The bit in the middle and why it's important: a computational analysis of the linguistic features of body paragraphs

John C. Myers · Philip M. McCarthy · Nicholas D. Duran · Danielle S. McNamara

Abstract This study examines the relationship between the linguistic characteristics of body paragraphs of student essays and the total number of paragraphs in the essays. Results indicate a significant relationship between the total number of paragraphs and a variety of linguistic characteristics known to affect student essay scores. These linguistic characteristics (e.g., semantic overlap, syntactic complexity) contribute to two underlying factors (i.e., textual cohesion and difficulty) that are used as dependent variables in mixed-effect models. Results suggest that student essays with 5-8 paragraphs tend to be more linguistically consistent than student essays with 3, 4, and 9 paragraphs. Essays with totals of 5-8 paragraphs, considered by many educators to contain an optimal number of paragraphs, may include functionally and structurally similar paragraphs. These findings could aid writing researchers and educators in obtaining a clearer view of the relationship between the total number of paragraphs comprising an essay and the linguistic characteristics that affect essay evaluation. Consequently, writing interventions may become better equipped to pinpoint student difficulties and facilitate student writing skills by providing more detailed and informed feedback.

J. C. Myers · N. D. Duran · D. S. McNamara
Department of Psychology, University of Memphis,
Memphis, USA

P. M. McCarthy
Department of English, University of Memphis,
Memphis, USA

D. S. McNamara (✉)
FedEx Institute of Technology, University of Memphis,
Room 410, 400 Innovation Drive,
Memphis, TN 38152, USA
e-mail: dsmcnamara1@gmail.com

An essay is not merely a concatenation of paragraphs. Each paragraph in an essay serves a purpose, or a rhetorical function. Thus, the purpose of the essay is likely to be conveyed to the reader only when the appropriate kinds of paragraphs are used in a meaningful order (Meyer & Freedle, 1984). The function of paragraphs in essays is analogous to the function of sentences in paragraphs. The topic sentence of a paragraph, typically the first sentence, establishes the theme of the paragraph. The sentences immediately following the topic sentence support, expand, and elaborate the theme. A final warrant sentence offers the relevance, importance, or significance of the issues discussed within the paragraph (McCarthy et al., 2008; Toulmin, 1969). This three-part division is similar to the function of paragraphs in a typical student essay, particularly the five-paragraph essay: the Introduction, the Body Paragraphs, and the Conclusion (Smith, 2006; Wiley, 2000). The first paragraph, the introduction, has the rhetorical function of providing readers with (usually) three ideas to be discussed in the essay. The following three paragraphs, comprising the body of the essay, serve the rhetorical function of supporting, explaining, and elaborating on one of the ideas presented in the introduction. The final paragraph, the conclusion, has the rhetorical function of restating and emphasizing the importance of the ideas presented in the preceding paragraphs (College Board, 2009; Nunnally, 1991; Smith, 2006; Wesley, 2000).

The rhetorical functions of introductions, body paragraphs, and conclusions, as described by the five-paragraph theme, have been adopted by other conceptions of essay structure. That is, students are taught the rhetorical

functions of paragraphs that operate within the five-paragraph theme, even when they are not taught to use the five-paragraph essay theme itself (see Wesley, 2000). The act of teaching students these rhetorical functions rests on the assumption that each body paragraph serves the same rhetorical function (College Board, 2009; Essay Info, 2009), and it disregards potential variance in the total number of paragraphs of the essays.

According to College Board, only 8% of the student essays of the SAT writing assessment test actually conformed to the five-paragraph theme. This frequent divergence from the five-paragraph theme suggests variance in the total number of paragraphs that students produce in response to essay prompts (College Board, 2009). Such variation in the total number of paragraphs of student essays may be related to variation in the rhetorical function (and therefore the linguistic characteristics) of the body paragraphs. For example, one student may write only three paragraphs in response to a prompt, while another may write five or six paragraphs. If both students in this hypothetical situation were responding to the same prompt, then their essays would be representative of their individual approaches to addressing the prompt, and the respective body paragraphs that they would produce may feature differing linguistic characteristics that are dependent upon those approaches.

