Idea Generation in Student Writing: Computational Assessments and Links to Successful Writing

Scott A. Crossley[1], Kasia Muldner[2], and Danielle S. McNamara[3]

Abstract

Idea generation is an important component of most major theories of writing. However, few studies have linked idea generation in writing samples to assessments of writing quality or examined links between linguistic features in a text and idea generation. This study uses human ratings of idea generation, such as *idea fluency, idea flexibility, idea originality*, and *idea elaboration*, to analyze the extent to which idea generation relates to human judgments of essay quality in a corpus of college student essays. In conjunction with this analysis, linguistic features extracted from the essays are used to develop a predictive model of idea generation to further understand relations between the language features in an essay and the idea generation scores assigned to that essay. The results indicate that essays rated as containing a greater number of ideas that were flexible, original, and elaborated were judged to be of higher quality. Two of these features (elaboration and originality) were significant predictors of essay quality scores in a regression analysis that explained 33% of the variance in human scores. The results also indicate that

[1]Department of Applied Linguistics/ESL, Georgia State University, Atlanta, GA, USA
[2]Institute of Cognitive Science, Carleton University, Ottawa, ON, Canada
[3]Department of Psychology, Arizona State University, Tempe, AZ, USA

Corresponding Author:
Scott A. Crossley, Department of Applied Linguistics/ESL, Georgia State University, 25 Park Place, Suite 1500, Atlanta, GA 30303, USA.
Email: scrossley@gsu.edu

idea generation is strongly linked to language features in essays. Specifically, the use of unique multiword units, more difficult words, semantic but not lexical similarities between paragraphs, and fewer word repetitions explained 80% of the variance in human scores of idea generation. These results have implications for writing theories and writing practice.

Idea generation is an important component of human innovation and creativity (Gist, 1989; Paulus & Yang, 2000). In the present study, we focus on idea generation in a specific context: college-level student essay writing. Existing cognitive models of writing highlight that the quality of a writing sample can be attributed to the number and types of ideas embedded in the sample (Bereiter & Scardamalia, 1987; Galbraith, Ford, Walker, & Ford, 2005; Kellogg, 1988, 1994). However, few if any studies have quantified idea generation in student writing or have linked idea generation in writing samples to assessments of writing quality. Importantly, to the best of our knowledge no work has explored the relations between linguistic features of student essays (i.e., lexical and cohesion features found in the text), elements of idea generation, and human ratings of essay quality (cf. Crossley & McNamara, 2016). Thus, our goal is to examine the number, type, novelty, and depth of ideas generated by student writers, investigate links between generated ideas and linguistic features in the text, and assess how the number and types of ideas in a text can predict writing success. Collectively, such approaches provide support for the inclusion of idea generation as a component of cognitive writing models as well as provide a better understanding of how elements of idea generation interact with writing quality.

Idea Generation and Creativity

Idea generation is a fundamental aspect of writing that is accounted for by the majority of cognitive models of writing (e.g., Galbraith, 2009; Hayes, 1996; Hayes & Flower, 1980; Kellogg, 1994), which account for both the generation of ideas and the ways these ideas are translated onto paper (Torrance, Thomas, & Robinson, 1996). However, idea generation has received less attention compared to other cognitive processes described in these models, such as translation, review, and revision (Berninger et al., 2009).

In traditional models of writing, such as Bereiter and Scardamalia's (1987) knowledge-telling and knowledge-transforming model and Flower and Hayes's (1980, 1981) problem-solving model, idea generation was part of the planning component and described as an explicit and effortful process that consists of writers strategically and deliberately searching for ideas (Bereiter & Scardamalia, 1987; Flower & Hayes, 1980). Idea generation involved the writer identifying a memory probe and using this probe to explore long-term memory. The probe is usually based on the topic of the writing, but writers could generate their own cues based on rhetorical demand. The writer then evaluates the output from the memory search. If the output is considered relevant and important, it is written down. This process is recursive and continues until the writer has enough content for the writing task. Later revisions of the Flower and Hayes model included the incorporation of working memory, motivation, and affect (Hayes, 1996, 2006). Working memory was added to the model to provide a means for writers to hold and process information retrieved from long-term memory. Additional components of the model included writers' motivation and affect, long-term memory, and cognitive processes such as reflecting, interpreting, and producing text.

Writers can use a variety of methods to retrieve and generate ideas. These methods include the generation of ideas prompted by reading the assignment, the activation of new ideas when ideas are translated into text, and the development of ideas while the text is structured or during the revision process (van den Bergh & Rijlaarsdam, 1999). However, idea generation can be a difficult prospect for many writers, especially for children and less skilled individuals. According to many writing models, this is because less skilled writers dedicate working memory resources to lower-level linguistic and text production processes such as transcription and vocabulary retrieval (Alamargot & Fayol, 2009; Berninger, Abbott, Abbott, Graham, & Richards, 2002; McCutchen, 1996, 2000; McCutchen, Covill, Hoyne, & Mildes, 1994) at the expense of higher-level processes such as coordinating ideas, planning, evaluation, and considering rhetorical elements such as genre, audience, and goals (Kellogg, 1996, 2001). More skilled writers have generally automated these lower-level processes and thus can dedicate working memory resources to the higher-level processes.

A number of studies that focus on children and adult writers provide some evidence for the role of working memory and automation in writing. However, the evidence is mixed. Studies have shown that children have a more difficult time with idea generation in written as compared to oral production (Hayes & Berninger, 2010) because, during writing, children dedicate cognitive resources to lower level writing skills such as finger movement, letter writing, spelling, and grammar, which are not needed in oral production (Bereiter

& Scardamalia, 1987). Support for this is found in positive, moderate correlations reported between writing quality and scores on working memory span tasks (i.e., speaking span) for elementary and middle school students, although only weak, positive correlations were reported with reading span tests (McCutchen et al., 1994). Similar findings were reported by Babayigit and Stainthorp (2011), who conducted a one-year longitudinal study of second grade and fourth grade students. They found that working memory scores (as measured by a modified version of the reading span task) were positively and moderately related to both reading comprehension and writing. However, working memory capacity scores failed to make unique contributions when other measures, such as vocabulary, were considered.

