*. Minneapolis: University of Minnesota Press.

Toglia, M., and Battig, W. 1978. *Handbook of Semantic Word Norms*. New York: Erlbaum.

Wilson, M. D. 1988. The MRC Psycholinguistic Database: Machine-Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments, and Computers* 201: 6-11.