Linguistic Signatures of Cognitive Processes during Writing

Laura K. Allen (LauraKAllen@asu.edu)
Cecile Perret (cperret@asu.edu)
Danielle S. McNamara (dsmcnama@asu.edu)
Arizona State University, Department of Psychology, P.O. Box 872111
Tempe, AZ 85281 USA

Abstract

The relationship between working memory capacity and writing ability was examined via a linguistic analysis of student essays. Undergraduate students (n = 108) wrote timed, prompt-based essays and completed a battery of cognitive assessments. The surface- and discourse-level linguistic features of students' essays were then analyzed using natural language processing tools. The results indicated that WM capacity was related to surface-level, but not discourse-level features of student essays. Additionally, the results suggest that these relationships were attenuated for students with high inferencing skills, as opposed to those with lower inferencing skills.

Keywords: writing; natural language processing; computational linguistics; strategies; working memory

Introduction

Writing is a complex cognitive and social process that involves the production of texts for the purpose of conveying meaning to others (Graham, 2006). This task involves cognitive processes such as accessing vocabulary knowledge and constructing grammatical sentences, knowledge (e.g., of language, the writing process, and the domain), and the ability to strategically use language to connect and present ideas in a meaningful way (Donovan & Smolkin, 2006; Graham, 2006; McNamara, 2013).

While a clear demarcation cannot be placed between levels of processes (McNamara, Jacovina, & Allen, 2015), one dominating question in the writing literature regards the relative roles of lower- and higher-level cognitive processes. Evidence considered for the role of lower-level processes generally comes from studies that examine relations between writing ability and working memory (WM) capacity (Kemper, Rash, Kynette, & Norman, 1990; Hoskyn & Swanson, 2003; McCutchen, Covill, Hoyne, & Mildes, 1994). Evidence considered for *higher-level* processes comes primarily from studies showing that skilled writers have more knowledge about writing norms and are more strategic, as well as from studies that show the effectiveness of writing strategy interventions (Graham & Perin, 2007).

Importantly, evidence for lower- and higher level skills tends to come from separate studies and different research camps in the writing literature. Few writing researchers have explored the *relative* influences of lower- and higher-level skills, nor have they considered the potential for higher-level skills to mitigate the impact of lower-level factors. The goal of this study is to address these gaps in the literature by examining the linguistic signatures of cognitive processes in students' essays and their relations to WM capacity.

Working Memory Capacity and Writing

Within the writing literature, WM capacity has received considerable attention and is commonly labeled as a central component of the writing process (Berninger & Swanson, 1994; Hayes, 1996; Kellogg, 2001; 2008; McCutchen, 1996). The measured capacity of an individual's WM has been theorized to relate to their writing ability because of the complex and resource-demanding nature of the task. Writing requires individuals to engage in multiple separable processes, such as accessing word knowledge, planning, activating prior knowledge about a particular domain, and making connections between ideas, all of which can place extreme demands on the cognitive system (Kellogg, 2001).

Notably, however, the link between WM capacity and writing skill has failed to be consistently supported by the literature. Although some studies have reported positive correlations between performance on WM tasks and writing quality (Babayigit & Stainthorp, 2011; Berninger & Swanson, 1994; Kellogg, 2008), others have found this relationship to be non-significant or even negative (Allen et al., 2014; Dixon, LeFevre, & Twilley, 1988).

One potential explanation for this conflicting evidence lies in the characteristics of the rubrics by which essays are assessed. The measured capacity of an individual's WM may be inconsistently related to essay quality because the definition of writing quality has not been operationalized consistently (i.e., it might focus more or less on different text properties). If this is the case, in order to understand what role (if any) WM capacity plays in the writing process, it is important to conduct analyses at multiple levels of discourse. Our supposition is that multi-dimensional discourse analyses can potentially deepen our understanding of individual differences in writing.