It is argued that as writers move out of childhood, they are more likely to automate lower level writing skills and thus can dedicate greater memory resources to idea generation, which can lead to the production of more written ideas (Bourdin, Fayol, & Darciaux, 1996). However, studies directly associating writing quality with working memory scores in college level adults have found that the working memory capacity (i.e., operation span scores) do not correlate with writing quality scores (whereas vocabulary knowledge shows strong associations with writing quality; Allen, Snow, Jackson, Crossley, & McNamara, 2014). Studies have shown an increase in the number of written ideas as a function of grade level (from second, fourth, sixth, and eighth grade), but links between this increase and working memory skills are scant.

While there is some evidence that memory resources play a role in idea generation, there is stronger evidence that writing strategies can help students focus their attention on a specific set of writing problems to reduce the difficulty of the writing process (Graham & Perrin, 2007). For instance, there are a number of planning strategies that can be used to increase the generation of ideas for all students. One of these strategies is outlining. Research on outlining has shown beneficial effects on text quality (Galbraith et al., 2005; Kellogg, 1988, 1994). For instance, Kellogg (1988, 1994) reported that outlining positively affected essay scores and attributed the positive relation to outlining enabling writers to separate idea generation and organization from text production. Galbraith et al. (2005) reported significant, positive correlations between writing quality and various elements of planning. These included positive correlations between essay quality and the amount of content in a writer's plan, the number of new ideas introduced during outlining, and the degree to which the plan included rhetorical groupings of content. Thus, Galbraith and colleagues provided some evidence that the number of ideas generated by writers can be linked to writing quality.

Another important component of idea generation is the value placed on the originality of ideas, the appropriateness of ideas, and the breadth of ideas

generated by the writer. One approach to evaluating these aspects is to consider the writer's problem space (all of the ideas considered by the writer) and the task solution set (all of the possible ideas for the task). Ideas can be evaluated as more original and of greater breadth if they are contained in the task solution set, but not generated by most writers (Torrance et al., 1996). This approach to establishing the values of ideas in a written text overlaps strongly with theories of creativity. In particular, these theories propose that creativity requires an idea(s) that is novel, because it is either different from other ideas produced by that individual or different from the ideas produced by other individuals (Amabile, 2013; Mayer, 1999).

How can the creativity of a given item be assessed? One method, referred to as the *consensual assessment technique*, involves using expert judges to evaluate the creativity of the target item (Baer & McKool, 2009). The judges are not given a precise definition of creativity but rather must rely on their own intuitions of what creativity entails in their domain. An alternative method involves decomposing the creativity construct into subscales, such as *fluency, flexibility, originality*, and/or *elaboration* components (e.g., Feldhusen & Goh, 1995; Runco & Okuda, 1991; Silver, 1997; Snyder, Mitchell, Bossomaier, & Pallier, 2004). These subscales overlap strongly with theories of planning and have been applied to evaluating the idea generation of writing in particular (Cheung, Tse, & Tsang, 2001; Majid, Tan, & Soh, 2003), as well as other creative endeavors in general (Antes & Mumford, 2009; Diakidoy & Constantinou, 2000-2001; Levav-Waynberg & Leikin, 2012; Muldner & Burleson, 2015; Sosik, Kahai, & Avolio, 1998). The subscales are operationalized below in terms of idea generation in writing:

- *Fluency:* the number of ideas generated in a text. This component can correspond to the initial idea generation phase during writing activities. In contrast to flexibility and originality components, fluency corresponds to the pure number of ideas produced, without assessing their creative potential.
- *Flexibility:* how different a given individual's ideas are from other ideas produced by that individual. This is typically captured by analyzing ideas according to higher-level categories to which they belong. For instance, given the common "think of all the uses of a brick" task, an individual whose ideas all involve building materials (category *a*) is less "flexible" than an individual whose ideas also involve art supplies (category *b*).
- *Originality:* how different an individual's ideas are from others' ideas. Originality differs from fluency by considering an individual's ideas with respect to the sample population.

   – *Elaboration:* the degree to which a given idea is elaborated or expanded upon.

Current Study

In the present article, we investigate the fluency, flexibility, originality, and elaboration of ideas generated by college student writers in produced text during a timed writing assignment. We also investigate links between generated ideas and linguistic features and assess how these link to human ratings of essay quality. The research questions that guide this study are the following:

1. Are human judgments of essay features related to idea generation associated with human ratings of essay quality?
2. What linguistic features of a text are predictive of idea generation?

The first question allows us to assess links between idea generation in text and successful student writing. The second question allows us to model idea generation such that a better understanding of the linguistic features related to idea generation emerges. Such models will inform our understanding of writing and, in turn, potentially inform the development of strategies and interventions to help students generate ideas more effectively while writing to produce higher quality texts.

Method

Corpus

To address our research questions, we used the corpus of essays from Crossley and McNamara (2011). This corpus comprised 313 timed independent essays written on SAT prompts (i.e., essays that require no background knowledge of integration of source material). The essays were written by undergraduate freshmen composition students at a large university in the southeastern United States. The students were given 25 minutes to write an essay, during which no outside referencing was allowed. Two SAT prompts were used in the data collection with students randomly assigned to either prompt. All students were native speakers of English.

Human Ratings

Each essay was read and scored by two trained raters on overall quality (i.e., a holistic score) and on idea generation and style (i.e., analytic scores).

Different raters were used for scoring holistic quality than were used for scoring idea generation and style features. The holistic grading scale was based on a standardized rubric commonly used in assessing SAT essays (see the appendix). The idea generation rubric corresponded to 6-point Likert-type scales, one each for fluency, flexibility, originality, and elaboration. The rubric also included a style section that comprised 6-point Likert-type scales for humor, metaphor and simile, and word play (see the appendix for details on each scale).

