

Say more and be more coherent: How text elaboration and cohesion can increase writing quality

Scott A. Crossley[°] & Danielle S. McNamara^{*}

[°]Georgia State University - Atlanta, GA | USA

^{*}Arizona State University - Tempe, AZ | USA

Abstract: This study examines links between essay quality and text elaboration and text cohesion. For this study, 35 students wrote two essays (on two different prompts) and for each, were given 15 minutes to elaborate on their original text. An expert in discourse comprehension then modified the original and elaborated essays to increase cohesion, resulting in a 2 (prompt) x 2 (original content, elaborated content) x 2 (original cohesion, improved cohesion) design. Expert raters scored the essays for overall quality and text coherence. In terms of overall essay quality, increasing text content (i.e., elaboration) and improving cohesion both led to significant gains in expert judgments of writing quality, and a combination of both elaboration and improved cohesion led to increased scores over increased cohesion alone. Judgments of text coherence were increased by improved cohesion (but not elaboration); and a combination of both elaboration and improved cohesion led to higher human ratings of coherence in comparison to the original and elaborated versions. The results have important implications for writing theories, writing success, writing pedagogy, and standardized testing.

Keywords: essay quality, cohesion, coherence, elaboration



Crossley, S. A. & McNamara, D. S. (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research*, 7(3), 351-370. doi: 10.17239/jowr-2016.07.03.02

Contact: Scott A. Crossley, Georgia State University, 25 Park Place, Suite 1500, GA 30303, Atlanta | USA - scrossley@gsu.edu

Copyright: Earli | This article is published under Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported license.

1. Introduction

Writing is of critical importance to accomplishments in both academia and in the workplace. Despite its importance, successful writing eludes many students. Hence, one important goal is to provide students with effective instruction that helps them meet expectations for good writing. Doing so necessitates developing an understanding of which aspects of writing are particularly key to writing quality.

One approach to understanding which features of writing are more or less important to writing quality has been to assess the linguistic characteristics of writing using natural language processing tools. Using this approach, a number of studies have shown that expert judgments of writing quality are best predicted by the amount of content in an essay (i.e., the production of more words/text elaboration; Crossley, Roscoe, et al., 2011; Ferrari, Bouffard, & Rainville, 1998; Haswell, 2000; McNamara, Crossley, & McNamara, 2010; McNamara, Crossley, & Roscoe, 2013). Beyond text elaboration, essay quality is also related to the sophistication of the words (e.g., more rare words), greater syntactic complexity, and the greater use of rhetorical features in essays (McNamara et al., 2010; McNamara et al., 2013). Similar studies examining teachers' expectations of writing quality have shown comparable results (Varner, Roscoe, & McNamara, 2013). In essence, better writers produce more words with more sophisticated text structures.

There is also a general assumption that cohesion is important to writing quality (Collins, 1998; DeVillez, 2003). In line with that assumption, expert judgments of text coherence operationalized in terms of text organization and cohesion are the strongest predictors of overall essay quality (Crossley & McNamara, 2010, 2011). However, a number of studies have reported that linguistic properties related to local cohesion (i.e., cohesion between sentence level units; Halliday & Hasan, 1976) are either unrelated or negatively related to essay quality (Crossley & McNamara, 2010, 2011; Crossley, Weston, McLain-Sullivan, & McNamara, 2011). Conversely, linguistic properties related to global cohesion (i.e., cohesion between larger chunks of texts such as paragraphs; Givón, 1995; Kintsch, 1995; Louwerse, 2005) have shown positive relations with essay quality (Crossley, Roscoe, McNamara, & Graesser, 2011; McNamara et al., 2013; McNamara, Crossley, Roscoe, & Dai, 2015). Hence, in terms of understanding essay quality, the results emerging from analyses of text cohesion properties have been inconclusive.

In the current study, our goal is to experimentally manipulate the linguistic features of writing samples by increasing text elaboration (i.e., text length) and text cohesion to examine links between these experimental manipulations and expert judgments of essay quality. We do so by analyzing whether expert raters score essays differently in terms of text coherence and overall essay quality when text content is elaborated, text cohesion is improved, or when both text content is elaborated and cohesion is improved. To experimentally modify texts, we asked undergraduate students to write

persuasive essays and, when finished, asked them to add additional content. An expert in discourse studies then increased the levels of cohesion in both the original essay and the essay with additional content. This approach differs from the majority of studies that investigate text features and essay quality because its method does not rely on natural language processing tools but rather experimental modifications. Such methods should help us better understand which linguistic features of a text lead to greater writing quality and, if such features can be isolated, identify whether developing writers can be provided with instruction and strategies to help them augment the writing success.

1.1 Linguistic Features Related to Writing Quality

Writing is an important component of both academic and professional success (Geiser & Studley, 2001; Powell, 2009), but the process of becoming a successful writer is a long, complex, and arduous undertaking (NAEP, 2011) that requires writers to coordinate a number of cognitive and knowledge skills. These skills include discourse awareness, linguistic abilities, goal setting, sociocultural knowledge, and memory management strategies (Flower & Hayes, 1981; Kellogg & Whiteford, 2009). The importance of writing in everyday activities and the sheer difficulty in becoming a successful writer has led many researchers to investigate differences in the skills exhibited between successful and unsuccessful writers (Applebee, Langer, Jenkins, Mullis, & Foertsch, 1990; Ferrari et al., 1998; McNamara et al., 2010).