This study examines the relationship between the total number of paragraphs in an essay and the linguistic characteristics of its body paragraphs. In approaching this question, we formed two contrasting hypotheses. The first hypothesis is that a greater total number of paragraphs is indicative of an essay in which ideas are more rhetorically explicit within each paragraph, because the use of more paragraphs implies more sophisticated explication and knowledge. These body paragraphs are likely written by students who are better equipped to address essay prompts thoroughly, and the paragraphs are more likely to demonstrate characteristics such as cohesion and elaboration (qualities that are shown to significantly improve essay scores) (Crossley, White, McCarthy, & McNamara, 2009). The contrasting hypothesis is that a greater total number of paragraphs is indicative of writers who are conceptually diverse and tend to express multiple ideas without elaborating (i.e., covering more ideas, but in less depth), resulting in paragraphs that are lower in cohesion and higher in characteristics of text difficulty. McNamara, Crossley, and McCarthy (2009) found that such essays tend to be graded significantly higher on scales of writing proficiency. In both cases, the presence of more paragraphs in student essays tends to increase the likelihood that they receive higher grades. This study, however, does not focus on essay quality, but rather on providing a detailed

observation of the linguistic features that characterize the body (middle paragraphs) based on the total number of paragraphs in the essays.

Linguistic characteristics

When examining the linguistic characteristics of text, cohesion often plays a very important role in distinguishing texts from one another. Cohesion refers to the extent to which ideas in a text are explicitly connected. Cohesion can be derived from a variety of sources, including word overlap (e.g., *noun overlap*, *content word overlap*), causal relations within text (e.g., *ratio of causal particles to causal verbs*), connective usage (e.g., *because*, *eventually*), and lexical diversity measures (low lexical diversity indicating higher lexical repetition; see McCarthy & Jarvis, 2010). Texts with higher cohesion tend to be more easily comprehended than texts with lower cohesion, as the relationships between ideas in cohesive texts are more explicit. These explicit connections between ideas allow readers to make fewer inferences in their efforts to understand a text, emphasizing the importance of expressing ideas coherently (e.g., McNamara, 2001; McNamara, Louwerse, McCarthy, & Graesser, 2010; Ozuru, Dempsey, & McNamara, 2009).

Research on reading comprehension has also shown that the presence of textual cohesion (or lack thereof) is a good indicator of the difficulty of a text (see Duran, Bellissens, Taylor, & McNamara, 2007). McNamara and colleagues claim that the readability formulae (e.g., *Flesch-Kincaid Grade Level*) that are most commonly used to select appropriate textbooks for students are not optimal due to their incorporation of only surface-structure textual characteristics (e.g., *number of syllables per word*). On the other hand, cohesion and other deep-structure characteristics (e.g., *word frequency measures*, *semantic features*, *word familiarity*) significantly correlate with ratings of text coherence and difficulty (Duran et al., 2007; Graesser & McNamara, in press).

Coh-metrix

Coh-Metrix is a computational tool that measures both the surface-structure and deep-structure linguistic characteristics of words, sentences, and discourse. This assessment is achieved by combining several linguistic databases, a syntactic parser, and a broad range of state-of-the-art textual analysis approaches. For instance, the MRC database provides psycholinguistic information about word meaningfulness and familiarity (Coltheart, 1981), and Latent Semantic Analysis (LSA) uses a statistical representation of world knowledge based on corpus analysis to calculate the

semantic similarities between units of texts (Landauer, McNamara, Dennis, & Kintsch, 2007). Graesser and colleagues provide an extensive overview of many of the linguistic indices supported by Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004; McNamara et al., 2010).

Numerous studies have shown that Coh-Metrix indices are powerful enough to detect even subtle differences in text and discourse, and many of these studies use Coh-Metrix to distinguish between different types of texts. For instance, Louwerse, McCarthy, McNamara, and Graesser (2004) used Coh-Metrix to find significant differences between spoken and written English. McCarthy, Lewis, Dufty, and McNamara (2006) showed that Coh-Metrix indices can successfully predict original authorship, even while considering significant shifts in authors' writing styles over time. And, Lightman, McCarthy, Dufty, and McNamara (2007) were able to distinguish between the *beginnings*, *middles*, and *ends* of chapters in a corpus of expository text books for high school students. Studies such as these indicate that Coh-Metrix is a valuable text analysis tool capable of analyzing and differentiating a variety of text types.