In all cases, the trained raters had either a master's or a doctoral degree in English, and each rater had at least 2 years of experience teaching university-level composition classes. The raters were informed that the distance between each score on a given scale (e.g., the fluency scale) was equal. The raters were first trained to use the rubric with 20 practice essays that were not part of the corpus. After they reached an interrater reliability of at least $r = .60$ for the analytic scores and at least $r = .70$ for the holistic score,[1] they then scored the 313 essays in the corpus independently. After scoring was completed, differences between raters were calculated. If the difference in ratings on a feature was less than 2 points (e.g., one rater assigned the essay a score of 4 and the second rater assigned the essay a score of 5), an average score was computed for that essay feature (assessed using the corresponding Likert-type scale, e.g., fluency scale). If the difference was 2 points or greater (e.g., one rater assigned the essay a score of 4 and the other rater assigned the essay a score of 6), the raters attempted to adjudicate the final rating through discussion. If no agreement was met, the score was not changed. Correlations and weighted kappa scores between the raters after adjudication for the holistic score and the analytic scales related to idea generation and style are reported in Table 1. Two features (*originality* and *word play*) reported a Cohen's kappa value below substantial (<.60). The holistic scoring rubric and the idea generation and style scoring rubric are located in the appendix.

Idea Generation and Linguistic Features

To explore the relations between idea generation and linguistic variables, we selected variables related to basic text properties, lexical sophistication, text cohesion, and rhetorical style from the Writing Assessment Tool (WAT; Crossley, Roscoe, & McNamara, 2013; McNamara, Crossley, & Roscoe, 2013), the Tool for the Automatic Assessment of Cohesion (TAACO; Crossley, Kyle, & McNamara, in press), and the Tool for the Automatic Assessment of Lexical Sophistication (TAALES; Kyle & Crossley, 2015). We selected indices related to basic text properties under the presumption that essays with greater fluency would be longer and essays with greater

Table 1. Interrater Reliability for Essay Scores.

Scale	Correlation	Weighted kappa
Holistic quality	.789	.798
Fluency	.801	.763
Flexibility	.647	.642
Originality	.573	.533
Elaboration	.707	.703
Humor	.718	.715
Metaphor and simile	.686	.683
Word play	.492	.488

originality would contain a greater number of unique n-grams (i.e., single words and double and triple word combinations). We selected indices related to lexical sophistication with the hypothesis that texts with greater fluency, originality, and flexibility would contain more sophisticated words (i.e., less frequent and more academic words). We selected text cohesion and rhetorical style indices under the hypotheses that creative discourse demonstrates greater text repetition (Carter, 2016) and that texts with greater elaboration would include more overlap of ideas and more connectives between ideas. A brief description of the linguistic indices used in this study is provided below. We refer the reader to McNamara et al. (2013), Crossley et al. (in press), and Kyle and Crossley (2015) for more in depth discussions of the indices.

WAT. WAT (Crossley et al., 2013; McNamara et al., 2013) was developed specifically to assess writing quality. As such, it includes a number of writing specific indices related to global cohesion, contextual cohesion, n-gram (i.e., multiword units) accuracy, lexical sophistication, key word use, and rhetorical features. Specifically, WAT calculates the semantic overlap between paragraphs (initial to middle paragraphs, middle paragraphs to final paragraph, and initial paragraph to final paragraph) and between the essay prompt and the essay. The semantic similarities among the paragraphs, essay, and prompt are calculated using latent semantic analysis (LSA) cosine values. WAT also calculates the cohesion between the text and the prompt by computing the semantic similarity between the prompt and the response using both LSA and key word analyses as well as reporting two indices of prompt-specific key word use that can be used to measure the topic development in text. WAT includes a number of lexical diversity indices that control for text length. These indices include MTLD (McCarthy & Jarvis, 2010) and D (Malvern, Richards, Chipere, & Durán, 2004), both of which report variations of

type-token ratio counts that control for text length effects. Last, WAT reports on a variety of lexical categories related to rhetorical style. These include amplifiers (*absolutely, fully*), hedges (*maybe*), indirect pronouns (*all, no one*), exemplification (*for instance*), copula verbs (*be, seem, appear*), private verbs (*assume, hear*), public verbs (*claim, promise*), persuasive verbs (*agree, request*), vague nouns (*whatever, whoever*), negations (*not, no*), modals (*can, would, should*), contractions (*I'm, I've*), and downtoners (*almost, barely*).

TAACO. TAACO (Crossley et al., in press) incorporates over 150 classic and recently developed indices related to text cohesion. For a number of indices, the tool incorporates a part of speech (POS) tagger from the Natural Language Tool Kit (Bird, Loper, & Klein, 2009) and synonym sets from the WordNet lexical database (Miller, 1995). The POS tagger affords the opportunity to look at content words (i.e., nouns, verbs, adjectives, and adverbs) as well as function words (i.e., determiners, propositions). TAACO provides linguistic counts for both sentence and paragraph markers of cohesion and incorporates WordNet synonym sets. Specifically, TAACO calculates type token ratio (TTR) indices (for all words, content words, function words, and n-grams), sentence overlap indices that assess local cohesion for all words, content words, function words, POS tags, and synonyms, paragraph overlap indices that assess global cohesion for all words, content words, function words, POS tags, and synonyms, and a variety of connective indices such as logical connectives (e.g., *moreover, nevertheless*), conjuncts (*however, furthermore*), and the incidence of *and*.

TAALES. TAALES (Kyle & Crossley, 2015) incorporates about 150 indices related to basic lexical information (i.e., the number of words and n-grams, the number of word and n-gram types), lexical frequency (i.e., how many times an item occurs in a reference corpus), lexical range (i.e., how many documents in which a reference corpus an item occurs), psycholinguistic word information (e.g., concreteness, familiarity, meaningfulness), and academic language (i.e., items that occur more frequently in an academic corpus than in a general use corpus) for both single words and multiword units (n-grams such as bigrams and trigrams). The frequency and range indices draw on the British National Corpus (2007), Thorndike-Lorge Corpus (Thorndike & Lorge, 1944), Brown corpus (Kucera & Francis, 1967), Brown verbal frequencies (Brown, 1984), which were compiled based on the London-Lund corpus of English Conversation (Svartvik & Quirk, 1980), and the SUBTLEXus corpus of subtitles (Brysbaert & New, 2009). Psycholinguistic word information indices draw on the Medical Research Council psycholinguistic database (Coltheart, 1981), which includes word scores for familiarity, concreteness, imageability, and

meaningfulness (how many associations a word has). Also included in TAALES are psycholinguistic word information indices that incorporate recently collected concreteness norms for single words and two-word units (Brysbaert, Warriner, & Kuperman, 2014) and age of acquisition norms for single words (i.e., at what age a word is estimated to be learned; Kuperman, Stadthagen-Gonzales, & Brysbaert, 2012). Academic language frequencies are based on the Academic Word List (Coxhead, 2000) and the Academic Formulas List (Simpson-Vlach & Ellis, 2010).

