Journal of Educational Psychology
2016, Vol. 108, No. 7, 911–924

0022-0663/16/$12.00 http://dx.doi.org/10.1037/edu0000109

The Narrative Waltz: The Role of Flexibility in Writing Proficiency

Laura K. Allen, Erica L. Snow, and Danielle S. McNamara
Arizona State University

A commonly held belief among educators, researchers, and students is that high-quality texts are easier to read than low-quality texts, as they contain more engaging narrative and story-like elements. Interestingly, these assumptions have typically failed to be supported by the literature on writing. Previous research suggests that higher quality writing is typically associated with *decreased* levels of text narrativity and readability. In this study, the authors present the hypothesis that writing proficiency is associated with an individual's *flexible* use of linguistic properties, rather than simply the consistent use of a particular set of linguistic properties. To test this hypothesis, the authors leveraged both natural language processing and dynamic methodologies to capture variability in students' use of narrative style across multiple essay prompts. Forty-five high school students wrote 16 essays across 8 laboratory sessions. Natural language processing techniques were first used to calculate the narrativity of each essay. Random walk and Euclidian distance measures were then used to visualize and classify students' flexibility in narrativity across essays. The results support the hypotheses that students who were flexible in their use of narrativity also wrote essays that were rated as having higher quality, whereas inflexible writers tended to write lower quality essays. Additionally, more flexible writers performed higher than the more inflexible writers on general assessments of literacy and prior knowledge. These results are important for researchers and educators, as they indicate that the link between textual properties and writing quality may fluctuate according to the context of a given writing assignment.

Keywords: writing, flexibility, dynamics, linguistics, individual differences

The study of writing proficiency typically involves the collection of essays that students have written in response to a particular topic, and the subsequent scoring of these essays is based on their linguistic and rhetorical properties. The score that a student receives on this essay is then presumed to serve as a strong proxy for their writing proficiency (Attali & Burstein, 2006). Importantly, however, this essay scoring process is extremely difficult and subjective—even for trained, expert raters—and therefore may not fully capture the construct of writing proficiency (Huot, 1990, 1996; Meadows & Billington, 2005). Accordingly, an important area of research regards *whether* and *how* writing proficiency can be more reliably captured, particularly emphasizing the specific characteristics of both the individual writers and the texts they produce (Crowhurst, 1990; McNamara, Crossley, & McCarthy, 2010; Rafoth & Rubin, 1984; Witte & Faigley, 1981). Findings from such research can inform our theoretical understanding of the writing process (Flower & Hayes, 1981; Hayes, 1996; Kellogg, 2008; McCutchen, 2000; Swanson & Berninger, 1996), as well as the development and automation of writing quality assessments (Attali & Burstein, 2006; McNamara, Crossley, & Roscoe, 2013; McNamara, Crossley, Roscoe, Allen, & Dai, 2015) and pedagogical interventions for struggling writers (Roscoe, Varner, Crossley, & McNamara, 2013; Shermis & Burstein, 2003).

One assumption that is commonly held among educators, researchers, and students is that more *proficient* writers produce texts that are easier to comprehend than less proficient writers. This assumption relies on the notion that narrative text properties, such as events, characters, and personal anecdotes, help authors to gain the attention of their readers and, subsequently, make texts more relatable (Newkirk, 1997, 2012). Indeed, prior research has confirmed that texts with more narrative elements are typically easier to *comprehend* than informational texts (Bruner, 1986; Graesser, Olde, & Klettke, 2002; Haberlandt & Graesser, 1985). Additionally, the degree to which a text is *narrative* as opposed to *informative* is indicative of its readability across a number of domains and grade levels (Graesser, McNamara, & Kulikowich, 2011). Interestingly, however, the link between narrativity and *essay quality* has failed to be supported by prior literature. Although narrative elements may sometimes be associated with high-quality writing (Crossley, Roscoe, & McNamara, 2014), the majority of research on essay quality suggests that higher quality writing is associated with *decreased* levels of text narrativity and measures of readability in general (Crossley, Weston, McLain Sullivan, & McNamara, 2011; McNamara et al., 2013).

One potential explanation for this conflicting evidence lies in the *situational* influence of narrative text elements on writing quality. In other words, it is possible that the frequency of specific linguistic or rhetorical text elements alone is not consistently indicative of essay quality. Rather, these effects may be largely driven by individual differences in students' ability to leverage the benefits of these elements in the appropriate contexts. In this article, we hypothesize that writing proficiency is associated with

This article was published Online First January 18, 2016.