The current study includes a subset of 13 Coh-Metrix indices (see Table 1) that have been shown to effectively represent the different levels of textual and semantic cohesion and difficulty (see McCarthy, Guess, & McNamara, 2009; McNamara et al., 2009, 2010). The first five Coh-

Metrix indices in Table 1 are indicators of cohesion. *Noun Overlap* measures the repetition of nouns across consecutive sentences; more cohesive texts tend to repeat the same nouns across sentences (McNamara et al., 2010). MED (*Minimal Edit Distance)* is an approach to measuring differences in the sentential positioning of content words. This measure produces values in the opposite direction of most measurements of cohesion, because a high value for MED indicates that content words are located in different places within sentences across the text (e.g., "*Elizabeth* is the queen of England." vs. "This castle belongs to *Elizabeth*."), suggesting lower structural cohesion (see McCarthy et al., 2009). The *Measure of Textual Lexical Diversity* (MTLD: McCarthy & Jarvis, 2010) evaluates the degree to which a text varies in terms of lexical deployment. That is, texts that use many different words (with little repetition) receive higher MTLD values than texts with greater repetition, and texts that have lower lexical diversity (those that use the same words throughout the text) tend to be more cohesive (McNamara et al., 2010). As previously mentioned, LSA uses a statistical representation of world knowledge to measure semantic similarities between units of texts (Landauer et al., 2007). The incidence of causal connectives (e.g., *so*, *because*) reflects the degree to which ideas are connected in the text using such causal connectives. Because understanding the causal relationships between objects within a text is

Table 1 Descriptive statistics of middle paragraphs as a function of the total number of paragraphs in the essay

Variable name	3–Paragraph essays	4–Paragraph essays	5–Paragraph essays	6–Paragraph essays	7–Paragraph essays	8–Paragraph essays	9–Paragraph Essays
Noun overlap	0.34 (0.19)	0.45 (0.28)	0.42 (0.23)	0.46 (0.27)	0.49 (0.29)	0.57 (0.34)	0.64 (0.35)
MED content words	0.91 (0.05)	0.89 (0.06)	0.87 (0.07)	0.87 (0.08)	0.86 (0.12)	0.868 (0.05)	0.83 (0.16)
MTLD	84.26 (23.09)	81.77 (23.80)	76.22 (20.72)	76.75 (24.55)	77.63 (27.86)	70.82 (19.64)	65.18 (23.38)
Latent semantic analysis	0.18 (0.06)	0.29 (0.10)	0.22 (0.09)	0.22 (0.11)	0.25 (0.11)	0.24 (0.10)	0.24 (0.13)
Causal connectives	33.39 (21.10)	33.87 (20.12)	35.74 (17.53)	39.99 (19.46)	37.06 (18.08)	42.03 (21.02)	41.23 (22.32)
Causal verbs	30.81 (16.12)	26.02 (19.00)	26.07 (14.15)	26.17 (13.02)	25.28 (14.26)	33.14 (16.29)	36.33 (14.78)
Words before main verb	10.76 (5.36)	14.77 (7.02)	12.02 (6.03)	11.82 (6.17)	10.52 (6.10)	8.40 (4.13)	10.50 (6.82)
Word concreteness	307.53 (13.69)	300.53 (16.56)	302.38 (15.26)	302.09 (15.07)	300.75 (11.59)	300.71 (11.19)	293.53 (12.54)
Word concreteness min.	200.16 (29.34)	189.73 (22.86)	189.90 (19.57)	192.87 (19.10)	189.43 (22.06)	201.93 (18.60)	207.81 (21.36)
CELEX word frequency	0.33 (0.12)	0.30 (0.10)	0.30 (0.12)	0.31 (0.12)	0.26 (0.11)	0.34 (0.17)	0.26 (0.14)
MRC word familiarity min.	224.40 (267.69)	160.96 (243.81)	226.83 (264.48)	170.66 (248.22)	162.06 (249.31)	213.56 (266.59)	125.92 (228.68)
Flesch-Kincaid grade level	9.16 (2.20)	9.97 (1.89)	9.41 (1.80)	9.77 (1.73)	9.94 (1.81)	9.50 (1.77)	9.79 (1.82)
Average sentence length	17.91 (4.62)	20.36 (4.75)	19.44 (6.41)	20.00 (5.56)	20.093 (5.01)	18.46 (5.31)	18.47 (4.03)
Number of words	133.00 (65.33)	165.76 (92.13)	147.30 (58.37)	132.98 (52.39)	113.13 (44.79)	97.23 (25.79)	77.19 (38.14)
Number of sentences	7.66 (3.70)	8.31 (4.79)	7.88 (3.35)	6.92 (2.78)	5.91 (2.53)	5.60 (1.94)	4.14 (1.75)

Note: Standard deviation in parentheses

integral for comprehension, a higher incidence of causal connectives suggests a repetition of causal relationships and serves as a measure of cohesion.