The Role of Higher-Level Skills

In addition to considering texts at multiple levels, researchers can benefit from a consideration of the interactions that potentially occur amongst the lower- and higher-level skills that students have developed. Despite the stronger emphasis on lower-level skills in cognitive writing research, evidence from educational research suggests that the development and use of higher-level skills (e.g., strategies) can significantly reduce the demands of the writing process and enhance writing performance.

For example, in a meta-analysis conducted on over 120 published studies of writing interventions, Graham and Perin (2007) found that strategy instruction was the most

effective form of writing instruction. Given these findings, it may be the case that the writing process is primarily constrained by WM in the absence of higher-level skills. However, once students have been trained to employ strategic processes during writing, this relationship between WM and writing ability may be significantly reduced.

Current Study

We adopt a different theoretical and methodological approach than is considered typical of studies in the writing literature. Our theoretical approach is motivated by research in cognitive science, which supposes that the development of higher-level skills, such as the ability to generate inferences, can help to make up for deficits in knowledge and cognitive capacity (e.g., McNamara & Scott, 2001).

Our methodological approach is inspired by research showing that the linguistic properties of texts reflect readers' potential levels of comprehension as measured using indices at multiple levels (e.g., Graesser & McNamara, 2011). To investigate writing, our approach is to consider the notion that there are multiple linguistic dimensions of the texts that students produce. *Surface-level* text features relate to the characteristics of the words and sentences in texts. Variations in these features can alter the style of the essay, as well as influence its readability and perceived sophistication. *Discourse-level* features, on the other hand, go beyond the individual words and sentences, and instead reflect aspects of the situation model portrayed by the text such as the degree of narrativity in the essay.

Our first hypothesis is that the surface-level features of students' essays will be related to their WM scores. Second, we hypothesize that the role of WM capacity in the writing process will be moderated by the development of strategic inferencing skills (i.e., higher-level cognitive skills). Thus, the relations between students' WM scores and the characteristics of their essays may be more or less pronounced depending on the degree to which they have developed abilities to think and write strategically.

We investigate these hypotheses through a linguistic analysis of student essays. We first examine relations between students' WM scores and the properties of their essays at both the surface- and discourse-levels of the text. Next, we examine whether these relationships differ as a function of students' inferencing skills.

Our research questions are listed below:

1) Do WM scores demonstrate significant relations to surface-level and/or discourse-level linguistic properties of students' essays?
2) Do these potential relations between WM scores and text properties vary as a function of students' inferencing skills?

Methods of Automated Text Analysis

To calculate the linguistic properties of students' essays (see Table 1 for the indices calculated), we used two natural language processing (NLP) tools: Coh-Metrix (McNamara, Graesser, McCarthy, & Cai, 2014) and the Writing Assessment Tool (WAT; McNamara, Crossley, & Roscoe, 2013). Both tools report hundreds of linguistic indices that relate to the structure of the text, its general readability, rhetorical patterns, lexical choices, and cohesion using a combination of components that are commonly used in NLP tools (Crossley, Allen, Kyle, & McNamara, 2014).

For the purposes of the current analysis, we pre-selected 20 indices from Coh-Metrix and WAT, all of which had theoretical links to writing quality. These indices related to two primary essay levels (each containing 10 variables): surface-level and discourse-level text features. The surface-level features relate primarily to the characteristics of the individual words and sentences in the text, whereas discourse-level features relate to text cohesion and rhetorical functions. It is important to note that, although we have grouped the linguistic features into two distinct categories, they lie more realistically on a continuum. Thus, some of the discourse-level features will tend more towards the surface level than others (and vice versa). The indices selected for these two categories are described briefly below. For more thorough descriptions of these indices and their theoretical links, see McNamara et al. (2014).

Surface-level Text Indices The surface-level text indices selected describe word- and sentence-level characteristics of students' essays. Indices were broadly selected to account for a multitude of independent constructs related to texts at these levels. These indices range from simple frequency counts for certain parts-of-speech to more informative measures that describe the types of words used in a text.