Statistical Analysis

We first examine links between the analytic idea generation and style scores and essay quality scores to identify which analytic scores correlate with essay quality scores. We use those analytic scores that correlate with essay quality in a subsequent linear regression analysis to examine how much of the variance in the human scores of essay quality is predicted by the analytic scores.

To refine the idea generation and style scores and increase their validity, we conducted an exploratory factor analysis using the human scores for the idea generation and style items to investigate potential subscales for the ratings. Our goal was to develop a weighted subscale for idea generation alone based on co-occurrence factors in the human ratings found in the idea generation and style rubric. To accomplish this, the average human ratings for the seven analytic items from the idea generation and style scoring rubric were entered into a factor analysis to identify which items co-occurred and could be combined to form an idea generation subscale. We then used this subscale to calculate a correlation between it and the human judgments of essay quality and as a dependent variable in a second regression analysis that examined links between the idea generation scores and the selected linguistic indices.

Prior to conducting both regression analyses, the essay corpus was divided into training and test sets using a 67/33 split (Witten, Frank, & Hall, 2011). Using the training set, the analytic features and those indices reported by WAT, TAACO, and TAALES that lacked normal distributions were discarded. Correlations were then calculated for the remaining variables to determine whether there was a statistical ($p < .05$) and meaningful (at least a small effect size, $r > .1$) relation between the selected indices and the dependent variable (either the human essay scores or the idea generation component score developed in the factor analysis). Indices that were highly collinear ($r \geq .900$) were flagged, and the index with the strongest correlation with human scores was retained while the other indices were removed. The remaining indices were included as predictor variables in a stepwise multiple regression to explain the variance in the idea generation component scores.

Table 2. Correlations Between Analytic Scores and Essay Quality.

Index	r	p
Elaboration	.560	<.001
Fluency	.520	<.001
Flexibility	.387	<.001
Metaphor and simile	.200	<.001
Originality	.177	<.010
Word play	.130	<.050
Humor	.055	>.050

The model from the stepwise regression was then used to predict the variance in the idea generation component scores for the essays in the test set.

Results

Relationship Between Analytic Idea Generation and Style Scores and Essay Quality Scores

To address Research Question 1, we conducted correlations between the analytic scores for idea generation and style (e.g., *fluency, elaboration, humor*) and the essay quality score for each essay in the corpus. As shown in Table 2, the positive correlations indicated medium or strong effects for scores related to idea generation, including *elaboration, fluency*, and *flexibility*, and small effects for scores related to style, including *metaphor, originality*, and *word play.* The correlations indicate that essays rated as containing more elaborated ideas, more ideas, different ideas, and unique ideas were scored higher in quality. In addition, essays scored as containing more metaphors and similes and greater word play were also scored higher in quality.

To analyze which analytic features best predicted the human judgments of essay quality, we conducted a stepwise regression analysis using the six significant indices as the independent variables. This yielded a significant model, $F(2, 212) = 53.182$, $p < .001$, $r = .580$, $R^2 = .336$. Two variables were significant and positive predictors of the essay quality scores: *elaboration* and *fluency*. These two variables explained 34% of the variance in the human scores of essay quality for the 214 essays in the training set (see Table 3 for additional information). When the model was applied to the test set, it yielded $r = .558$, $R^2 = .311$, indicating that the two variables together explained 31% of the variance in the essay quality scores for the 99 essays in the test set and that the model is stable and generalizable.

Table 3. Stepwise Regression Analysis and Significance Values for Analytic Features Predicting Essay Quality Scores.

Entry	Index added	r	R^2	R^2change	B	B	SE	t
1	Elaboration	.561	.315	.315	0.474	0.373	0.117	4.044**
2	Fluency	.580	.336	.021	0.269	0.238	0.104	2.587*

Note: *B* = unstandardized β; B = standardized; *SE* = standard error. Estimated constant term is 3.901.
*$p < .05$. **$p < .001$.

Table 4. Factor Analysis: Eigen Loadings for Components.

Item	Idea generation component	Humor component
Fluency	.890	
Flexibility	.832	
Elaboration	.809	
Originality	.535	
Metaphor and simile	.509	
Humor		.824
Word play		.615

Idea Generation Component Score

To obtain an overall idea generation score, we conducted an exploratory factor analysis using the human scores for the seven idea generation and style items to investigate potential components representing the ratings. A Bartlett's test of sphericity was statistically significant ($p <.001$), and the Kaiser-Meyer-Olkin measure of sampling adequacy reported .693 indicating underlying structures. The scree plot suggested the extraction of two components, which was also supported by the percentage of variance explained by the initial eigenvalues between the second and third factors. The principal axis factoring using a varimax rotation also identified two components. The items that loaded onto the first component, which we labeled *idea generation*, included elements from the original *idea generation* (*fluency, flexibility, originality, elaboration*) and one element for the *style* scale (*metaphor*). The items that loaded onto the second component, which we labeled *humor*, were *humor* and *word play*. All items loaded onto their respective factors with eigenvalues > .500 (see Table 4). The components for both idea generation and humor were calculated by weighting the items based on their eigenweights in the factors and averaging these weighted scores across the items for each factor. For the present analysis,

we focus on the idea generation component score, which we used in a subsequent regression analysis along with selected linguistic variables to examine the potential for language features in the text to explain idea generation.

Relationship Between Idea Generation Component Scores and Essay Quality

To further explore the relationship between idea generation and essay quality, we calculated the correlation between the idea generation component scores and the essay quality scores. The correlation, $r(313) = .526$, $p < .001$, provides confirmation for the previous regression analysis by indicating a strong relationship between the idea generation component scores and human ratings of essay quality. Specifically, essays judged as containing greater idea generation were also judged to be of higher quality.