The next eight Coh-Metrix indices in Table 1 are indicators of text difficulty, while the last two are measures of text length (i.e., number of words and number of sentences). As opposed to the incidence of causal connectives, the incidence of causal *verbs* is inversely related to text difficulty (Graesser, McNamara, & Kulikowich, 2010). Causal verbs (e.g., *pour, break*) represent state changes in text (e.g., from intact to broken), and they can be associated with narrativity (i.e., the extent to which a text expresses events rather than pure information), which is easier to process than informative text (Graesser & McNamara, in press; Haberlandt & Graesser, 1985). Among the indicators of text difficulty, *Maximum Words before Main Verb* is a measure of syntactic complexity. Typically, basic sentences in English express one idea and consist of a subject, followed by a verb, followed by an object (e.g., "The dog ate the bone."). More complex sentences (e.g., "The dog that we saw in the park yesterday ate the bone.") tend to contain more words before the mention of the main verb, increasing working memory load and, consequently, the difficulty of the text to be comprehended (Just & Carpenter, 1992). *Word Concreteness* measures the extent to which the meanings of the words in a text are clear and objective (McNamara et al., 2010). For instance, the word *disk* is more concrete than the word *pleasure*. Texts that feature fewer concrete terms than abstract ones tend to receive higher values in terms of difficulty. *Word Concreteness Minimum* is also a measure of lexical concreteness within a text. However, this measures the minimum lexical concreteness within each sentence.

Together, these indices provide a more comprehensive assessment of word concreteness within the texts, without being highly inter-correlated. *CELEX (Content) Word Frequency* calculates the likelihood of occurrence for content words (e.g., *table* = high frequency; *cognition* = low frequency) within the CELEX corpus. Texts with many low frequency words are likely to be more difficult to read. The MRC database derives its *word familiarity minimum* scores from Toglia and Battig (1978) and Gilhooly and Logie (1980). Higher scores, based on human ratings, indicate words with greater familiarity (e.g., *hat*), as opposed to lower scores, which indicate less familiar words (e.g., *abdicate*). Texts composed of more familiar words are likely to be more easily comprehended than texts composed of less familiar words. The *Flesch-Kincaid Grade Level* measure is commonly used to estimate the difficulty of student text books. The formula is based on the number of words per sentence and the number of syllables per word. Texts with longer sentences often present readers with more ideas, and texts with longer words imply higher semantic complexity, jointly adding more difficulty to the text overall.

As mentioned previously, the 13 Coh-Metrix indices used in this study were selected based on prior research indicative of their importance. According to that research, the variables accurately represent their respective linguistic constructs (e.g., LSA as a measure of semantic overlap), without being highly inter-correlated. Table 2 provides correlations between the 13 variables of this study. Although many of the correlations between the variables are significant, none of them are strong enough for any of the variables to be excluded ($r \geq \pm 0.70$).

Table 2 Correlations among Coh-Metrix indices

Variable Name	AWPS	CC	CV	CWF	FKGL	LSA	WMV	MTLD	MED	NO	WC	WCM
CC	0.063											
CV	-0.426**	-0.093**										
CWF	-0.269**	-0.045	-0.069									
FKGL	0.555**	0.129**	-0.255**	-0.225**								
LSA	0.074	-0.009	-0.024	-0.170**	0.181**							
WMV	0.206**	0.046	-0.213**	-0.042	0.276**	0.028						
MTLD	0.024	-0.074	-0.005	0.097**	0.147**	-0.362**	-0.051					
MED	-0.328**	-0.069	-0.115**	0.238**	-0.065	-0.309**	-0.079	0.378**				
NO	0.231**	0.040	-0.094**	-0.252**	0.296**	0.556**	-0.075	-0.267**	-0.246**			
WC	-0.078	-0.126*	-0.138	-0.018	-0.257*	0.049	-0.100*	-0.046	-0.005	-0.061		
WCM	-0.018	0.033	-0.134**	-0.057	0.060	-0.021	-0.052	-0.061	-0.063	-0.058	0.073	
MRC	-0.276**	-0.010	-0.073	0.244**	-0.398**	-0.150**	-0.019	-0.079	0.061	-0.228**	0.082*	-0.067