Linguistically, more skilled writers have better control over language and know more about language in general and, more specifically, how to use language in written discourse. More skilled writers have stronger syntax, grammar, lexicon, punctuation, and spelling skills (Applebee et al., 1990). For instance, as writers develop, they begin to produce more complex syntactic structures (McCutchen, Covill, Hoyne, & Mildes, 1994) with the trend toward more complex structures extending from the first grade through college (Haswell, 2000; Stewart, 1978). Specifically, writers at the college-freshman level produce more syntactically complex sentences than 9th grade writers (Crossley Weston et al., 2011) and developing writers use longer sentences and longer clauses as a function of time (Haswell, 2000). In general, writers who produce higher scored essays use fewer verb base forms (Crossley, Roscoe, et al., 2011) and produce sentences that contain more words before the main verb phrases (McNamara et al., 2010).

Higher scored essays can also be predicted based on a writer's lexical knowledge. For instance, higher scored essays generally contain more infrequent words (Crossley, Roscoe, et al., 2011; McNamara et al., 2010; McNamara et al., 2013). Developmentally, less skilled middle school students demonstrate lower lexical generation than skilled middle school students (McCutchen, 1986; McCutchen et al., 1994) and more advanced writers, as a function of grade level, write essays that contain longer words (Haswell, 2000), less concrete words, and more infrequent and ambiguous words (Crossley, Weston, et al., 2011). Lastly, essays scored as higher quality generally contain fewer errors in grammar, punctuation, and spelling (Ferrari et

al., 1998). Together, these studies indicate that developing writers and those writers who are judged to be more proficient produce writing that contains greater lexical sophistication and syntactic complexity, while, concomitantly, containing fewer errors.

A number of studies have demonstrated that text generation (i.e., elaboration) is strongly associated with essay quality and is generally the strongest predictor of essay quality (McNamara et al., 2013). These studies generally show that successful writers produce longer texts (Crossley, Roscoe, et al., 2011; Ferrari et al., 1998; Haswell, 2000; McNamara et al., 2010; McNamara et al., 2013). Theoretically, this link is related to the amount of content that an essay contains and, in addition, likely indicates that an increased amount of relevant information aids the reader in understanding the topic or argument of the essay (i.e., more information equals better clarification) and that elaboration on a topic or argument implies that the writer is a more knowledgeable and trustworthy expert on the topic of the essay.

Improved writing skills are also linked to text coherence. Cohesion and coherence are strongly linked with cohesion defined, in principle, as the presence or absence of explicit cues in the text that allow the reader to make connections between the ideas in the text (Halliday & Hasan, 1976). Thus, cohesion is specific to the text. Coherence, on the other hand, is specific to the reader and refers to the understanding that the reader derives from the text (i.e., coherence is in the mind of the reader). Coherence depends on a number of factors, which may include explicit and implicit cohesion cues, but also nonlinguistic factors such as prior knowledge and reading skill (McNamara, Kintsch, Songer, & Kintsch, 1996; O'Reilly & McNamara, 2007). For young writers, explicit cohesion devices that are local in nature are often used to link sections of text together (e.g., referential pronouns and connectives; King & Rentel, 1979). However, around the 8th grade, developing writers begin to use fewer explicit cohesion cues to organize text (McCutchen, 1986; McCutchen & Perfetti, 1982). This trend continues into high school and beyond. For adolescent and adult writers, the use of explicit local cohesion cues is generally associated with less proficient writing. For instance, less proficient college-level writers tend to use a greater repetition of words than more proficient writers (McNamara et al., 2010) and have greater word overlap between sentences (McNamara et al., 2013). Developmentally, Crossley, Weston, et al. (2011) reported that less proficient writers produced texts with a greater repetition of words, greater word overlap between sentences, and greater use of connectives than more proficient writers. However, the results are different when examining human ratings of essay quality based on global cohesion features. As an example, Crossley, Roscoe, et al. (2011) found that two indices of global cohesion (semantic similarity between initial and middle paragraphs, and semantic similarity between initial and final paragraphs) significantly and positively correlated with essay quality. Similar findings have been reported in a number of follow-up studies (McNamara et al., 2013, 2014).

Thus, research suggests that essays scored higher by expert raters contain a greater number of words per text. At the same time, essays scored higher by expert raters contain fewer explicit cohesive devices related to local cohesion, but produce more

implicit links related to global text cohesion (i.e., semantic similarity between paragraphs). However, one limitation of these approaches is that they presume a linear model for predicting experts' judgments of writing quality. Recent analyses have demonstrated that there are a variety of different styles that can be used to produce a high quality essay (Crossley, Roscoe, & McNamara, 2014). These styles include action and depiction style (i.e., the use of verbs and adjectives), academic style (i.e., longer texts that are also more linguistically complex), accessible style (i.e., texts that are more cohesive), and lexical styles (i.e., the use of more infrequent words). Each style can be used to produce a successful essay indicating that expert raters can attend to a variety of different linguistic styles when assigning a quality score and not solely to combinations features such as text length, lexical sophistication, syntactic complexity, or text cohesion.

1.2 Text Cohesion and Text Coherence

Beyond links between essay quality and measurements of text cohesion, researchers are also interested in how human ratings of text coherence can be predicted based on the production of cohesion features. Such research seeks to gain a better understanding of how accurately cohesion features in the text relate to coherence in the mind of expert raters. Importantly, such studies first assess the power text coherence ratings to explain ratings of overall essay quality. These studies report that text coherence is the strongest indicator of overall essay quality when compared to other analytic features such as strength of thesis, strength of argument, strength of conclusion, grammatical accuracy, and other indicators of text quality (Crossley & McNamara, 2010, 2011).