Laura K. Allen, Erica L. Snow, and Danielle S. McNamara, Learning Sciences Institute, Department of Psychology, Arizona State University.

Correspondence concerning this article should be addressed to Laura K. Allen, Learning Sciences Institute, Department of Psychology, Arizona State University, P.O. Box 872111, Tempe, AZ 85287-2111. E-mail: LauraKAllen@asu.edu

an individual's *flexible* use of text properties, rather than simply the consistent use of a particular set of properties. Some researchers have cited flexibility as a characteristic of strong writers (Graham et al., 2012; Graham & Perin, 2007). Graham and Perin (2007), for instance, claimed "proficient writers can adapt their writing flexibly to the context in which it takes place" (p. 9). However, few studies (if any) have empirically tested this claim. In the current study, we address this research gap by investigating how writing proficiency relates to students' flexible use of *narrativity* across multiple essay prompts.

Writing Proficiency

Writing is a complex and demanding activity that requires individuals to coordinate a number of cognitive skills and knowledge sources through the process of setting goals, solving problems, and strategically managing their memory resources (Flower & Hayes, 1981; Hayes, 1996). Importantly, this writing process differs across individuals. Each student brings different strengths and weaknesses to a given writing task, and these variables interact to affect their unique writing processes, as well as the strategies and procedures they utilize to produce effective writing. Individual differences can encompass a broad range of characteristics, from students' degree of prior knowledge (e.g., word and content knowledge) to their daily and overall affect (e.g., their motivation to succeed). Indeed, many models of writing proficiency attempt to account for the influence of individual differences among students, such as knowledge, skill, and working memory capacity (e.g., Kellogg, 2008; McCutchen, 2000; Swanson & Berninger, 1996).

One important difference between skilled and less skilled writers is their level of reading comprehension skill. Reading and writing are tightly connected cognitive processes (Allen, Snow, Crossley, Jackson, & McNamara, 2014; Fitzgerald & Shanahan, 2000; Shanahan & Tierney, 1990; Tierney & Shanahan, 1991); therefore, students who are better at comprehending texts (as well as those who read more frequently) also tend to be better at generating high-quality texts. Similarly, writing proficiency can be influenced by differences in students' vocabulary knowledge (Allen, Snow, Crossley et al., 2014; Graham & Perin, 2007). Students who have access to a greater number of vocabulary words have a greater number of options regarding how they convey ideas.

Strong writers also differ from weak writers in their knowledge of the writing process, including their understanding of writing goals and strategies. For example, Saddler and Graham (2007) found that less skilled writers demonstrated a weaker understanding of writing goals ($d = -1.13$), were less knowledgeable of the differences between strong and poor writing ($d = -.98$), and had less knowledge of efficient writing strategies ($d = -1.10$). Additionally, these less skilled writers wrote lower quality and shorter essays.

Finally, individual differences in prior world knowledge may influence writing proficiency (McCutchen, 1986; Olinghouse, Graham, & Gillespie, 2015). Olinghouse and colleagues (2015), for instance, recently examined the role of discourse and topic knowledge in the quality and characteristics of fifth grade students' stories, persuasive essays, and informational text. The results of this study suggested that discourse and topic knowledge were important elements of young students' writing skills. Specifically, they found that each of the two forms of knowledge made unique, significant contributions to a prediction of writing quality. These results are important, as they indicate that variability in knowledge can influence the quality of a written text. This is important, particularly in the context of persuasive essay writing, because students who know more about the world can, theoretically, develop stronger arguments, as they have greater access to supporting examples and evidence.

Linguistic Properties of High-Quality Writing

Many of these characteristics of skilled writers (e.g., strong reading comprehension skills, etc.) are directly related to their production of specific linguistic properties in essays (Deane, 2013). In particular, more sophisticated linguistic text properties (e.g., cohesion, complex syntax) are related to higher cognitive functioning. Thus, their presence in an essay is indicative of a student's ability to more easily produce complex text, which allows them to place a greater focus on higher level rhetorical and conceptual text properties (Deane, 2013). To this end, many researchers have sought to identify the linguistic properties that relate to high-quality writing (e.g., Applebee, Langer, Jenkins, Mullis, & Foertsch, 1990; Crossley, Roscoe, McNamara, & Graesser, 2011; Ferrari, Bouffard, & Rainville, 1998; McNamara et al., 2010; Varner, Roscoe, & McNamara, 2013; Witte & Faigley, 1981). In these studies, trained, expert human raters typically score essays based on a standardized rubric (e.g., the SAT rubric). The essays are then analyzed for specific linguistic properties, either using computational text analysis tools or human coding. Finally, statistical techniques (e.g., regression analyses, ANOVAs, discriminant function analyses) are employed to determine whether there are specific linguistic properties that systematically relate to these human judgments of essay quality.