Word Information. Coh-Metrix and WAT calculate multiple indices that describe the specific types of words used in texts. Word frequency measures, for instance, are used to assess how frequently certain words occur in the English language. Coh-Metrix reports indices of word frequency that are taken from the CELEX database. Coh-Metrix additionally reports the logarithm of word frequency for all words in a text and the minimum log word frequency for content words. An index of log frequency is calculated because reading times are typically linearly related to the logarithm of word frequency (rather than the raw word frequency; Haberlandt & Graesser, 1985). Coh-Metrix, therefore, provides the average minimum log frequency of words across sentences (minimum log word frequency). Average measures of word frequency are important indicators of lexical knowledge and can inform text readability because frequent words are more easily accessed and decoded than less frequent words (Perfetti, 1985).

Additionally, Coh-Metrix employs WordNet to calculate polysemy and hypernymy scores for all content words in a text. *Polysemy* scores denote the number of senses that are associated with a given word (ambiguous words have more senses). Polysemy scores provide indications of lexical proficiency (McNamara et al., 2014). *Hypernymy* is indicative of the specificity of a given word – as defined by its location within a conceptual hierarchy (e.g., *dog* would

have a lower hypernymy value than *poodle*). Hypernymy scores have been linked to lexical knowledge and production (Crossley & McNamara, 2009).

Part-of-speech and Sentence Information. Coh-Metrix and WAT additionally contain multiple indices that describe the features of the sentences in texts, such as the parts of speech they contain and the complexity of their constructions. Coh-Metrix reports incidence scores for all of the part-of-speech tags in the Penn Tree Bank Tag Set. This set includes tags at the word and phrase levels for content items, as well as for function items. These tags have been used in previous studies to classify high and low quality essays (Crossley & McNamara, 2011).

In the current study, we selected part-of-speech tags that were related to function items (to assess the surface-level grammatical structures of the text, rather than specific content) and that target independent grammatical constructs. We examine the incidence of modals (might, could), the incidence of prepositions (on, in, around, between) and subordinating conjunctions (after, because, whereas, unless), the incidence of first person pronouns (I, me), the incidence of third person pronouns (he, she), the incidence of causative subordinators (because, as), and the incidence of phrasal coordinators (and, but).

Sentence complexity is measured in Coh-Metrix with multiple indices. Higher quality essays typically contain more complex syntactic constructs (McNamara et al., 2014), which can increase WM load on readers (Graesser et al., 2006). However, it is unclear whether these syntactic constructions are related to the WM of the *writer.* We used the index *mean number of words before the main verb* as a proxy for sentence complexity.

Discourse-level Text Indices The discourse-level indices relate to the cohesion and semantic properties of texts. As with the surface-level indices, these measures were broadly selected in order to target multiple constructs. The final set of indices ranges from basic measures of cohesion to more robust component indices related to the style of the text.

Connectives. Coh-Metrix provides an incidence score for the number of connectives that are contained in a text. Connectives increase cohesion because they explicitly link ideas and clauses (Longo, 1994). Coh-Metrix also provides indices related to specific categories of connectives. Because the essays in this study were argumentative essays that rely on logical argumentation, we chose to analyze the incidence of logical connectives to serve as a measure of cohesion.

Lexical Overlap. Cohesion is also calculated through indices of overlap for certain parts-of-speech. Relevant to the current study, argument overlap calculates how often two sentences share nouns with common stems. Lexical overlap increases the readability of a given text (Kintsch & van Dijk, 1978).

Situation Model Cohesion. Coh-Metrix calculates multiple indices that attempt to tap into text cohesion beyond the word level. For instance, WordNet is used to assess spatial cohesion using two forms of information: location information and motion information. Location information is represented through nouns such as *school* or *Indiana*, whereas motion information is represented through verbs such as *fight* or *run*. Similarly, intentional cohesion is measured by the ratio of intentional particles to intentional verbs. Both of these cohesion measures help to increase the readability of a text by promoting the successful generation of inferences by readers (McNamara et al., 2014).