Linguistic Indices' Correlations With Idea Generation Component Scores

To address Research Question 2, we conducted correlations between the selected linguistic indices and the idea generation component score. Of the available variables that were normally distributed, 44 demonstrated significant correlations with the idea generation component score while not demonstrating significant multicollinearity with other variables ($r \geq .900$). The correlations for these variables are reported in Table 5.

Regression Analysis to Predict Idea Generation Component Score

To analyze which linguistic features predicted the idea generation component score, we conducted a stepwise regression analysis using the significant indices as the independent variables. This yielded a significant model, $F(6, 212) = 163.235$, $p < .001$, $r = .906$, $R^2 = .820$. Six variables, three reported by TAACO, two reported by WAT, and one reported by TAALES, were significant predictors of the idea generation component scores. These variables were *number of bigrams types, age of acquisition: content words, semantic similarity: introduction to body paragraphs, function word TTR, incidence of exemplifications*, and *adjacent overlap adjectives: sentences.* The first three variables were positive predictors of the idea generation component scores, while the last three were negative predictors.

Table 5. Correlations Between Idea Generation Component Scores and Computational Indices.

Index	Tool	Construct	*r*	*p*
Number of bigrams types	TAALES	Lexical	.879	<.001
Function word TTR	TAACO	Cohesion	−.733	<.001
Trigram written frequency	TAALES	Lexical	.644	<.001
Incidence of logical connectives	TAACO	Cohesion	−.637	<.001
Adjacent overlap function words: Paragraphs	TAACO	Cohesion	.605	<.001
Semantic similarity: Introduction to body paragraphs	WAT	Cohesion	.517	<.001
Academic word list	TAALES	Lexical	.496	<.001
Semantic similarity: Body to conclusion paragraphs	WAT	Cohesion	.478	<.001
Type token ratio	TAACO	Cohesion	−.474	<.001
Incidence of "and"	TAACO	Cohesion	.442	<.001
Adjacent overlap nouns: Paragraphs	TAACO	Cohesion	.428	<.001
Incidence of indirect pronouns	WAT	Rhetoric	−.423	<.001
Semantic similarity: Body paragraphs	WAT	Cohesion	.414	<.001
Incidence of negations	WAT	Rhetoric	.407	<.001
Adjacent overlap verbs: Paragraphs	TAACO	Cohesion	.395	<.001
Adjacent overlap pronouns: Paragraphs	TAACO	Cohesion	.383	<.001
Semantic similarity verbs: Paragraphs	TAACO	Cohesion	.374	<.001
Lexical diversity D	WAT	Cohesion	.350	<.001
Academic formula list	TAALES	Lexical	.339	<.001
Adjacent overlap adverbs: Paragraphs	TAACO	Cohesion	.329	<.001
Adjacent overlap content words: Sentences	TAACO	Cohesion	−.300	<.001
Incidence of conjuncts	TAACO	Cohesion	.268	<.001
Incidence of downtoners	WAT	Rhetoric	.266	<.001
Incidence of persuasive verbs	WAT	Rhetoric	.239	<.001
Incidence of emphatics	WAT	Rhetoric	.234	<.001
Semantic similarity: Introduction to conclusion paragraphs	WAT	Cohesion	.229	<.001
Adjacent overlap adjectives: Sentences	TAACO	Cohesion	−.227	<.001
Word imageability	TAALES	Lexical	.211	<.001
Incidence of exemplifications	WAT	Rhetoric	−.207	<.001
Content word type token ratio	TAACO	Cohesion	−.204	<.001
Incidence of public verbs	WAT	Rhetoric	.204	<.001
Lexical diversity MTLD	WAT	Cohesion	.195	<.001
Range score content words: SUBTLEXus	TAALES	Lexical	−.181	<.001

(continued)

Table 5. (continued)

Index	Tool	Construct	r	p
Age of acquisition: Content words	TAALES	Lexical	.179	<.001
Incidence of modals	WAT	Rhetoric	−.176	<.010
Word familiarity: Content words	TAALES	Lexical	−.171	<.010
Incidence of amplifiers	WAT	Rhetoric	.165	<.010
Incidence of seem	WAT	Rhetoric	.150	<.010
Bigram type token ratio	TAACO	Cohesion	−.150	<.010
Concreteness: Content words	TAALES	Lexical	.118	<.050
Meaningfulness: All words	TAALES	Lexical	−.118	<.050
Incidence of contractions	WAT	Rhetoric	.117	<.050
Incidence of vague nouns	WAT	Rhetoric	−.111	<.050
Incidence of hedges	WAT	Rhetoric	.101	<.050

Note: TAACO = Tool for the Automatic Assessment of Cohesion; TAALES = Tool for the Automatic Assessment of Lexical Sophistication; TTR = type token ratio; WAT = Writing Assessment Tool.

Table 6. Stepwise Regression Analysis and Significance Values for Linguistic Indices Predicting Idea Generation Component Scores.

Entry	Index added	r	R^2	R^2change	B	B	SE	t
1	Number of bigrams types	.874	.764	.764	0.015	0.001	0.651	13.322**
2	Age of acquisition: Content words	.884	.781	.017	0.903	0.197	0.135	4.589**
3	Semantic similarity: Introduction to body paragraphs	.893	.797	.016	0.515	0.126	0.136	4.074**
4	Function word TTR	.899	.809	.011	−5.024	1.362	−0.174	−3.689**
5	Incidence of exemplifications	.903	.816	.007	−0.418	0.132	−0.094	−3.174*
6	Adjacent overlap adjectives: Sentences	.906	.82	.004	−9.629	4.289	−0.067	−2.245*

Note: B = unstandardized β; B = standardized; SE = standard error; TTR = type token ratio. Estimated constant term is 3.901.
*$p < .05$. **$p < .001$.

　　The six variables together explained 82% of the variance in the idea generation component scores for the 218 essays in the training set (see Table 6 for additional information). When we applied the model to the test set, the model yielded $r = .904$, $R^2 = .817$, indicating that the six variables together explained 82% of the variance in the idea generation component scores for the 97 essays in the test set and that the model is stable and generalizable.