In terms of text features, Crossley and McNamara (2010) examined links between local cohesive devices (e.g., causal cohesion, word overlap, semantic co-reference, spatial cohesion, connectives and logical operators, temporal cohesion, anaphoric resolution) and human judgments of coherence (e.g., overall text coherence and ease of understanding.) and writing quality. They reported that the majority of these local cohesion features demonstrated no correlation with human judgments of text coherence and of those that did yield significant correlations were negative. Thus, the production of more local cohesion features (in this case causal cohesion, anaphoric reference, connectives, and lexical overlap) was an indication that a writing sample was judged to be less coherent by expert raters. A follow up study (Crossley & McNamara, 2011) examined both local and global indices of cohesion. The global cohesion indices focused on lexical and semantic overlap between the prompt, the essay as a whole, and parts of the essay, such as the initial, middle, and final paragraphs. As in the initial study (Crossley & McNamara, 2010), most local cohesion indices failed to demonstrate significant correlations with human judgments of text coherence and those that did were negatively correlated. In contrast, however, many of the global indices were positively correlated with judgments of text coherence. These indices included semantic similarity scores between middle and final paragraphs, initial to middle paragraphs, and initial to final paragraphs. Together, these two studies

provide evidence that text coherence is an important indicator of essay quality and that judgments of coherence on the part of expert raters are not explained by local cohesive devices in the text, but rather through global cohesive devices.

2. The Current Study

The goal of the current study is to test the hypotheses that increasing text length and local and global cohesion will positively affect experts' judgments of that text's overall quality and experts' judgments of text coherence. We predict that increasing text elaboration and local and global cohesion will lead to increased judgments of writing proficiency. We also predict that text elaboration will interact with cohesion such that elaboration will help improve low cohesion essays to a greater extent than high cohesion essays. Lastly, we test the hypothesis that the incidence of local and global cohesion features in a text are associated with expert ratings of essay quality and coherence. We predict that global cohesion features will be more strongly associated with human ratings of essay quality and coherence than local cohesion features.

3. Method

The goal in this study is to examine the effects of elaborated content, improved cohesion, and the combination of elaborated content and improved cohesion on expert judgments of text coherence and expert judgments of overall text quality. Specifically, we collected a corpus of 280 essays from 35 freshman writers. Each writer wrote two persuasive essays, each within a 25-minute time period. For each essay, the students were given 15 minutes to elaborate on the essay. A trained expert then improved the cohesion on both versions of the two essays. Thus, unlike previous studies, we focus on experimental manipulations of text. We first investigate linguistic differences in the conditions in terms of cohesion cues (both local and global), and lexical and syntactic complexity (which are also related to essay quality). We next assess the differences between these conditions in terms of expert ratings of text coherence and essay quality. We lastly examine links between the expert ratings and incidence of local and global cohesion in the essays.

3.1 Study Design

This study uses a 2x2x2 repeated measures design. The first variable is prompt including two prompts, Fitting in and Winning (see Appendix B). The order of presentation of the two essay prompts was counterbalanced across students. The second independent variable is Essay Elaboration (i.e., Original Content vs. Elaborated Content). The third independent variable is Essay Cohesion (i.e., Original Cohesion vs. Improved Cohesion). The dependent variables include the expert ratings for each essay. We also assessed whether text manipulations increased linguistic features in the text by examining difference in the incidence of cohesion, lexical, and syntactic features

among text conditions. We presumed cohesion would increase as a function of manipulation, but we also examined if the text manipulations led to changes in the texts' lexical sophistication, and syntactic complexity.

3.2 Corpus

The essay corpus used in this study comprised 280 essays written by 35 freshman students. The students were enrolled in Introductory Psychology courses at a Midwestern university in the United States. The students received extra credit for their participation. The students included 27 females and 8 male students. The reported ethnic make-up for the students was 31.5% African American, 40.5% Caucasian, and 5.5% who identified themselves as other. The students' average age was 19 years with an age range of 17 to 31. Each of the 35 students composed two original essays on two different essay prompts and revised each essay to elaborate the content. These four essays were then each revised by a discourse expert to increase the cohesion of the essay. Hence, for each student, the corpus included two original essays, two original essays with elaboration (by the student), two original essays with improved cohesion (by an expert), and two essays with elaboration (by the student) and improved cohesion (by an expert).

3.3 Essay Prompts

Two essay prompts were selected from 20 possible prompts. Six expert writing teachers had evaluated these prompts. These experts were asked to rate the appropriateness and quality for each prompt using a scale from 1 (a poor prompt) to 4 (a good prompt) and asked to provide reasons for prompts they rated as poor. In particular, the teachers were asked to consider the degree to which the prompt would be appropriate and comprehensible for a high school student, not require domain specific background knowledge to answer, and induce a variety of ideas. The two prompts that writing teachers rated as highest in appropriateness and quality were used for this study (the scores were 3.45/4 and 3.55/4 for the two prompts). These two prompts are provided in Appendix B. These prompts are similar to those used on the SAT (sat.collegeboard.org) writing subtest. Although SAT prompts are standardized, a number of studies have indicated that the language of a given writing prompt can influence the construction of essays written on that prompt (Huot, 1990). Thus, we included prompt as a potential interaction in our statistical analyses.