Semantic Cohesion. Semantic cohesion is calculated in Coh-Metrix using Latent Semantic Analysis (LSA). Semantic overlap is calculated between the paragraphs in the essay (McNamara et al., 2014). Similarly, WAT uses LSA to calculate verb overlap. This measure of verb cohesion calculates the average LSA cosine between verbs in adjacent sentences. We used two indices of semantic cohesion: LSA overlap amongst all paragraphs (LSA paragraph-to-paragraph) and LSA overlap between verbs. These indices are indicative of the extent to which certain concepts are repeated across sections of essays.

Easability Component Scores. Coh-Metrix calculates five text Easability Components that were developed to account for the multiple dimensions of text difficulty (see Graesser, McNamara, & Kulikowich, 2011, for more information). In this study, we analyzed the components related to text *narrativity, referential cohesion, and deep cohesion,* because they relate to discourse-level text properties and essay quality. Narrativity captures the genre/style of a text by calculating the amount of story-like and familiar elements it contains. Referential cohesion measures how words and ideas are repeated between sentences. Deep cohesion refers to how ideas connect throughout the text (i.e., through causal and logical relationships).

Method

Participants This study included 108 college students from a large university campus in the Southwest United States. These students were, on average, 19.75 years of age (range: 18-37 years), with the majority of students reporting a grade level of college freshman or sophomore. Of the 108 students, 52.9% were male, 53.7% were Caucasian, 22.2% were Hispanic, 10.2% were Asian, 3.7% were African-American, and 9.3 % reported other nationalities. The data for two students were lost due to a computer failure.

Study Procedure This study consisted of a 2-hour session during which students completed the following assessments (in this order): demographics questionnaire, timed-essay, vocabulary test, comprehension test, WM task, and a component processes task.

Essays All students wrote a timed (25-minute), prompt-based, argumentative essay that resembles what they would see on the SAT (previously referred to as the Scholastic Aptitude Test; sat.collegeboard.org). Students were not allowed to proceed until the entire 25 minutes had elapsed.

Working Memory Capacity Students' WM capacity was assessed using the Automated Operation Span (Aospan; Unsworth, Heitz, Schrock, & Engle, 2005). Students'

overall Aospan score reflects the total number of letters they correctly recognized and correctly ordered.

Component Processes Students' inferencing ability was measured using Hannon and Daneman's (2001) component processes task. This test assesses individual differences in higher-level cognitive skills: text memory, text inferencing, knowledge access, and knowledge integration. Previous studies have used this task to identify the processes involved in reading comprehension (Hannon & Daneman, 2001) and writing proficiency (Allen et al., 2014).

Results

Correlation and regression analyses were conducted to examine relations between students' WM scores and the surface- and discourse-level properties of their essays. We first investigate these relations for all students, and then investigate whether and how these relationships differ for students with low and high inferencing skills.

Table 1. Pearson correlations between linguistic indices and AOSPAN scores for all students, low inference students, and high inference students

	Linguistic Index	All Ss	Low Inf	High Inf
Surface-level	Word frequency[1]	**.22***	**.40****	.04
	Word polysemy	**-.23***	**-.34***	-.07
	Word hypernymy	.03	-.07	.06
	Modals[2]	-.18	**-.35***	.05
	Subordinating conj.[3]	**.34****	**.30***	**.40****
	1st person pronouns[2]	.07	.26	-.08
	3rd person pronouns[2]	-.01	-.01	-.04
	Causative subordinators[2]	**.21***	**.32***	.06
	Phrasal coordinators[2]	-.15	**-.28***	-.07
	Word before main verb	**.21***	.24	.17
Discourse-level	Logical connectives[2]	.18	**.34***	.05
	Argument overlap	.03	-.02	.20
	Location nouns[2]	.00	.07	-.08
	Motion verbs[2]	-.18	**-.34***	.03
	Intentional ratio[4]	-.13	**-.33***	.16
	LSA Paragraph overlap	.08	.03	.15
	LSA Verb overlap	.06	.04	.14
	Narrativity[5]	-.01	.09	-.09
	Referential cohesion[5]	.11	.03	.23
	Deep cohesion[5]	.03	-.02	.10