Discussion

Idea generation is an important component of most major theories of writing. Writing researchers have argued that the number of ideas generated and the originality of those ideas make important contributions to writing success (Alamargot & Chanquoy, 2001; Galbraith, 2009; Hayes, 1996; Hayes & Flower, 1980; Kellogg, 1994). Likewise, theories of creativity indicate that success can be, at least in part, explained by the use of novel ideas that are appropriate for the task at hand. Specifically, creativity theories define successful ideas as being fluent, flexible, original, and elaborated (Feldhusen & Goh, 1995; Snyder et al., 2004). Here, we empirically investigate these concepts by analyzing relations between idea generation in writing samples and human judgments of writing quality. In addition, we explore links between idea generation in essays and the linguistic features produced by writers. The results indicate that student essays that were rated as containing a greater number of ideas that were flexible, original, and elaborated were judged to be higher quality texts. In addition, essays that contained greater use of metaphors and word play were also judged to be of higher quality. However, only two of the features under investigation (*elaboration* and *fluency*) were significant predictors of essay quality in a regression analysis.

Our results also indicate that successful idea generation is strongly predicted by language features in essays, including a greater production of bigrams, the use of more difficult words, semantic but not lexical similarities between paragraphs (but not sentences), fewer word repetitions, and the fewer exemplifications. Indeed, the six variables together accounted for 82% of the variance in the idea generation as judged by expert raters. This finding provides strong promise for the use of automated linguistic tools to assess the amount and quality of idea generation in writing.

The results of our study have important implications for writing theories and, in particular, notions of idea generation within cognitive models of writing. Perhaps the most important finding is the strong link between essay quality and features related to idea generation. Prior to this study, there was little empirical evidence that essay quality and idea generation were related. Though most major cognitive models of writing include idea generation and link it to successful writing, idea generation as defined within these models is relatively broad (e.g., Galbraith, 2009; Hayes, 1996; Hayes & Flower, 1980; Kellogg, 1994). Thus, although the models acknowledge that the number of ideas is important, specific definitions of the types of ideas related to essay quality are typically not provided. This study takes a step in filling this gap, by showing that all elements of idea generation and style found in the analytic rubric (with the exception of *humor*) correlated with essay

quality. Two features related to idea generation (*elaboration* and fluency) showed strong effect sizes, and another feature of idea generation (*flexibility*) showed a medium effect size. In a regression analysis, *elaboration* and *fluency* were significant predictors of essay quality with *elaboration* explaining the majority of the variance (over 95%). This finding indicates that, for essay quality, it is not only the number of ideas found in the essay, but how elaborated the ideas are. This makes sense on an intuitive level because most essays written in the United States (especially independent essays such as those used in this study) are based on a five-paragraph theme (Johnson, Thompson, Smagorinsky, & Fry, 2003; Nunnally, 1991) in which a few ideas (usually three) are elaborated on in depth.

It is somewhat surprising that elements of idea originality were not significant predictors in the regression model, particularly considering their theoretical importance to writing quality (Torrance et al., 1996). Nonetheless, features related to novel ideas did show significant correlations with essay quality and when idea originality scores were combined with other analytic features from the rubric into an idea generation component score, the component score correlated with essay quality. This correlation reported a strong relation between essay quality and the combined feature score, demonstrating that the number of ideas in conjunction with the flexibility, elaboration, originality, and metaphorical use of those ideas did correlate with higher judgments of writing quality.

The linguistic analysis of idea generation scores in this study can provide evidence for the types of linguistic features in essays that lead to higher human scores of idea generation. Specifically, the linguistic analysis indicates that essays judged as containing greater idea generation have more unique word combinations (i.e., a greater number of individual bigrams as compared to a greater number of bigrams). These bigrams likely relate both to the number of ideas generated and the uniqueness of the ideas. We also found that idea generation was linked to the difficulty of the words included in the essays. Specifically, essays judged to have higher idea generation contained more difficult words (i.e., words that are acquired later), while essays judged to have lower idea generation contained easier words. Idea generation was also linked to cohesion, with essays judged to have higher idea generation showing greater similarity between larger chunks of language (i.e., paragraphs), but less overlap of words between sentences and lower repetition of function words (i.e., TTR). Thus, essays judged to have less idea generation tended to repeat words across sentences, a signature of local cohesion (Halliday & Hasan, 1976). In contrast, essays judged to have higher idea generation had greater overlap of ideas across paragraphs, a signature of global cohesion (Crossley & McNamara, 2011). Last, idea generation was

related to the use of exemplifications such as *for instance, for example*, and *in particular*, wherein a greater use of these terms signaled lower idea generation scores. Hence, these explicit exemplification markers are more likely to be used by writers when they have fewer ideas, but are providing more information or examples related to those ideas.

Overall, the findings have important implications for cognitive writing models and our understanding of links between language features and idea generation. The analysis of the human ratings provides evidence that the ability of writers to elaborate on the ideas presented in an essay is a stronger indicator of successful writing than the number of ideas generated in that essay. Thus, our results challenge a number of writing theories that propose that idea generation alone is important. The linguistic analysis indicates that the number of unique bigrams in a text, the use of global but not local cohesion, and the use of sophisticated words are the best predictors of the number of ideas generated.

The findings also have implications of theories of creativity, which have posited that creativity results from producing a greater number of ideas that are flexible (i.e., different from one another), original, and elaborated. Our factor analysis supports these assertions and adds another element that may help explain creativity: the use of metaphors. In our component score analysis, human ratings of metaphors correlated strongly with traditional features of creativity such as flexibility and originality and metaphor scores were combined with these features to produce an idea generation component score. The notion that metaphor may be an important element of idea generation is also supported in theories of critical discourse, which hold that metaphors are essential for the development of new terms and word meanings (Anderson, 1996) and the expression of new ideas (Stern, 1965). Thus, theories of creativity may need to be broadened to include the use of metaphors under the assumption that writers who produce a greater number of metaphors may be more creative.

Our findings also have important implications for writing practice in that they highlight a number of ways in which students could be taught writing strategies intended to increase idea generation. For instance, research shows that teaching students to generate ideas through strategy practices such as freewriting exercises can have beneficial effects of text quality (Crossley & McNamara, 2016; Elbow, 1979; Hinkle & Hinkle, 1990; Reynolds, 1984). The results from our study indicate that teachers could also teach students strategies to elaborate on the ideas generated as well as how to create ideas through the use of metaphors (e.g., via a brainstorming strategy). In addition, our results indicate that idea generation was associated with text cohesion, introducing the possibility of developing idea generation interventions that

are directly linked to text cohesion. For instance, a potentially useful direction may involve instructing students about the importance of using global as compared to local cohesion when connecting ideas across text elements. The findings from this study could lead to the development and testing of instructional interventions that could potentially help students generate a greater number of creative ideas and show students how to link those ideas throughout an essay.