3.4 Procedure

Students were informed that they would write two essays on a laptop computer. The first essay assignment and prompt was presented at the top of an open text document on the computer and was not visible to students until all instructions were given. Students were told to develop and write their response to the first prompt using existing knowledge. Students were not allowed to use notes, the Internet, or ask questions about

what to write. Each student was allotted 25 minutes to compose the original response to the prompt. The student was then provided an additional 15 minutes to elaborate on the essays they had just written. Specifically, the students were told to add at least two additional paragraphs of about 4-5 sentences. The paragraphs were to provide additional examples to clarify the main idea in their essay. Students were not told where in the essay to add this information. If students asked where to add the information (e.g., at the end of the essay), they were instructed to add the content anywhere within the essay. The second essay prompt was then presented to the student. The original essay and elaborated essay were collected using the same procedure as the first essay prompt. The order of essay prompts (Fitting in; Winning) was counterbalanced across students. Students were not told beforehand that they would be given extra time to revise the essays they wrote.

3.5 Cohesion Revisions

An expert in discourse comprehension revised each of the four essays composed by the students in order to increase text cohesion at the local and global level. We relied on expert manipulations because cohesion, unlike elaboration, is an advanced strategy that requires specific training. Misspellings were corrected for each of the essays prior to making the revisions so that cohesion revisions accurately reflected the intended meaning of the essay as well as to increase the accuracy of the automated linguistic analyses of the essays. For example, corrected misspellings included incorrect use of homophones (due vs. do), abbreviations (U vs. you), transposition (form vs. from), additional or missed letters (and vs. an or he vs. the), or incorrect word choice (lose vs. lose).

One expert implemented the modifications within the essays and a second experimenter checked the modifications to ensure that the modifications adhered to the following guidelines. First, the cohesion within the essay was increased by adding word overlap across sentences and paragraphs (e.g., linking text segments together by repeating common words already used in the essay). Increasing links between sentences was meant to increase local cohesion while increasing links between paragraphs was meant to increase global cohesion. Second, referents were specified when anaphors were used (e.g., this, that). No other modifications were made to the essays. These changes were made while attempting to maintain the writer's original meaning, mechanics, voice, and word choices.

To assess the modifications made to the essays, we used the Tool for the Automatic Analysis of Cohesion (TAACO; Crossley, Kyle, & McNamara, in press) and Coh-Metrix (McNamara, Graesser, McCarthy, & Cai, 2014). These tools were used to examine if linguistic difference between the essays differed (i.e., did the modifications lead to significant differences). We used two indices of cohesion from TAACO. The first measured local cohesion (i.e., lemma overlap between sentences) and the second index measured global cohesion (i.e., lemma overlap between paragraphs). We included two Coh-Metrix indices related to lexical sophistication (i.e., word frequency) and syntactic

complexity (i.e., number of words before the main verb) that have been associated with writing quality in past studies (e.g., McNamara et al., 2010). We examined textual features related to lexical sophistication and syntactic complexity to ensure that prompt, text elaboration, and the text manipulations did not affect textual features other than cohesion.

We conducted a 2x2x2 repeated measures ANOVA to analyze effects of prompt, elaboration, and cohesion on the linguistic features of the essays. Descriptive statistics for the essay scores are provided in Table 1. For sentence and paragraph lemma overlap, there were no significant differences as a function of prompt. There was a significant effect of elaboration on local and global cohesion with elaborated essays containing more local cohesion and global cohesion than non-elaborated essays. There was also a significant effect for cohesion, with essays manipulated for local and global cohesion containing greater cohesion than essays not manipulated for cohesion. For cohesion indices, there were no significant interactions between elaboration, cohesion, and prompt. For lexical and syntactic features, no significant differences or interactions were reported as a function of prompt, elaboration, or cohesion.

Table 1: Descriptive statistics: Linguistic features for text prompt, elaboration, and cohesion

Feature	Original content mean (SD)		Elaborated content mean (SD)		Improved cohesion mean (SD)		Elaborated + Cohesion mean (SD)	
	Fitting in	Winning	Fitting in	Winning	Fitting in	Winning	Fitting in	Winning
Local cohesion index: Lemma overlap between sentences	0.149 (0.029)	0.138 (0.024)	0.146 (0.023)	0.147 (0.023)	0.161 (0.025)	0.149 (0.020)	0.167 (0.025)	0.162 (0.024)
Global cohesion index: Lemma overlap between paragraphs	0.184 (0.075)	0.174 (0.078)	0.219 (0.065)	0.228 (0.066)	0.202 (0.040)	0.207 (0.054)	0.250 (0.043)	0.245 (0.048)
Lexical sophistication index: Minimum content word frequency by sentence	1.313 (0.225)	1.316 (0.227)	1.254 (0.180)	1.276 (0.173)	1.278 (0.204)	1.271 (0.165)	1.311 (0.241)	1.321 (0.231)
Syntactic complexity index: Number of words before main verb	4.517 (1.622)	4.753 (1.760)	4.416 (1.302)	4.400 (1.252)	4.442 (1.641)	4.439 (1.598)	4.520 (1.574)	4.506 (1.701)

3.6 Essay Evaluation

Twelve raters with at least one year's experience teaching composition classes at a large university rated the 280 essays in the corpus using a holistic grading scale based on a standardized rubric commonly used in assessing SAT essays (see Appendix A for the SAT rubric) and a rubric that assessed individual features of the text including text coherence feature (i.e., Continuity; see Appendix C). The holistic grading scale and the rubric had a minimum score of 1 and a maximum score of 6. The raters were informed that the distance between each score was equal. Accordingly, a score of 5 is as far above a score of 4 as a score of 2 is above a score of 1.