Notes: ** $p < .001$; * $p < .05$; [1]Minimum log word frequency; [2]Incidence; [3]Incidence of prepositions and subordinating conjunctions; [4]Ratio of intentional particles to intentional; [5]Easability Percentile Score

Working Memory and Text Features

Students' scores on the WM (M = 56.44, SD = 11.79) and inferencing (M = 61.16, SD = 15.24) measures were not significantly correlated ($r = .14$, $p = .17$). Therefore, it can be inferred that the tasks measured independent skills. Students' essays contained an average of 410.44 words (SD = 152.50), ranging from a minimum of 84 words to a maximum of 984 words. Table 1 presents the Pearson correlations between scores on the WM test and the surface-level and discourse-level essay properties selected for the NLP analysis. WM scores were primarily related to the surface-level properties of students' essays, such as word frequency and syntactic complexity. However, the discourse-level indices demonstrated much weaker relations with WM, with only two marginally significant correlations demonstrated (i.e., motion verbs, logical connectives).

A stepwise regression was conducted to determine which of these text properties were most predictive of WM scores. The indices that demonstrated significant or moderately significant correlations were regressed onto the WM scores, yielding a significant model [$F(3, 97) = 9.85$, $p < .001$; $R^2 = .23$] with three significant predictors: incidence of prepositions and subordinating conjunctions [$B = .34$, $t(1, 97) = 3.75$, $p < .001$], minimum log word frequency [$B = .32$, $t(1, 97) = 3.45$, $p = .001$], and number of words before the main verb [$B = .23$, $t(1, 97) = 2.47$, $p = .015$]. The results suggest that students with higher WM scores generated texts that contained more complex sentence structures, less familiar words, and a greater incidence of transition words. Importantly, all variables retained in the analysis had been classified a priori as surface-level text properties. Thus, the results provide evidence that the relationship between WM and writing may be strongest at surface levels of the text, such as in the sophistication of the words and the complexity of the sentences.

Role of Inferencing Skills

Our second research question focused on the interplay between WM and inferencing skills during the writing process. To investigate this research question, we conducted a median split on participants' component processes task scores, which resulted in two groups: low (n = 53; M = 17.49, SD = 3.09) and high inference ability students (n = 53; M = 26.55, SD = 3.07). Separate correlation and regression analyses were then conducted on the two groups to determine whether the relations between WM and the linguistic indices were weaker for the high inference ability students than for the low inference ability students.

Low inference ability students Pearson correlations were calculated between the linguistic indices and students' WM scores (see Table 1). Similar to the previous analyses, performance on the WM test was most strongly related to the surface-level linguistic indices. The surface-level linguistic indices that most strongly correlated with WM scores were: minimum log word frequency, incidence of modals and word polysemy. Thus, for low inference ability students, higher WM scores were associated with more frequent words, along with language that was less abstract and more direct (modals, such as *might*, serve as a means through which writers can hedge their arguments).

Notably, fewer (only three) discourse-level indices related to WM scores: motion verbs, ratio of intentional particles to intentional verbs, and logical connectives. Thus, students with higher WM scores produced essays with fewer motion

verbs, less intentional cohesion, and more logical connectives.