These previous analyses have provided critical information about the linguistic properties of high-quality writing (particularly in the context of academic essays; Applebee et al., 1990; Crossley et al., 2011; Ferrari et al., 1998; McNamara et al., 2010; Witte & Faigley, 1981). For instance, skilled writers tend to produce longer essays (Crossley, Weston, et al., 2011; Ferrari et al., 1998; Haswell, 2000; McNamara et al., 2010; McNamara et al., 2013) that contain fewer spelling and grammar errors (Ferrari et al., 1998). At the word level, more proficient writers (i.e., writers that produce higher quality essays and writers in higher grades) use longer words (Haswell, 2000) that are less frequent and concrete, but are more abstract (Crossley, Weston, et al., 2011; McNamara et al., 2010; McNamara et al., 2013). Similarly, previous research has demonstrated that more advanced writers produce essays that contain more complex sentence structures (McCutchen et al., 1994). Haswell (2000), for instance, reported that advanced writers produced essays that contained longer sentences and clauses, and McNamara and colleagues (2010) reported that higher quality essays contained sentences that had a greater number of words before the main verb phrase (i.e., more complex sentence structures).

Finally, specific rhetorical and stylistic text properties have been associated with higher quality essays. Past studies have found that human ratings of essay quality tend to be negatively related to the frequency of narrative text properties, but positively related to the number of rhetorical structures that focus on contrasted ideas, explicitly stated arguments, conditional structures, and reported speech (Crossley, Weston, et al., 2011; McNamara et al., 2013). Overall, previous research studies reveal that more sophisticated writers (defined by both essay scores and higher grade levels) tend to produce essays that are longer and contain properties that are more indicative of sophisticated lexical, syntactic, and rhetorical choices.

Situational Variability of Writing Quality

Recently, researchers have noted that the text properties associated with essay quality often vary across different raters, authors, assignments, and contexts (e.g., Allen, Snow, & McNamara, 2014; Crossley, Allen, & McNamara, 2014; Crossley, Varner, & McNamara, 2013; Crossley, Varner, Roscoe, & McNamara, 2013; Crossley, Weston, et al., 2011; Varner et al., 2013). Crossley and colleagues (2014), for instance, argued that high-quality essays can take on a number of different forms—in other words, these essays can range quite broadly in their combinations of linguistic properties. To investigate this argument, they employed a cluster analysis approach for the purpose of identifying multiple linguistic *profiles* of successful essays. Their analysis revealed four distinct profiles of successful writers, which were linguistically distinct from one another. They argued that these results provided evidence that successful writing cannot be simply defined by one set of predefined linguistic properties—rather, successful writing can manifest in a number of different ways.

Our hypothesis is that writing proficiency is related (at least in part) to students' *sensitivity* to these different writing styles and, consequently, their ability to flexibly adapt the properties of their essays according to the specific context of the writing task. Writing proficiency, in other words, is partially characterized by an individual's ability to assess the context of their writing task and flexibly call upon various linguistic tools, given their knowledge of the constraints and demands of that surrounding environment. For example, if a writer has a strong degree of prior knowledge about the topic for a particular writing assignment, they may not need to employ narrative, story-like properties in order to persuade the reader to take their side on a given argument. On the other hand, if writer is presented with a topic on which they know few explicit facts, they might leverage these narrative story elements for the purpose of engaging their readers and eliciting emotional reactions. Writers in both of these examples could potentially develop successful essays (e.g., they might persuade their readers to take a particular side on an argument); however, the two essays would be composed of vastly different writing styles.

Here, we define writing flexibility as an individual's ability to adapt specific components of their writing in order to craft more effective text. Our argument is that quality texts should not be assessed using a one-size-fits-all formula; rather, successful text communication will depend on a large number of contextual factors, such as the prior knowledge and motivations of the writer and the audience, as well as specific characteristics of the assignment. Importantly, these characteristics of the writing task interact with each other to impact the demands of a particular writing assignment. Thus, writers must assess each writing task on an individual basis to determine the most appropriate strategies and approaches for completing an assignment. In this vein, we argue that more proficient writers will exhibit flexibility in their writing styles across different writing assignments. Our proposal in this article is that we can measure linguistic flexibility (i.e., the degree to which individuals vary their linguistic style across multiple essays) to serve as a proxy for this broader notion of writing flexibility.