The raters were first trained to use the survey instrument with 20 essays. A Pearson correlation for each essay evaluation was conducted between all possible pairs of raters' responses. If the correlations between all raters did not exceed $r = .70$ (which was significant at $p < .001$) on all items, the ratings were reexamined until scores reached the $r = .70$ threshold. After the raters had reached an inter-rater reliability of at least $r = .70$, each rater was assigned to a group with two other raters (four groups of three raters each). Each group was given a selection of 70 essays from the corpus. The essays were counterbalanced such that each group did not score more than one essay from each writer on each prompt and so that each group scored a similar number of essays from each essay type (original, original with elaboration, original with added cohesion, and original with elaboration and cohesion). The raters were blind to condition as well as to the variables of focus in the study.

Inter-rater reliability among the raters for the holistic essay score was $r = .80$. For the cohesion score (i.e., Continuity), inter-rater reliability among the raters was $r = .59$ (in line with the expectation that agreement on analytic features is more difficult to obtain; Weigle, 2002). Because the correlation was below $.70$, an outside, expert rater adjudicated all essays that had an average difference of 2 or greater among the three raters. After adjudication, the reported inter-rater reliability was $r = .73$.

3.7 Statistical Analysis

We conducted repeated-measures analyses of variance (ANOVAs) to examine the effects of text elaboration and increased text cohesion on judgments of essay quality, and judgments of text cohesion. These ANOVAs were 2x2x2 with the levels including prompt, elaboration, and cohesion. Lastly, we conducted correlations between the expert ratings and the cohesion indices reported by TAACO to assess which indices were stronger predictors of essay quality and text coherence.

4. Results

4.1 Essay Quality Scores

A 2x2x2 repeated measures ANOVA was conducted to analyze effects of prompt, elaboration, and cohesion on the human evaluations of essay quality. Descriptive

statistics for the essay scores are provided in Table 2. There were no significant differences as a function of prompt (all $F < 1$). There was a significant effect of elaboration with elaborated essays scored higher ($M = 2.97$, $SD = 0.72$) than non-elaborated essays ($M = 2.64$, $SD = 0.85$), $F(1, 34) = 24.16$, $p < .001$, $h_p^2 = .42$. There was a significant effect of cohesion, with essays containing added cohesion scored higher ($M = 2.92$, $SD = 0.78$) than essays without added cohesion ($M = 2.69$, $SD = 0.80$), $F(1, 34) = 14.30$, $p < .001$, $h_p^2 = .296$.

Table 2: Descriptive statistics: Human evaluation for text prompt, elaboration, and cohesion

Human evaluation	Original content mean (SD)		Elaborated content mean (SD)		Improved cohesion mean (SD)		Elaborated + Cohesion mean (SD)	
	Fitting in	Winning	Fitting in	Winning	Fitting in	Winning	Fitting in	Winning
Essay Scores (n = 35)	2.467 (0.751)	2.476 (0.887)	2.867 (0.687)	2.952 (0.868)	2.791 (0.901)	2.829 (0.88)	3.048 (0.691)	3.000 (0.642)
Cohesion scores (n = 35)	3.476 (0.634)	3.457 (0.983)	3.705 (0.803)	3.705 (0.651)	3.761 (0.807)	3.943 (0.773)	3.905 (0.629)	3.857 (0.585)

There was also a significant interaction between elaboration and cohesion, $F(1, 34) = 8.52$, $p < .010$, $h_p^2 = .20$. Essentially, both elaboration and cohesion equally increased the perceived quality of the essays, with a combined benefit for increasing both. This interpretation is substantiated by significant differences in scores between the original essays ($M = 2.47$, $SD = 0.82$) and original essays with elaboration ($M = 2.91$, $SD = 0.78$; $F(1, 34) = 36.04$, $p < .001$, $h_p^2 = .52$) and between the original essays and original essays with added cohesion ($M = 2.81$, $SD = 0.89$; $F(1, 34) = 19.21$, $p < .001$, $h_p^2 = .36$). The essays with both elaboration and cohesion ($M = 3.02$, $SD = 0.67$) were perceived as higher quality than were the original versions with added cohesion ($M = 2.80$, $SD = 0.89$), $F(1, 34) = 7.15$, $p < .050$, $h_p^2 = .17$, but not of higher quality than the elaborated essays without added cohesion ($M = 2.91$, $SD = 0.78$; $F(1, 34) = 3.15$, $p > .050$, $h_p^2 = .09$). There were no significant interactions involving prompt (all $F < 1$).

4.2 Essay Coherence Scores

A 2x2x2 repeated measures ANOVA was conducted to analyze effects of prompt, elaboration, and cohesion on the human evaluations of essay coherence (i.e., Continuity; see Appendix C). Descriptive statistics for the coherence scores are provided in Table 1. There were no significant differences as a function of prompt (all $F < 1$). There was no significant effect of elaboration with elaborated essays scored no higher on coherence ($M = 3.79$, $SD = 0.67$) than non-elaborated essays ($M = 3.66$, $SD = 0.77$), $F(1, 34) = 3.75$, $p > .050$, $h_p^2 = .10$ (although the finding did approach significance, $p = .06$). There was a significant effect of cohesion, with essays containing

improved cohesion scored higher for coherence ($M = 3.87$, $SD = 0.70$) than essays without added cohesion ($M = 3.59$, $SD = 0.75$), $F(1, 34) = 16.90$, $p < .001$, $h_p^2 = .33$.