To determine which of the indices were most predictive of WM scores, a stepwise regression was calculated with the variables that demonstrated significant or marginally significant correlations. This yielded a significant model [$F(4, 45) = 9.61$, $p < .001$; $R2 = .46$], with four predictors: minimum log word frequency [$B = .40$, $t(1, 45) = 3.63$, $p = .001$], word polysemy [$B = -.32$, $t(1, 97) = -2.84$, $p = .007$], logical connectives [$B = .32$, $t(1, 45) = 2.88$, $p = .006$], and incidence of modals [$B = -.24$, $t(1, 45) = -2.09$, $p = .043$]. These results suggest that, for the low inference ability students, text properties accounted for nearly *half* (i.e., 46%) of the variance in WM scores. Additionally, performance on the WM assessment was more strongly predicted by surface-level text properties than discourse-level text properties.

High inference ability students Similar analyses were conducted for the high inference ability students (see Table 1). As with the low inference ability students, performance on the WM test was more strongly related to surface-level linguistic indices than discourse-level linguistic indices. Unlike the low inference ability students, however, only one of the indices was significantly correlated with WM scores (incidence of prepositions and subordinating conjunctions). A follow-up regression was conducted and yielded a significant model [$F(1, 49) = 9.15$, $p = .004$; $R^2 = .16$], with incidence of prepositions and subordinating conjunctions [$B = .40$, $t(1, 49) = 3.02$, $p = .004$] as a significant predictor.

Taken together, the analyses of the two student groups provide confirmatory evidence for our hypothesis that inferencing skills helped to attenuate the effects of WM capacity on the writing process. Additionally, across the analyses, surface-level text properties provided the most predictive power for WM scores, suggesting that this individual difference predominantly manifests in the surface-level features of texts. These results potentially suggest that when students develop strong inferencing skills, the consequences of WM deficits may be reduced.

Discussion

In this study, we examined the relationship between WM capacity and writing ability through a linguistic analysis of student essays. The results first confirmed the notion that WM capacity is related to the features of texts produced by writers. Namely, students with higher WM scores produced essays that contained more sophisticated vocabulary and complex sentence constructions. The discourse-level properties of the essays, on the other hand, did not vary according to students' WM capacities. Thus, students with higher WM capacities did not necessarily produce essays that were more coherent or informative than their peers. Taken together, these results emphasize the importance of investigating writing at multiple levels of the text. The differential relations between WM and surface- and discourse-level text properties may shed light on the inconsistent findings regarding WM in previous research. In particular, depending on the nature of the rubric, WM may be more or less related to the scores assigned to essays.

Importantly, the results additionally revealed information about the interactive influence of students' lower- and higher-level skills on the writing process. When considering all of the students in the analysis, the linguistic features of the essays accounted for approximately a quarter (23%) of the variance in WM scores, which may suggest that WM potentially serves an important role in the writing process – at least with respect to the surface-level text properties. This relationship differed, however, once inferencing skills were taken into consideration. For low inference ability students, the linguistic properties accounted for nearly half (46%) of the variance in WM scores, whereas they only accounted for 16% of the variance in high inference ability students' WM scores. These results suggest that higher-level inferencing skills can potentially reduce the negative effects of WM constraints during writing.

These results are also important for writing researchers and educators, as they indicate that the link between cognitive skills and the writing process may fluctuate according to the degree to which students have developed strong inferencing skills. Accordingly, writing proficiency is not only influenced by the cognitive capacity of a given student, but is also (and arguably more importantly) closely related to the degree to which that student has developed strategic skills. This finding is particularly important in the context of education, as students can be taught to generate inferences. Although educators have little means to modify a student's WM capacity, they can help students enhance their strategic skills. Thus, higher-level strategy training may be a powerful intervention tool and potentially offset the negative effects of cognitive constraints.