Conclusion

This study provides us with a better understanding of relations among idea generation, linguistic features in an essay, and human judgments of essay quality. The study finds strong links between all these elements, providing us with a better understanding of writing, writing models, theories of idea generation, and the importance of language features in explaining idea generation. Overall, we find that essays that contain greater elaboration and more ideas are scored more highly and that idea generation is best predicted by linguistic features related to the number of ideas, the uniqueness of ideas, and coherence between those ideas at the global level. Certainly, future studies are warranted to further test these findings on a variety of other writing tasks that go beyond independent writing, such as integrated writing, creative writing, or research reports, and in a variety of context besides timed writing. Larger corpora that are more representative of the sample population would also help generalize our findings. Further work is also needed to better understand the connections between metaphor use and idea generation. Last, while the developed rubric led to acceptable interrater reliability in most cases, the analytic feature of originality may require additional training on the part of raters or rewording to ensure reliability. Nevertheless, in light of these limitations, this study takes an important step toward elucidating the interactions between idea generation and human ratings of essay quality, as well as identifying the linguistic elements that are predictive of idea generation.

Appendix

SAT Scoring Rubric

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.

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*).

We present here a description of the score as a guide using the example of *does not meet the set criterion in any way* versus *meets the set criterion in every way*. For example, 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*).

Score	Definition
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

Part	Score
1. Ideas	
1.1 Fluency	1 2 3 4 5 6
The essay contains many unique ideas within the essay.	
1.2 Flexibility	1 2 3 4 5 6
The essay contains a variety of different ideas (e.g., many different categories of ideas).	

(continued)

Part	Score
1.3 Originality	1 2 3 4 5 6
The essay contains ideas that are unique across essays.	
1.4 Elaboration	1 2 3 4 5 6
The essay includes information that expands on the main idea(s) contained in the essay.	
2. Style	
2.1 Humor	1 2 3 4 5 6
The essay attempts to provoke laughter or amusement.	
2.2 Metaphor & Simile (cognitive style)	1 2 3 4 5 6
The essay involves original comparisons that construe entities outside of their content domain(s).	
2.3 Word Play (linguistic style)	1 2 3 4 5 6
The essay includes the use of sounds, meanings, or forms of words that are unexpected or original.	

Note

1. We set different thresholds for holistic and analytic scores because holistic scores are easier to agree on than analytic scores.

References

Alamargot, D., & Chanquoy, L. (2001). Development of expertise in writing. In D. Alamargot & L. Chanquoy (Eds.), *Through the models of writing* (pp. 185-218). Dordrecht, Netherlands: Kluwer.

Alamargot, D., & Fayol, M. (2009). Modelling the development of written composition. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The Sage handbook of writing development* (pp. 23-47). Thousand Oaks, CA: Sage.

Allen, L. K., Snow, E. L., Jackson, G. T., Crossley, S. A., & McNamara, D. S. (2014). Reading components and their relation to writing. *L'Année psychologique/Topics in Cognitive Psychology*, *114*(4), 663-691.

Amabile, T. (2013). Componential theory of idea generation. In E. Kessler (Ed.), *Encyclopedia of management theory* (pp. 135-140). Thousand Oaks, CA: Sage.

Anderson, J. M. (1996). Historical linguistics. In K. Malmkjaer (Ed.), *The linguistics encyclopedia* (pp. 189-216). London, UK: Routledge.

Antes, A., & Mumford, M. (2009). Effects of time frame on creative thought: Process versus problem-solving effects. *Idea Generation Research Journal*, *21*, 166-182.

Babayigit, S., & Stainthorp, R. (2011). Modeling the relationships between cognitive-linguistic skills and literacy skills: New insights from a transparent orthography. *Journal of Educational Psychology*, *103*(1), 169-189.

Baer, J., & McKool, S. (2009). Assessing creativity using the consensual assessment technique. In C. Schreiner (Ed.), *Handbook of assessment technologies, methods, and applications in higher education* (pp. 1-13). Hershey, PA: IGI Global.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.

Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of Learning Disabilities*, *35*, 39-56.

Berninger, V. W., Fuller, F., & Whitaker, D. (1996). A process model of writing development across the life span. *Educational Psychology Review*, *8*, 193-218.

Berninger, V. W., Richards, T., Stock, P., Abbott, R., Trivedi, P., Altemeier, L., & Hayes, J. R. (2009). fMRI activation related to nature of ideas generated and differences between good and poor writers during idea generation. B*ritish Journal of Educational Psychology Monograph Series II*, *6*, 77-93.

Bird, S., Loper, E., & Klein, K. (2009). *Natural language processing with Python*. Sebastopol, CA: O'Reilly Media.

Bourdin, B., Fayol, M., & Darciaux, S. (1996). The comparison of oral and written modes on adults' and children's narrative recall. In G. Rijlaarsdam, H. Van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 121-141). Amsterdam, Netherlands: Amsterdam University Press.

Brown, G. D. A. (1984). A frequency count of 190,000 words in the London-Lund Corpus of English Conversation. *Behavioural Research Methods Instrumentation and Computers*, *16*, 502-532.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977-990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904-911.

Carter, R. (2016). *Language and creativity: The art of common talk* (2nd ed.). New York, NY: Routledge.

Cheung, W. M., Tse, S. K., & Tsang, W. H. H. (2001). Development and validation of the Chinese creative writing scale for primary school students in Hong Kong. *Journal of Creative Behavior*, *35*, 249-260.

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

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Crossley, S. A., Kyle, K., & McNamara, D. S. (in press). The Tool for the Automatic Analysis of Text Cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*.

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., & 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.

Crossley, S. A., Roscoe, R., & McNamara, D. S. (2013). Using automatic scoring models to detect changes in student writing in an intelligent tutoring system. In P. M. McCarthy & G. M. Youngblood (Eds.), *Proceedings of the 26th International Florida Artificial Intelligence Research Society (FLAIRS) conference* (pp. 208-213). Menlo Park, CA: AAAI Press.