Note: Correlations. *$p < 0.05$, **$p < 0.01$; Average Words Per Sentence (AWPS), Causal Connectives (CC), Causal Verbs (CV), CELEX Word Frequency (CWF), Flesch-Kincaid Grade Level (FKGL), Latent Semantic Analysis (LSA), Words Before Main Verb (WMV), Measure of Textual Lexical Diversity (MTLD), MED Content Words (MED), Noun Overlap (NO), Word Concreteness (WC), Word Concreteness Minimum (WCM), MRC Word Familiarity Minimum (MRC)

Corpus

Our corpus consists of 811 body (middle) paragraphs that were extracted from 1,418 paragraphs of 311 essays. The essays were collected from introductory English composition courses at Mississippi State University (n = 189) (Crossley & McNamara, 2009; McNamara et al., 2010), an introductory psychology course at Northern Illinois University (n = 60) (Wolfe, Britt, & Butler, 2009), and College Board (n = 62). Each essay was written by a different student, and each student wrote in response to one of five essay prompts (e.g., "Does fame bring happiness, or are people who are not famous more likely to be happy?"). Four of the five prompts were designed to mimic prompts created by College Board, while the fifth was an actual College Board prompt (i.e., "Is the world changing for the better?"). The prompts addressed creativity, television, equal rights, cell phone usage, and whether the world is changing for the better or worse. The students were instructed to persuade their readers to take a certain position on the respective prompts.

We used Coh-Metrix to analyze each paragraph. However, it is important to emphasize beforehand that each paragraph was analyzed as an individual text unit. That is, the unit of analysis is the paragraph, not the essay. Each paragraph was categorized according to the total number of paragraphs of the source essay (i.e., the essay in which the paragraph occurred). We refer to this variable as the *total number of paragraphs*. Paragraphs from texts with two or fewer paragraphs were excluded from the analysis because they did not have enough paragraphs with which to occupy all three positions of an essay (i.e., first, middle, last), and thus there were no middle paragraphs in such essays. Additionally, paragraphs from texts with totals of 10 or more paragraphs were excluded due to their infrequency (n = 8, 0.009% of the data). Table 3 provides the frequency distribution of the middle paragraphs. The table shows that essays with five paragraphs were the most frequent, suggesting the continuing prominence of the five-paragraph essay. As such, the 345 middle paragraphs from

Table 3 Frequency distribution of middle paragraphs

Total number of paragraphs	Middle paragraphs	Percentage
3	38	4.69
4	78	9.62
5	345	42.54
6	188	23.18
7	90	11.10
8	30	3.70
9	42	5.18
Total	811	100.00

essays with five paragraphs comprise 42.54% of the total number of paragraphs.

Analysis 1

Table 1 provides the Coh-Metrix descriptive statistics for the middle paragraphs as a function of the number of paragraphs in the essay. We conducted a principal component analysis (PCA) to extract the predicted underlying factors of cohesion and text difficulty. PCA is designed to extract a specified number of latent, unobserved constructs from a set of observed variables. The variables in Table 1 were analyzed using the extraction method 'Principal component' (SPSS), with orthogonal *varimax rotation*. Varimax rotation is a commonly used statistical technique that attempts to reduce the complexity of extracted factor(s) by maximizing the variance of each variable that contributes to the factor(s) (hence the term 'varimax'). This technique results in a simpler interpretation of the relationship(s) between the observed variables and the extracted factor(s), and it ensures that the extracted factors are not inter-correlated (StatSoft, 2010).

Based on the contributions of the variables in Table 1, we extracted two factors through PCA. The rotated (varimax) component matrix shown in Table 4 shows the correlations between each variable and the two extracted factors (i.e., factor loadings). Table 4 also provides the eigenvalues and the percentage of overall variance explained by the factors. The factor loadings aid in interpreting the latent constructs extracted from the variables, and they offer insight into how a variable contributes to a factor. Conventionally, variables that moderately correlate with their respective factor ($r \geq \pm 0.400$) are considered to be important contributors to that factor, while eigenvalues above 1.000 and the percentage of variance explained indicate the 'significance' of the extracted factor. Factor 1 shows moderate correlations with many of the cohesion variables from Table 1, suggesting that those particular variables contribute to the underlying factor of cohesion. Factor 2 also features moderate correlations with the variables predicted to indicate text difficulty, suggesting its contribution to that particular underlying factor – difficulty.