An additional strength of the current study is that it employs NLP techniques to analyze the linguistic properties of the students' essays. Although previous studies have investigated the role of individual differences in the writing process, they have largely relied on human judgments of essay quality or subjective human coding of specific essay elements. Here, we leveraged NLP tools to automatically calculate the surface- and discourse-level features of students' essays. These analyses afforded us the opportunity to investigate the role of WM capacity at a much finer grain size. Thus, rather than simply concluding that WM is an important component in essay quality (according to certain essay rubrics), we can claim that WM is most strongly related to the production of essays that contain sophisticated words and sentences. Overall, these fine grain linguistic analyses can serve as powerful tools for writing researchers, as they can provide more thorough descriptions for the various components of the writing process.

Acknowledgments

This research was supported in part by the Institute for Educational Sciences, US Department of Education, through Grant R305A120707. 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.

References

Allen, L. K., Snow, E. L., Crossley, S. A., Jackson, G. T., & McNamara, D. S. (2014) Reading components and their relation to the writing process. *L'Année Psychologique/Topics in Cognitive Psychology, 114*, 663-691.

Babayigit, S., & Stainthorp, R. (2011). Modeling the relationships between cognitive–linguistic skills and literacy skills: New insights from a transparent orthography. *Journal of Educational Psychology*, *103*(1), 169.

Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writers. In E. C. Butterfield (Ed.), *Children's writing: Toward a process theory of the development of skilled writing* (pp. 57-81). Greenwich, CT: JAL.

Crossley, S. A., Allen, L. K., Kyle, K., & McNamara, D.S. (2014). Analyzing discourse processing using the Simple Natural Language Processing Tool (SiNLP). *Discourse Processes, 51*, 511-534.

Crossley, S. A. & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing, 18,* 119-135.

Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *IJCEELL, 21,* 170-191.

Dixon, P., LeFevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology, 80*, 465-472.

Donovan, C. A., & Smolkin, L. B. (2006). Children's understanding of genre and writing development. *Handbook of Writing Research,* 131-143.

Graesser, A. C., Cai, Z., Louwerse, M., & Daniel, F. (2006). Question Understanding Aid (QUAID): A web facility that helps survey methodologists improve the comprehensibility of questions. *Public Opinion Quarterly*, *70*, 3–22.

Graesser, A. C. & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science, 2*, 371-398.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40,* 223-234.

Graham, S. (2006). Writing. In P. Alexander & P. Winne (Eds.), *Handbook of Educational Psychology* (pp. 457-478). Mahwah, NJ: Erlbaum.

Graham, S., & Perrin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445-476.

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General, 114*, 357-374.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding the individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93,* 103-128.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In Levy, C. M., and Ransdell, S. (eds.), *The science of writing: Theories, methods, individual differences, and applications*, Lawrence Erlbaum Associates, Publishers, Mahwah, NJ, pp. 1-27.

Hoskyn, M., & Swanson, H. L. (2003). The relationship between working memory and writing in younger and older adults. *Reading and Writing*, *16*(8), 759-784.

Kellogg, R. T. (2001). Long-term working memory in text production. *Memory & Cognition*, 29, 43-52.

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research, 1*, 1-26.

Kemper, S., Rash, S., Kynette, D., & Norman, S. (1990). Telling stories: The structure of adults' narratives. *European Journal of Cognitive Psychology*, *2*, 205-228.

Kintsch, W. & van Djik, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363-394.

Longo, B. (1994). The role of metadiscourse in persuasion. *Technical Communication*, *41*, 348– 352.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8*, 299-325.

McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology*, *86*, 256.

McNamara, D. S. (2013). The epistemic stance between the author and the reader: A driving force in the cohesion of text and writing. *Discourse Studies, 15*, 575-592.

McNamara, D. S., Crossley, S. A., & Roscoe, R. D. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods, 45,* 499-515.

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix.* Cambridge: Cambridge University Press.

McNamara, D. S., Jacovina, M. E., & Allen, L. K. (2015). Higher order thinking in comprehension. In P. Afflerbach (Ed.), *Handbook of individual differences in reading: Text and context* (pp. 164-176). Taylor & Francis, Routledge: NY.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29*, 10-17.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498-505.