There was a significant interaction between elaboration and cohesion, $F(1, 34) = 4.41$, $p < .050$, $h_p^2 = .12$, such that essays with both elaboration and cohesion ($M = 3.88$, $SD = 0.61$) were perceived as containing greater coherence than were the original versions ($M = 3.47$, $SD = 0.76$), $F(1, 34) = 20.96$, $p < .001$, $h_p^2 = .38$, and the original versions with elaboration ($M = 3.71$, $SD = 0.73$), $F(1, 34) = 4.16$, $p < .050$, $h_p^2 = .11$. No significant differences in coherence scores were reported between essays with both elaboration and cohesion and cohesion alone ($M = 3.85$, $SD = 0.79$), $F(1, 34) = 0.10$, $p > .050$, $h_p^2 = .003$, nor were there significant interactions involving prompt (all $F < 1$).

4.3 Correlations between Indices and Expert Ratings

Pearson product moment correlations between the local and global indices of cohesion reported by TAACO and the expert ratings of essay quality and text coherence demonstrated that the global cohesion index (i.e., lemma overlap between paragraphs) was the stronger predictor of the human judgments (see Table 3), with a significant medium effect size. The local cohesion index reported small effect sizes that were significant for the coherence score but not the holistic score.

Table 3: Correlations (r) between human judgments and cohesion indices

Index	Holistic score	Coherence score
Lemma overlap between sentences	.115	.123*
Lemma overlap between paragraphs	.489**	.469**

* $p < .050$, ** $p < .001$

5. Discussion

Writing is a difficult task and one that takes practice, training, and time. Indeed, these features are likely the reasons that writing is a critical part of academic success and success in the workplace. In light of these ideas, it is important to understand what linguistic elements lead to successful writing and how these elements might lead to improvements in writing instruction and the use of writing strategies. This study addresses two key components of written communication: text elaboration and cohesion. The study examines how these components individually and collectively can predict human ratings of essay quality and text coherence. The study also addresses how the manipulation of these components can affect text structure and linked text features to human ratings of essay quality. In tandem, these analyses can inform theories of cohesion and coherence and their effects on discourse processing. In addition, the results have important implications for writing theories, writing success, writing pedagogy, and standardized testing.

In terms of overall essay quality, both increasing text content (i.e., elaboration) and improving cohesion led to significant gains in human judgments of writing quality. Neither elaboration nor increasing cohesion led to gains over and above one another, but a combination of both elaboration and increased cohesion led to increased gains in essay scores over increased cohesion alone. The interaction was such that writing gains were greater when lower cohesion texts were elaborated as compared to when high cohesion texts were elaborated. For human judgments of text coherence, elaboration did not lead to significant differences in the human ratings, but increasing cohesion did. Both increasing cohesion and elaborating on content also led to higher ratings of coherence over the original and elaborated versions. No interactions were reported for high cohesion texts and elaboration (i.e., a high cohesion text was scored similarly coherent whether it was elaborated or not). However, low cohesion texts benefited from elaboration and were scored as more coherent when elaborated. The follow up linguistic analysis determined that the manipulations created by the students and by the researcher led to linguistic differences in the texts such that elaborating on content and improving cohesion created texts that contained greater local and global cohesion. However, such changes did not concomitantly lead to increases in lexical sophistication or syntactic complexity.

In sum, then, asking students to elaborate on ideas in an essay leads to both increased judgments of essay quality and increased local and global cohesion within the essays. Notably, quality and cohesion increased solely as a result of students adding content, following fairly simple instructions. This finding indicates that writing quality does not emerge from individual differences alone. If individuals' skills were the principal factor driving the scores, then simply giving writers an extra 15 minutes to elaborate on their writing would not be effective. On the contrary, successful writing, to some extent, depends on more than just individual skills. It relies on the student knowing or understanding what is required or expected in the task (Varner et al., 2013). In this case, informing the student to add more content went a long way in helping the student to improve the essays. These results have a simple implication for writing pedagogy: students should be given the opportunity or be required to revise an essay prior to submission. General guidance that prompts the student to write at least an additional two paragraphs that provide additional examples to clarify the main idea seems to be sufficient instruction to significantly increase essay scores as well as the local and global cohesion of an essay. While the notion that asking students to revise essays can increase essay quality is not revolutionary, the findings from this study provide an important reminder of the effect revision has on essay quality. In addition, the findings quantify, to a degree, the amount text and time that statistically increase writing quality. Also, the findings from this study provide evidence that provide time and guidelines for revision cannot only increase human judgments of essay quality, but also human judgments of text coherence.

Having an expert improve the cohesion of the essays also increased local and global cohesion (as expected), and in turn led to gains in judgments of quality and

coherence. In terms of writing pedagogy, this finding indicates that writers will gain from instruction on cohesion, and specifically on how to increase word overlap across text segments and specifying unclear anaphors. These skills are relatively simple and should allow for successful classroom interventions based solely on asking students to repeat key words and phrases across paragraphs and specifying anaphoric reference (e.g., changing *This is important* where *this* refers to *showering* to *Showering is important*). The inclusion of such writing strategies in student instruction should further increase the quality and the coherence of the essay.