Analysis 2

To examine the effect of the total number of paragraphs in an essay on the cohesion and difficulty of its middle paragraphs, we used the two extracted factors (the standardized variable scores multiplied by their respective loading coefficients for each factor) as dependent variables in two *Hierarchical Linear Models (HLMs)* using the SPSS procedure 'MIXED'. This statistical technique hierarchically

Table 4 Rotated (varimax) component matrix (correlations with extracted factors 1 and 2)

Variable name	Factor 1 (Cohesion)	Factor 2 (Difficulty)
Indices of cohesion		
Noun overlap	**0.693**	0.269
MED content words	**-0.651**	-0.112
Measure of textual lexical diversity	**-0.6s73**	0.212
LSA	**0.750**	0.078
Causal connectives	0.037	0.190
Indices of difficulty		
Causal verbs	0.023	**-0.568**
Max. words before main verb	-0.141	**0.462**
Word concreteness	0.158	-0.349
Word concreteness minimum	0.162	-0.068
CELEX (content) word frequency	-0.399	-0.328
Flesch-Kincaid grade level	0.081	**0.817**
MRC word familiarity minimum	-0.168	**-0.497**
Average words per sentence	0.178	**0.758**
Eigenvalues	1.858	2.791
Variance explained	14.295%	21.471%

Note: Factor loadings over 0.40 appear in bold

structures independent variables to account for variance and interactions at multiple levels simultaneously. HLM is quite often used in classroom studies (with students embedded in classrooms, and classrooms within schools), and its usage on such occasions better accounts for multi-level variance and interactions than does standard multiple regression and analysis of variance (Richter, 2006; Baayen, 2008). Furthermore, HLM is appropriate because it is designed to accommodate unequal observations, which is the case in our corpus. In the HLM, we controlled for random variance in *essay topic, participant (student who wrote the essay)*, and the *interaction between topic and participant*. We also included text length (number of words in each paragraph) as a covariate in order to account for the variance associated with the length of the paragraphs. Factor 1 (cohesion) produced a significant result: $F(6, 221.508) = 2.986$, $p = 0.008$, whereas Factor 2 (difficulty) yielded a marginally significant result: $F(6, 216.427) = 2.051$, $p = 0.060$.

Table 5 displays pairwise comparisons using *Least Significant Difference (LSD)*, a post hoc test that shows patterns in data by analyzing the variance between two variables at a time. LSD tests were conducted for each dependent variable (cohesion, difficulty) used in the HLM. The tests revealed that the differences in the linguistic characteristics of middle paragraphs were most acute in essays that featured 3, 4, and 9 total paragraphs. In other words, essays with 5 to 8 total paragraphs tended to demonstrate no significant differences between one another in terms of cohesion. Thus, these essay lengths (i.e., 5-8 paragraphs) might be deemed prototypical, with respect to the variables of analysis and the frequency of essays within this range (see Tables 1 and 3).

The results suggest that students who write more than 8 paragraphs in response to College Board prompts may be over-explaining their ideas. Example Paragraph 1 in the Appendix illustrates this notion. The example is a body paragraph from a student essay with 9 paragraphs. The paragraph contains 129 words, 54 of which are repeated at least once, and of the 62 content lemmas, 19 are repeated (about 1:3). This repetition results in high overlap values and low lexical diversity values, which are both related to higher cohesion.

In contrast to Example Paragraph 1, Example Paragraph 2 is from an essay with 3 paragraphs. The student does not seem to have sufficiently explained the ideas in the text. The paragraph might benefit from greater cohesion and perhaps an expansion upon the multiple ideas presented in such a short text. Example Paragraph 2 features 89 words, only 27 of which are repeated; in terms of content lemmas, 52 are unique with a mere 8 repeated (about 1:7).

The results from the second analysis offer some evidence that, in order to write cohesively, student writers may benefit by seeking to write essays of between 5 and 8 paragraphs in length. These findings suggest that essay prompts may only warrant 5-8 paragraphs of information, a concept that largely corresponds with the expectations of the College Board and similar organizations.