Current Study

The goal of the current study is to test the hypothesis that better writing is associated with increased flexibility of writing style, rather than only a set of static linguistic characteristics. This concept of "flexible" writers is in direct contrast to writers who use a fixed set of linguistic properties within the majority of their essays—in other words, they are *inflexible*. There have been mixed empirical findings regarding the relationship between text *narrativity* (and readability, more broadly) and essay quality. In this study, we suggest that this may result, in part, from the various demands of the writing assignment. In other words, different writing prompts and assignments may call on different skills and knowledge sources, which can differentially affect the writing strategies and processes engaged by individuals. Thus, we additionally suggest that this flexibility in writing style may result as a function of individual differences related to literacy skills, such as vocabulary knowledge, comprehension ability, and prior world knowledge. Our primary research questions are:

1. How is writing proficiency related to students' flexible use of narrativity?

2. How does this flexible use of narrativity vary as a function of individual differences among students?

We first hypothesize that greater writing proficiency will be positively associated with students' linguistic flexibility across the essays. In particular, we hypothesize that students who vary in their use of narrative language across multiple essays will also produce essays that are rated as higher quality texts.

Second, we hypothesize that this measure of narrative flexibility will vary as a function of individual differences among the students. This hypothesis follows from the assumption that writing flexibility is a strategic behavior that relates to students' literacy abilities and prior knowledge of a given topic. Thus, students who have developed strong literacy skills will be more likely to assess when it is appropriate to employ specific linguistic and rhetorical devices within individual writing assignments.

This study combines both natural language processing and dynamical techniques to characterize the degree to which students vary in their use of narrativity across 16 timed, argumentative, prompt-based essays. Thus, writing flexibility is measured here in a very specific context. We chose to specifically focus on the narrativity within the essays because of the previously mixed empirical findings regarding the construct of narrativity in text quality. Crossley and colleagues (2014), for instance, found that one profile of high-quality writing related to a more narrative, story-like style, whereas a separate profile of essays (of equally high quality) were related to more informative, academic text. Thus, an important research question is whether more proficient writers are able to leverage the benefits of both narrative and informative styles according to the demands of specific writing assignments. For instance, one skilled writer might have little fact-based domain knowledge with which to develop evidence on a particular prompt. Therefore, this writer might construct an essay that relies on personal anecdotes and descriptions that are engaging to the reader. On the other hand, another skilled writer might rely more heavily on fact-based evidence to answer the prompt. In this essay, the writer would use facts to argue a particular perspective on the prompt question. In both scenarios, the resulting essays are high quality and successfully able to argue a particular point to the reader. However, the two writers simply used different strategies to achieve this goal.

An additional note is that this study solely focuses on timed, prompt-based essays. Although we argue that this investigation of narrativity is important across a number of different writing genres, we chose to focus our initial analysis on this genre because these essays do not require prior content knowledge of a particular domain. This allows us to more easily tease apart our results in terms of their relationship to writing proficiency, rather than greater knowledge of a particular domain.

Methods of Automated Text Analysis

To address our research questions, we use a combination of natural language processing and dynamic methodologies to examine students' use of narrativity across multiple argumentative essays. Text narrativity is a key component of text readability; therefore, it provides a strong foundation on which to build an understanding of the relations between text readability and essay quality. In this study, we chose to leverage automated text analysis tools to provide a measure of text narrativity. Automated indices provide a quick and reliable alternative to the subjective coding of essays by humans.

Automated measures of text readability and narrativity. In the current study, we employed Coh-Metrix (McNamara & Graesser, 2012; McNamara, Graesser, McCarthy, & Cai, 2014) to automatically assess the degree to which students' essays were more *narrative* or *informative.* The principal method for automatically measuring text difficulty is the use of standardized "readability" formulas (Hiebert, 2002). These formulas provide a single metric by which the relative syntactic and semantic difficulty of texts can be compared. One of the most common readability formulas is the Flesch-Kincaid Grade Level (FKGL; Kincaid, Fishburne, Rogers, & Chissom, 1975), which calculates word and sentence length to determine text difficulty. This score is a single index that maps onto the grade levels in the U.S. school system. Unidimensional measures, such as FKGL, can simplify the text assignment process by providing teachers a single metric to select grade-appropriate texts for their students.