From a theoretical perspective, the findings from this study support the notion that increasing both local and global cohesion leads to increased judgments of writing proficiency. Like previous research, this study finds stronger links between human scores of essay quality and text coherence for a global index of text cohesion than a local index, indicating that global cohesion is a more important indicator of text quality than local cohesion (Crossley & McNamara, 2011; Crossley, Roscoe, et al., 2011; McNamara et al., 2013; McNamara et al., 2014). However, our local cohesion index was positively correlated with holistic scores and coherence scores and, in the case of the latter, demonstrated a significant correlation (albeit small). This is counter to several previous studies (Crossley & McNamara, 2010, 2011; Crossley, Weston, et al., 2011) and is likely the result of expert manipulations of cohesion as compared to individual differences in the use of cohesion. One can assume that the use of cohesive cues in writing depends on a host of other factors, which in this study were controlled because the essays were written by the same individual.

6. Conclusion

This study has demonstrated that both elaborating on essay content and increasing the cohesion of an essay lead to gains in human judgments of essay quality and coherence. Since both techniques are relatively simple, the techniques can be taught as strategies to student writers to potentially increase expert ratings of writing proficiency and, hopefully, academic success. The findings also provide researchers with a better understanding of how textual features related to local and global cohesion help develop coherence for expert raters.

Future studies should consider similar methodologies, but focus on non-expert ratings. The effects of local and global cohesion devices on the development of coherent mental representations of text for non-experts would help distinguish the types of textual features that lead to more readable and comprehensible texts. Future research should also consider the practical effects of the pedagogical strategies suggested by this study on a number of different populations such as middle and high-school students. Specifically, it remains to be seen if specific strategies related to cohesion are useful to younger writers. Such studies will help cement the findings found here and extend the results to a broader range of populations. The end results, hopefully, will be improved writing proficiency at both the academic and professional levels.

Acknowledgments

This research was supported in part by the Institute for Education Sciences (IES R305A080589 and IES R305G20018-02). Ideas expressed in this material are those of the authors and do not necessarily reflect the views of the IES. We thank various members of the SoLET lab at Arizona State University (Jianmin Dai and Laura Varner specifically) and members of the Institute for Intelligent Systems at The University of Memphis (Carl Cai and Art Graesser specifically). Most importantly we thank Mary Jane White for early work on this project. Without her, this article would have never been possible.

References

- Applebee, A. N., Langer, J. A., Jenkins, L. B., Mullis, I., & Foertsch, M. A. (1990). *Learning to write in our nation's schools*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Collins, J. L. (1998). *Strategies for struggling writers*. New York, NY: The Guilford Press.
- Crossley, S. A. & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 984-989). Austin, TX: Cognitive Science Society.
- Crossley, S. A., & McNamara, D. S. (2011). Text coherence and judgments of essay quality: Models of quality and coherence. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society*. (pp. 1236-1241). Austin, TX: Cognitive Science Society.
- Crossley, S. A., Roscoe, R., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication, 31* (2), 184-215. doi: 10.1177/0741088314526354
- Crossley, S. A., Roscoe, R. D., McNamara, D. S., & Graesser, A. (2011) Predicting human scores of essay quality using computational indices of linguistic and textual features. In G. Biswas, S. Bull, J. Kay, and A. Mitrovic (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education*. (pp. 438-440). New York: Springer. doi: 10.1007/978-3-642-21869-9_62
- Crossley, S. A., Weston, J., McLain-Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication, 28* (3), 282-311. doi: 10.1177/0741088311410188
- DeViliez, R. (2003). *Writing: Step by step*. Dubuque, IO: Kendall Hunt.
- Ferrari, M., Bouffard, T., & Rainville, L. (1998). What makes a good writer? Differences in good and poor writers' self-regulation of writing. *Instructional Science, 26*, 473-488. doi:10.1023/A:1003202412203
- Flower, L. S., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31-50). Hillsdale, NJ: Lawrence Erlbaum.
- Geiser, S., & Studley, R. (2001). UC and SAT: *Predictive validity and differential impact of the SAT I and SAT II at the University of California*. Oakland: University of California. doi: 10.4324/9780203463932_uc_and_the_sat
- Givón, T. (1995). Coherence in the text and coherence in the mind. In M.A. Gernsbacher & T. Givón, *Coherence in spontaneous text* (pp.59-115). Amsterdam: Benjamins.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Haswell, R. H. (2000). Documenting improvement in college writing: A longitudinal approach. *Written Communication, 17*, 307-352. doi: 10.1177/0741088300017003001
- Kellogg, R., & Whiteford, A. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist, 44*, 250-266. doi: 10.1080/00461520903213600

- King, M., & Rentel, V. (1979). Toward a theory of early writing development. *Research in the Teaching of English, 13*, 243-253
- Kintsch, W. (1995). How readers construct situation models for stories: the role of syntactic cues and causal inferences. In M.A. Gernsbacher & T. Givón, *Coherence in spontaneous text* (pp.139-160). Amsterdam: Benjamins. doi: 10.1075/tsl.31.06kin
- Louwerse, M. M. (2005). Un modelo conciso de cohesión en el texto y coherencia en la comprensión [A concise model of cohesion in text and coherence in comprehension]. *Revista Signos, 37*, 41-58.
- McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. *Journal of Memory and Language, 25*, 431-444. doi: 10.1016/0749-596X(86)90036-7
- McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology, 86*, 256-266. doi:10.1037/0022-0663.86.2.256
- McCutchen, D., & Perfetti, C. (1982). Coherence and connectedness in the development of discourse production. *Text, 2*, 113-139. doi: 10.1515/text.1.1982.2.1-3.113
- McNamara, D. S. (2001). Reading both high and low coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55*, 51-62. doi: 10.1037/h0087352
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). The linguistic features of quality writing. *Written Communication, 27* (1), 57-86. doi: 10.1177/0741088309351547
- McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural Language Processing in an Intelligent Writing Strategy Tutoring System. *Behavior Research Methods, 45* (2), 499-515. doi: 10.3758/s13428-012-0258-1
- McNamara, D. S., Crossley, S. A., Roscoe, R., Allen, L., Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing, 23* (1), 35-59. doi: 10.1016/j.asw.2014.09.002
- 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., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43. doi: 10.1207/s1532690xci1401_1
- National Assessment of Educational Progress. (2011). The nation's report card: Writing 2011. Retrieved from <http://nces.ed.gov/nationsreportcard/writing/>
- O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43*, 121-152. doi: 10.1080/01638530709336895
- Powell, P. (2009). Retention and writing instruction: Implications for access and pedagogy. *College Composition and Communication, 60*, 664-682.
- Varner (Allen), L. K., Roscoe, R. D., & McNamara, D. S. (2013). Evaluative misalignment of 10th-grade student and teacher criteria for essay quality: An automated textual analysis. *Journal of Writing Research, 5*, 35-59. doi: 10.17239/jowr-2013.05.01.2