Discussion

This study provides writing researchers and educators with valuable information concerning the relationship between the total number of paragraphs in an essay and the cohesion

Table 5 Least significant difference (LSD) post hoc tests

Factor 1 (Cohesion)

Total number of paragraphs	4-Paragraphs	5-Paragraphs	6-Paragraphs	7-Paragraphs	8-Paragraphs	9-Paragraphs
3-Paragraphs	-0.300	-0.440**	-0.450*	-0.600**	−0.710*	-1.030**
4-Paragraphs		-0.140	-0.160	-0.300	-0.420	-0.730*
5-Paragraphs			-0.010	-0.160	-0.270	-0.590*
6-Paragraphs				-0.150	-0.260	-0.580*
7-Paragraphs					-0.110	-0.430
8-Paragraphs						-0.320
Factor 2 (Difficulty)						
3-Paragraphs	-0.576**	-0.266	-0.452*	-0.578*	-0.113	-0.407
4-Paragraphs		0.310*	-0.124	-0.002	0.463	-0.169
5-Paragraphs			-0.186	-0.312	0.154	-0.141
6-Paragraphs				-0.126	0.339	-0.045
7-Paragraphs					0.465	-0.170
8-Paragraphs						-0.295

Note: Mean differences. *$p < 0.05$, **$p < 0.01$

and difficulty of its middle paragraphs. The relationships suggest that the cohesion and difficulty of middle paragraphs are not consistent. That is, when providing instruction to students on how to write cohesive paragraphs or when grading student essays, it may be advantageous to consider how much they plan to write, or how much they have written.

Differences in cohesion between the middle paragraphs across student essays seem to be most prominent when the total number of paragraphs is 3, 4, or 9. These differences suggest that students who write fewer than 5 paragraphs may not be writing enough to adequately convey the ideas about the prompts and that students who write more than 8 paragraphs are being repetitive (see Example Paragraphs 1 and 2).

McNamara et al. (2009) found that ratings of writing proficiency are not related to indices of cohesion. Interestingly though, text difficulty variables (e.g., Words before Main Verb) were significantly predictive of writing proficiency ratings. It is possible that the results of McNamara and colleagues were slightly altered by an interference of the linguistic features of first and last paragraphs; if differing linguistic features across paragraphs are considered, then cohesion may be found to play an important role in assessments of writing proficiency.

The results of this study also suggest that computational assessments of student essays would benefit by considering not only *how much* the student has written in terms of the number of paragraphs, but *what* the student has written in terms of linguistic characteristics. That is, analyzing essays by parts (i.e., first, middle, last paragraphs) may increase the accuracy of computational essay assessment by considering and accommodating differing rhetorical functions across paragraphs based on their linguistic characteristics.

Our results also confirmed that student essays most commonly include five paragraphs (42.54%). To some extent, this is to be expected because students are often encouraged to write five paragraphs in response to essay prompts, whether by teachers or by organizations such as College Board (Nunnally, 1991; Smith, 2006). Including five paragraphs in an essay seems optimal in relation to the goal presented by the task, which is to write brief essays that address a particular prompt. Unfortunately, the unwavering predominance of the 5-paragraph essay results in an unequally distributed corpus. Nonetheless, it is an aspect of our data that cannot be remedied, even given that HLM better accommodates the data set. That is, if the ecological validity of the data set is to be maintained, the number of paragraphs will necessarily vary. Given the target task examined in this study, it is likely that the number of paragraphs of essays would have an unequal distribution (with a predominance of 5-paragraph essays).

Our hope is that having a clearer view of the relationship between the total number of paragraphs of an essay and the linguistic characteristics that affect essay scores may better equip writing researchers and educators to pinpoint specific issues underlying student difficulties and to subsequently provide them with more appropriate feedback. Moreover, an examination of the relationship between the total number of paragraphs and essay scores sheds light on the constituents of essay quality and the rhetorical values (i.e., preferences) of essay graders

Appendix

Example Paragraph 1

Today the government puts regulations on television programs trying to manipulate the mind of all Americans, so we think we know exactly what is going on in the world today, but really we have no idea. Do you think they would show us everything about the war or terrorists, no because they are trying to comfort us and tell us everything will be ok when really we might get blown up tomorrow and have no idea. Television gives false sense of power and knowledge to the people of today. The violence we watch on television is the exact violence the government wants us to watch. In a way, the government acts like an opium to the people giving us a false sense of reality and what is really happening.