Despite their simplicity, traditional readability formulas lack the sophistication needed to represent the multiple levels of text difficulty. One problem is that these formulas typically measure the surface-level characteristics of texts, which are solely predictive of students' superficial text comprehension (i.e., their understanding of the individual words and sentences; Davison, 1984). Most contemporary models of reading comprehension suggest that there are multiple levels of understanding that contribute to the comprehension process (Graesser & McNamara, 2011). However, standard readability formulas often fail to identify the text characteristics that impact students' understanding at deep levels (e.g., deep cohesion). Further, they provide teachers little guidance on how to diagnose and remediate students' difficulties. In particular, they give no information on which text properties may be challenging or helpful to individual students.

Coh-Metrix (McNamara & Graesser, 2012; McNamara et al., 2014) is a computational text analysis tool that was developed, in part, to provide stronger measures of text difficulty (Duran, Bellissens, Taylor, & McNamara, 2007). This tool analyzes texts at the word, sentence, and discourse levels; thus, it can potentially offer more information about the specific challenges and linguistic scaffolds contained in a given text. Previous work with Coh-Metrix suggests that multiple dimensions coordinate within texts to affect subsequent comprehension performance (McNamara, Graesser, & Louwerse, 2012). To account for these multiple text dimensions, Graesser and colleagues (2011) developed the Coh-Metrix Easability Components. These components offer a detailed glance at the primary levels of text difficulty and are well aligned with an existing multilevel framework (Graesser & McNamara, 2011).

Narrativity. The degree of narrativity versus informational content provided within an essay is assessed using the narrativity component score provided by Coh-Metrix (Graesser et al., 2011; McNamara, 2013). The narrativity of a text reflects the degree to which a story is being told, using characters, places, events, and other elements that are familiar to readers. This measure is highly related to the use of familiar words, greater world knowledge, and oral language style. Combining many narrative elements within a text can be used to sustain readers' attention by creating uncertainty, excitement, or building suspense (Barab, Gresalfi, Dodge, & Ingram-Goble, 2010; Cheong & Young, 2006; Vorderer, Wulff, & Friedrichsen, 1996). Additionally, narrativity allows readers to connect and comprehend action sequences, making it easier to keep track of main characters, plot points, and cause-and-effect relationships (Bruner, 1986; Schank & Abelson, 1995). The degree to which a text is narrative is strongly associated with word familiarity, world knowledge, and oral language.

Because of their engaging and familiar properties, highly narrative texts are considerably easier to read, comprehend, and recall than informative texts (Graesser & McNamara, 2011; Haberlandt & Graesser, 1985). Within the context of essay writing, however, the role of narrativity is less clear. Persuasive essays written with lower degrees of narrativity are typically rated as having higher quality (as judged by expert human raters who use standardized rubrics) than more narrative essays (although not consistently), include more content words (e.g., nouns), and discuss more unfamiliar topics. The use of facts and data as evidence in an essay (as opposed to, e.g., personal anecdotes) is associated with more refined rhetorical strategies on the part of the writer, which may serve to explain negative correlations between narrativity and essay scores.

The narrativity component score is calculated in Coh-Metrix based on the results of a previous, large-scale corpus analysis (Graesser et al., 2011). In this study, the Touchstone Applied Science Associates (TASA) corpus was used to provide a representative sample of the types of texts that are commonly seen from kindergarten through 12th grade. This corpus consists of 37,520 texts (average of 288.6 words per text, $SD = 25.4$) that have been classified according to genre and assigned an appropriate grade level. To develop the narrativity score (and the other Easability components), Graesser and colleagues (2011) first used Coh-Metrix to analyze the linguistic characteristics of the texts in the

TASA corpus (53 measures were used; see Graesser et al., 2011, for more specific information about these indices). These indices ranged from basic word level information (e.g., word frequency) to higher level information about semantic text cohesion. A principal component analysis (PCA) was conducted to reduce these indices to a smaller number of dimensions. The Coh-Metrix measures converged on the PCA with eight principle component scores, accounting for 67.3% of the variability among the texts.