Diakidoy, I., & Constantinou, C. (2000-2001). Idea generation in physics: Response fluency and task specificity. *Idea Generation Research Journal*, *13*(3-4), 401-410.

Elbow, P. (1973). *Writing without teachers*. New York, NY: Oxford University Press.

Feldhusen, J., & Goh, B. E. (1995). Assessing and accessing idea generation: An integrative review of theory, research, and development. *Idea Generation Research Journal*, *8*(3), 231-247.

Flower, L. S., & Hayes, J. R. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication*, *31*, 21-32.

Flower, L. S., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*, 365-387.

Galbraith, D. (2009). Writing as discovery. *Teaching and Learning Writing*, *2*(6), 5-26. doi:10.1348/978185409X421129

Galbraith, D., Ford, S., Walker, G., & Ford, J. (2005). The contribution of different components of working memory to planning in writing. *L1 Educational Studies in Language and Literature*, *15*, 113-145.

Gist, M. E. (1989). The influence of training method on self-efficacy and idea generation among managers. *Personnel Psychology*, *42*, 787-805.

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

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, UK: Longman.

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

Hayes, J. R. (2006). New directions in writing theory. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 28-40). New York, NY: Guilford.

Hayes, J. R., & Berninger, V. W. (2010). Relation between idea generation and transcription. In C. Bazerman, R. Krut, K. Lunsford, S. McLeod, S. Null, P. Rogers, & A. Stansell (Eds.), *Traditions of writing research* (pp. 166-180). New York, NY: Routledge.

Hayes, J., & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3-30). Hillsdale, NJ: Lawrence Erlbaum.

Hinkle, S., & Hinkle, A. (1990). An experimental comparison of the effects of focused freewriting and other study strategies on lecture comprehension. *Teaching of Psychology*, *17*, 31-35.

Johnson, T. S., Thompson, L., Smagorinsky, P., & Fry, P. G. (2003). Learning to teach the five-paragraph theme. *Research in the Teaching of English*, *38*, 136-176.

Kellogg, R. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 355-365.

Kellogg, R. T. (1994). *The psychology of writing*. New York, NY: Oxford University Press.

Kellogg, R. T. (1996). A model of working memory in writing. In M. C. Levy & S. E. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 57-71). Hillsdale, NJ: Lawrence Erlbaum.

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

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Kuperman, V., Stadthagen-Gonzales, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*, *44*(4), 978-990.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757-786.

Levav-Waynberg, A., & Leikin, R. (2012). The role of multiple solution tasks in developing knowledge and idea generation in geometry. *Journal of Mathematical Behavior*, *31*, 73-90.

Majid, D., Tan, A., & Soh, K. (2003). Enhancing children's idea generation: An exploratory study on using the Internet and SCAMPER as creative writing tools. *Korean Journal of Thinking & Problem Solving*, *13*(2), 67-81.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Houndmills, UK: Palgrave Macmillan.

Mayer, R. E. (1999). Fifty years of idea generation research. In R. J. Sternberg (Ed.), *Handbook of idea generation* (pp. 449-458). Cambridge, UK: Cambridge University Press.

McCarthy, P., & 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.

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

McCutchen, D. (2000). Knowledge, processing, and working memory: Implication for a theory of writing. *Educational Psychologist*, *35*, 13-23.

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.

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.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, *38*(11), 39-41.

Muldner, K., & Burleson, W. (2015). Utilizing sensor data to model students' creativity in a digital environment. *Computers in Human Behavior*, *42*, 127-137.

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

Paulus, P. B., & Yang, H.-C. (2000). Idea generation in groups: A basis for idea generation in organizations. *Organizational Behavior and Human Decision Processes*, *82*, 76-87.

Reynolds, M. (1984). Freewriting's origin. *English Journal*, *73*, 81-82.

Runco, M. A., & Okuda, S. M. (1991). The instructional enhancement of the flexibility and originality scores of divergent thinking tests. *Applied Cognitive Psychology*, *5*, 435-441.

Silver, E. (1997). Fostering idea generation through instruction rich in mathematical problem solving and problem posing. *ZDM*, *29*(3), 75-80.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics (Oxford)*, *31*(4), 487-512.

Snyder, A., Mitchell, J., Bossomaier, T., & Pallier, G. (2004). The idea generation quotient: An objective scoring of ideational fluency. *Idea Generation Research Journal*, *16*, 415-420.

Sosik, J., Kahai, S., & Avolio, B. (1998). Transformational leadership and dimensions of idea generation: Motivating idea generation in computer-mediated groups. *Idea Generation Research Journal*, *11*(2), 111-121.

Stern, G. (1965). *Meaning and change of meaning: With special reference to the English language*. Bloomington: Indiana University Press.

Svartvik, J., & Quirk, R. (1980). *A corpus of English conversation*. Lund, Sweden: Gleerup.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York, NY: Teachers College, Columbia University.

Torrance, M., Thomas, G. V., & Robinson, E. J. (1996). Finding something to write about: Strategic and automatic processes in idea generation. In M. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 189-206). Mahwah, NJ: Lawrence Erlbaum.

The British National Corpus. (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Retrieved from http://www.natcorp.ox.ac.uk/ (accessed October, 2013).

van den Bergh, H., & Rijlaarsdam, G. (1999). The dynamics of idea generation during writing: An online study. In M. Torrance, L. Waes, & D. Galbraith (Eds.), *Writing and cognition* (pp. 125-150). Amsterdam, Netherlands: Elsevier.

Witten, I. A., Frank, E., & Hall, M. A. (2011). *Data mining*. San Francisco, CA: Elsevier.

Author Biographies

Scott A. Crossley is an Associate Professor of Applied Linguistics at Georgia State University. His interests intersect natural language process, machine learning, and adult literacy.

Kasia Muldner is an Assistant Professor in the Institute of Cognitive Science at Carleton University. Her research interests include investigating the cognitive and affective aspects that impact learning and creativity.

Danielle S. McNamara is a Professor in Cognitive Science at Arizona State University. Her research interests involve better understanding the cognitive processes relating to comprehension and writing.