Example Paragraph 2

The War in Iraq has stimulated many controversial thoughts in the minds of Americans. Why are we sending troops half-way across the world, to fight for the freedom of a foreign country? Foreign by distance, by culture. By race, by religion for many. Many view this war as nothing more than a plethora of death and destruction. American sit, eagerly before their televisions, horribly enthralled in news program after news program, breaking news of the progress of the troops, of the terrible transgressions condemning the innocent citizens of Baghdad.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R* (1st ed.). Cambridge, England: Cambridge University.

College Board (2009). *College Board Announces Scores for New SAT® with Writing Section*. Retrieved from http://www.collegeboard.com/press/releases/150054.html on October 25, 2009.

Coltheart, M. (1981). MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33*, 497–505.

Crossley, S. A., & McNamara, D. S. (2009). Computationally assessing lexical differences in second language writing. *Journal of Second Language Writing, 17*(2), 119–135.

Crossley, S. A., White, M. J., McCarthy, P. M., & McNamara, D. S. (2009). *The effects of elaboration and cohesion on human evaluations of writing proficiency. Poster presentation at the, Society for Computers in Psychology Conference (SCiP)*. Boston: Massachusetts.

Duran, N., Bellissens, C., Taylor, R., & McNamara, D. (2007). Qualifying text difficulty with automated indices of cohesion and semantics. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the cognitive science society* (pp. 233–238). Austin, TX: Cognitive Science Society.

Essay Info (2009). *Five-paragraph Essay Writing*. Retrieved from http://essayinfo.com/essays/5-paragraph essay.php on October 25, 2009.

Gilhooly, K. J., & Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behaviour Research Methods & Instrumentation, 12*, 395–427.

Graesser, A.C. & McNamara, D.S. (in press). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science.*

Graesser, A.C., McNamara, D.S., & Kulikowich, J.M. (2010). Theoretical and automated dimensions of text difficulty: How easy are the texts we read? Manuscript sumbitted for publication.

Graesser, A. C., McNamara, D. S., Louwerse, M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36*, 193–202.

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.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum.

Lightman, E. J., McCarthy, P. M., Dufty, D. F., & McNamara, D. S. (2007). The structural organization of high school educational texts. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the twentieth international Florida artificial intelligence research society conference* (pp. 235–240). Menlo Park, California: The AAAI.

Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variation in language and cohesion across written and spoken registers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual cognitive science society* (pp. 843–848). Mahwah, NJ: Erlbaum.

McCarthy, P. M., Guess, R. H., & McNamara, D. S. (2009). The components of paraphrase evaluations. *Behavioral Research Methods, 41*, 682–690.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*, 381–392.

McCarthy, P. M., Lewis, G. A., Dufty, D. F., & McNamara, D. S. (2006). Analyzing writing styles with Coh-Metrix. In *Proceedings of the Florida Artificial Intelligence Research Society International Conference (FLAIRS)* (pp. 764-770).

McCarthy, P. M., Renner, A. M., Duncan, M. G., Duran, N. D., Lightman, E. J., & McNamara, D. S. (2008). Identifying topic sentencehood. *Behavioral Research Methods, 21*, 364–372.

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55*, 51–62.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2009). Linguistic features of writing quality. *Written Communication, 27*(1), 57–86.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*, 292–330.

Meyer, B. F., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal, 21*, 121–143.

Nunnally, T. E. (1991). Breaking the five paragraph theme barrier. *English Journal, 80*, 67–71.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19*, 228–242.

Richter, T. (2006). What is wrong with ANOVA and multiple regression? Analyzing sentence reading times with hierarchical linear models. *Discourse Processes, 41*, 221–250.

Smith, K. (2006). In defense of the five-paragraph essay. *English Journal, 95*, 16–17.

StatSoft, Inc. (2010). Electronic Statistics Textbook. Tulsa, OK: StatSoft. Retrieved from http://www.statsoft.com/textbook/ on February 16, 2010.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

Toulmin, S. (1969). *The uses of argument*. Cambridge, England: Cambridge University.

Wesley, K. (2000). The ill effects of the five paragraph theme. *English Journal, 90*, 57–60.

Wiley, M. (2000). The popularity of formulaic writing (and why we need to resist). *English Journal, 90*, 61–67.

Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication, 26*(2), 183–209.