The narrativity Easability Component score consists of 17 Coh-Metrix indices, with loadings ranging from 0.53 to 0.92. These indices provide critical information about the differences between narrative and informational texts. First, narrative texts include more descriptions of actions and events; thus, the narrativity Easability Component assigns its scores (in part) based on the notion that more narrative texts contain more main verbs, adverbs, and intentional events, actions, and particles. Informational texts, on the other hand, are characterized by more unfamiliar content words, often in the form of nouns. An additional characteristic of narrative texts is that they share many characteristics of oral language (Biber, 1988), as evidenced by the increased frequency of familiar words and pronouns in the narrativity Easability Component, as well as the use of simpler sentence constructions.

The resulting narrativity Easability Component score is calculated in the form of a percentile score (ranging from 0% to 100%), with higher scores indicating that the text is more narrative than informative (and likely easier to read) than other texts in the TASA corpus. For instance, a percentile score of 85% means that 85% of the texts in the TASA corpus are likely more difficult than the particular text (at least in terms of its narrativity), and 15% are likely easier to read. Overall, the Coh-Metrix narrativity Easability Component score can serve as a measure of text readability, specifically regarding the degree of story-like elements that are present within an individual text.

Dynamic Analyses

In the current study, we use dynamic systems theory and its associated analysis techniques to analyze the *flexible* relations between the narrative properties of essays and students' writing proficiency. Dynamic methodologies offer researchers a means with which they can characterize patterns that emerge from students' behaviors or interactions (e.g., writing, dialect, or choices) during a learning task. Unlike more traditional statistical measures, dynamic methodologies place a strong emphasis on the role of time in the assessment of behavioral patterns and change. In other words, dynamic analyses focus on the individual fluctuations that occur across time, as opposed to treating behavior as a static (i.e., inflexible) process, as is customary in many traditional statistical approaches (i.e., self-reports). Dynamic methodologies can, therefore, help to contextualize students' behaviors and offer educators and researchers a means of capturing important fine-grained patterns across time.

Although the current study is one of the first to use dynamic analyses to assess *writing flexibility*, these techniques have previously been used across a wide variety of domains as a means to understand the complex patterns that manifest in individuals' behaviors over time (Snow, Allen, Russell, & McNamara, 2014; Snow, Likens, Jackson, & McNamara, 2013; Soller & Lesgold, 2003; Zhou, 2013). Here, we utilize two dynamic methodologies—*random walks* and *Euclidian distances*—to visualize and classify the extent to which students demonstrate a flexible use of narrative properties across time. Random walks are mathematical tools that are used to *visualize* fine-grained patterns that emerge in categorical data over time (Nelson & Plosser, 1982; Snow et al., 2013). Researchers have used this technique in a variety of domains, such as psychology (Allen, Snow, & McNamara, 2014; Collins & De Luca, 1993), genetics (Lobry, 1996), ecology (Benhamou & Bovet, 1989), and the learning sciences (Snow et al., 2013). For example, geneticists have utilized random walk analyses to investigate how patterns of disease form within gene sequences (Arneodo et al., 1995; Lobry, 1996), and learning scientists have used this methodology to visualize how students' choice patterns within computer-based learning environments vary as a function of their prior skills (Snow et al., 2013).

In order to validate the visualizations offered by these random walk analyses, researchers need to *quantify* these fine-grained patterns of behavior. Euclidian distance analyses offer a metric that is embedded within the random walks that can quantify students' fluctuations as they unfold over time (Allen, Snow, & McNamara, 2014). In this calculation, Euclidian distances for each "step" or movement within a random walk analysis are used to create a distance time series. This time series serves as a quantification for the movements in the categorical patterns visually represented in the random walk.

Method

Participants

The data presented here were collected as part of a larger study ($n = 86$), which compared the Writing Pal intelligent tutoring system (ITS) to an Automated Writing Evaluation (AWE) system (Allen, Crossley et al., 2015; Allen, Crossley, Snow, & McNamara, 2014; Crossley, Varner, Roscoe, et al., 2013; Roscoe & McNamara, 2013). In this study, we focus on the participants who engaged with the AWE system ($n = 45$). All participants were high school students recruited from an urban environment located in the southwestern United States. These students were, on average, 16.4 years of age, with a mean reported grade level of 10.5.

Of the 45 students, 66.7% were female and 31.1% were male. Students self-reported ethnicity breakdown was as follows: 62.2% were Hispanic, 13.3% were Asian, 6.7% were Caucasian, 6.7% were African American, and 11.1% reported "other." All students were recruited from local high schools and publically posted flyers. These students received $10.00 for their participation in each session of this experiment. Additionally, the students' money was doubled for completing all 10 of the sessions. Thus, the participants in this study each received $200 for their participation.