Appendix A: Holistic Scoring for Essays

SCORE OF 6

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

SCORE OF 5

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

SCORE OF 4

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

SCORE OF 3

An essay in this category demonstrates developing mastery, and is marked by ONE OR MORE of the following weaknesses: develops a point of view on the issue, demonstrating some critical thinking, but may do so inconsistently or use inadequate examples, reasons, or other evidence to support its position is limited in its organization or focus, or may demonstrate some lapses in coherence or progression of ideas displays developing facility in the use of language, but sometimes uses weak vocabulary or inappropriate word choice lacks variety or demonstrates problems in sentence structure contains an accumulation of errors in grammar, usage, and mechanics.

SCORE OF 2

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

SCORE OF 1

An essay in this category demonstrates very little or no mastery, and is severely flawed by ONE OR MORE of the following weaknesses: develops no viable point of view on the issue, or provides

little or no evidence to support its position is disorganized or unfocused, resulting in a disjointed or incoherent essay displays fundamental errors in vocabulary demonstrates severe flaws in sentence structure contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning.

Appendix B: Essay Prompts Provided to Students

Fitting In Prompt

Think carefully about the issue presented in the following excerpt and the assignment below.
From the time people are very young, they are urged to get along with others, to try to "fit in." Indeed, people are often rewarded for being agreeable and obedient. But this approach is misguided because it promotes uniformity instead of encouraging people to be unique and different. Differences among people give each of us greater perspective and allow us to make better judgments.

Is it more valuable for people to fit in than to be unique and different?

Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations.

Winning Prompt

Think carefully about the issue presented in the following excerpt and the assignment below.

From talent contests to the Olympics to the Nobel and Pulitzer prizes, we constantly seek to reward those who are "number one." This emphasis on recognizing the winner creates the impression that other competitors, despite working hard and well, have lost. In many cases, however, the difference between the winner and the losers is slight. The wrong person may even be selected as the winner. Awards and prizes merely distract us from valuable qualities possessed by others besides the winners.

Do people place too much emphasis on winning?

Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observations.

Appendix C: Analytical Rating Form

Read each essay carefully and then assign a score on each of the points below. For the following evaluations you will need to use a grading scale between 1 (minimum) and 6 (maximum). A grade of 1 would relate to not meeting the criterion in any way, and a grade of 4 would relate to somewhat meeting the criterion. The distance between each grade (e.g., 1-2, 3-4, 4-5) should be considered equal. Thus, a grade of 5 (*meets the criterion*) is as far above a grade of 4 (*somewhat meets the criterion*) as a grade of 2 (*does not meet the criterion*) is above a grade of 1 (*does not meet the criterion in any way*).

- 1 Does not meet the criterion in any way
- 2 Does not meet the criterion
- 3 Almost meets the criterion but not quite
- 4 Meets the criterion but only just
- 5 Meets the criterion
- 6 Meets the criterion in every way

Structure

The essay contains a clear division into introduction (one paragraph), argumentation (more than one paragraph) and conclusion (one paragraph).

Continuity

The essay exhibits coherence throughout the essay by connecting ideas and themes within and between paragraphs.

Introduction

The essay contains a clear introductory sentence.

Thesis Statement

The thesis statement is presented in the introduction and briefly states the argument and the position the writer intends to take. (quality matters here).

Reader Orientation

The essay is easy to understand and is coherent overall.

Topic Sentences

Each argumentative paragraph (excluding the introduction and the conclusion) has an identifiable topic sentence. (makes a claim, 1st sentence, tends to be shorter than other sentences).

Evidential Sentences

Each argumentative paragraph (excluding the introduction and the conclusion) contains evidential sentences that support the topic sentence or the point of the paragraph.

Relevance

The argumentation only contains relevant information (i.e., information that helps support the writer's thesis).

Appropriate Registers

The vocabulary used in the essay follows the expectations for the register.

Grammar, Spelling, and Punctuation

The essay demonstrates good use of grammar, spelling, and punctuation.

Conclusion

The essay contains a clear conclusion.

Conclusion Type

The conclusion follows one of the seven identifiable styles. (Anecdote, question, further research, recommendation, speculation, importance, restatement of thesis).

Conclusion Summary

The conclusion summarizes the arguments and the thesis found in the paper.

Closing

It is clear that the essay is finished, for example by a closing statement. There are no loose ends left.