Study Procedure

The current study was a 10-session experiment that lasted approximately three weeks. During the first session, students completed a pretest that contained measures of writing ability, prior knowledge, reading ability, and literacy skills. Training occurred during the following eight sessions, in which students engaged with the AWE system. During Session 10, students completed a posttest, which contained measures similar to the pretest. Previous analyses have indicated that students increased their essay quality, motivation, perceptions of improvement, and self-assessment accuracy across the training sessions (for more thorough information on the results of the training study, see Allen, Crossley, et al., 2015).

Pretest. During Session 1, students completed a pretest that lasted approximately one hour in duration and contained a battery of individual difference measures. These measures included demographics, prior knowledge test, writing proficiency (25-min SAT-style essay), and literacy skills.

Training. During training (Sessions 2 to 9), students practiced writing 25-min timed essays on SAT-style prompts. During each of the eight training sessions students wrote and revised two timed essays (i.e., 16 essays). Upon completion of each essay, the AWE system provided students with automated formative feedback. After students examined this feedback they were allotted 10 min to revise their essay based on the feedback presented.

Posttest. During Session 10, all participants completed a posttest. The posttest comprised measures similar to the pretest, including a writing proficiency test (25-min SAT-style essay).

Materials and Measures

Prior reading ability. Students' reading ability was assessed using the Gates-MacGinitie reading skill test (4th ed.; MacGinitie & MacGinitie, 1989). This 48-item multiple-choice test assessed students' reading comprehension ability by asking students to read short passages and then answer two to six questions about the content of the passage. These questions were designed to measure both shallow- and deep-level comprehension. All students were given standard instructions, which included two practice questions. This test was a timed task that gave every student 20 min to answer as many questions as possible. The Gates-MacGinitie Reading Test is a well-established measure of student reading comprehension, which provides information about students' literacy abilities (α = .85–.92; Phillips, Norris, Osmond, & Maynard, 2002).

Vocabulary knowledge. Students' vocabulary knowledge was assessed using the Gates-MacGinitie vocabulary test (4th ed.; MacGinitie & MacGinitie, 1989; see previous section for reliability). This test includes 45 simple sentences, each with an underlined vocabulary word. Students are asked to read the sentence and choose the word most closely related to the underlined word within the sentence from a list of five choices. All students' were given standard instructions, which included two practice questions. This test was a timed task that gave every student 10 min to answer as many questions as possible.

Prior knowledge. Students' prior science knowledge was assessed using a 30-item measure of prior knowledge designed for use with high school students. This task has been used previously in work related to reading comprehension and strategy skill acquisition (Roscoe, Crossley, Snow, Varner, & McNamara, 2014). The 30-item multiple-choice measure assesses students' knowledge in the areas of science, literature, and history. The test shows high reliability, with α ranging from .72 to .81. The measure is a modified version of a knowledge assessment used in several studies and validated with over 4,000 high school and college students (McNamara, O'Reilly, Best, & Ozuru, 2006; O'Reilly, Best, & McNamara, 2004; O'Reilly & McNamara, 2007; O'Reilly, Taylor, & McNamara, 2006). This version of the assessment was developed in prior work by including items with moderate difficulty (i.e., 30%–60% of students could answer correctly) that were correlated with individual difference measures (e.g., reading skill) and performance on comprehension tests. Additional items were obtained from high school textbooks. In this process, 55 multiple-choice questions (i.e., 18 science, 18 history, and 19 literature) were piloted with 15 undergraduates to test item performance. Thirty questions (10 per domain) were selected such that no items selected exhibited either a ceiling (>.90) or floor effect (<.25, chance level). Examples are provided in Table 1.

Pretest and posttest essay quality. Students writing proficiency was assessed at both pretest and posttest through the use of timed (25-min) and counterbalanced SAT-style essays (the two essay prompts can be found in the Appendix). The pretest and posttest essays were assessed on a 6-point scale by two independent expert human raters. These raters had previous experience scoring academic essays and were compensated for their time. Additionally, they were college composition instructors with at least three years of experience teaching writing. The holistic rating scale was developed in order to assess the quality of each essay on a scale from 1 to 6.[1] The raters were given specific instruction on this rubric and given example essays for each score in the rubric (i.e., they were given an example of an essay that had received a score of "1" and another essay that had received a score of "2"). Additionally, they were told that the distance between each score was equal (i.e., a score of 5 is as far above a score of 4 as a score of 3 is above a score of 2). After receiving instruction on the rubric, the raters practiced using the rubric on a sample set of SAT style essays written on the same prompts as the essays in the current study. The raters were expected to continue with practice until their interrater reliability reached a correlation of r = .70. After the raters had reached an interrater reliability of r = .70, each rater then evaluated the entire set of essays. Thus, each essay received two essay scores. Once these ratings were collected, differences between the raters' scores were calculated. All score differences between the raters were less than 2 (i.e., the raters demonstrated 100% adjacent agreement with the final set). Thus, holistic scores for pretest and posttest essays were calculated by averaging the scores between raters. For the final set, the raters demonstrated a 57% exact accuracy and a 100% adjacent accuracy. Additionally, the raters' final essay scores were significantly correlated, $r = .55$, $p < .001$.

Training essay performance. Training performance in this study was defined as students' average essay score across the 16 essays that were composed in the AWE system. All of the essays that students wrote in this AWE system were timed, SAT-style essays, with prompts that were similar to those given at pretest and posttest (for a list of the prompt topics and the order they were assigned, see Table 2). To score these essays, we used a previously developed algorithm to assign holistic writing scores to these written essays. The algorithm uses variables from Coh-Metrix, the Writing Assessment Tool and Linguistic Inquiry and Word Count (Pennebaker, Booth, & Francis, 2007) to assign essay scores on a scale from 1 to 6. These indices range from word-level properties of the essays, such as the number of infinitives, to higher level

[1] For a copy of the SAT rubric, see http://sat.collegeboard.org/scores/sat-essay-scoring-guide.

Table 1
Examples of Questions and Answers in Prior Knowledge Assessment

Domain	Question and answer choices
Science	The poisons produced by some bacteria are called . . . (a) antibiotics, (b) toxins, (c) pathogens, (d) oncogenes
History	A painter who was also knowledgeable about mathematics, geology, music, and engineering was . . . (a) Michelangelo, (b) Cellini, (c) Titian, (d) da Vinci
Literature	Which of the following is the setting used in *The Great Gatsby* . . . (a) New York, (b) Boston, (c) New Orleans, (d) Paris

properties, such as the semantic similarity of the paragraphs within the essay. The algorithm was developed using correlation and discriminate function analyses to categorize 1,243 student essays that had been previously scored by expert human raters. The resulting models reported exact matches between the human scores and the predicted essay scores with 55% accuracy. Additionally, the models reported 92% accuracy for adjacent matches (see McNamara et al., 2015, for a more thorough description of the algorithm used in this study).

Assessment of narrative flexibility. We used random walk analyses to investigate the flexibility of students' use of narrativity across time. Random walk analyses are mathematical tools that are used to provide visual representations of patterns in categorical data as they manifest across time (Benhamou & Bovet, 1989; Lobry, 1996; Nelson & Plosser, 1982; Snow et al., 2013). In the current study, we first used Coh-Metrix to compute a *narrativity percentile score* (range from 0 to 100) for each essay. We then used this narrativity percentile score to classify each essay into four orthogonal categories (see Table 3). This classification was organized based on the degree of narrativity present in each essay (using the percentile score provided by Coh-Metrix). Each orthogonal category was then assigned to a vector that fell along a basic scatterplot. Therefore, if an essay received a narrativity score below 25%, this essay was assigned to the vector (−1, 0), whereas an essay that received a score that was greater than 75% narrative was assigned to the vector (0, −1). Once each essay had been assigned to a vector, we calculated a random walk for each student that began at the origin of the scatterplot (0, 0). For each subsequent essay that a student wrote, the walk would "step" in the direction that was consistent with the assigned vector. The resulting walk would represent each student's use of narrativity across the 16 training essays.

Figure 1 provides an example of what a random walk might look like for a student who wrote four training essays. All walk sequences begin at the origin of the scatterplot (see #0 in Figure 1). The first essay written by the student was low in narrativity (i.e., narrativity percentile score <25%); thus, the walk takes a step left along the *x*-axis (see #1 in Figure 1). The second essay written by the student received a narrativity percentile score between 25% and 50%; this means that the walk takes a step up along the *y*-axis (see #2 in Figure 1). The student wrote a third essay that had a narrativity percentile score between 50% and 75% narrativity. Therefore, the walk takes a step to the right along the *x*-axis (see #3 in Figure 1). The fourth and final essay written by the student received a narrativity percentile score between 25% and 50%, which again makes the walk step up along the *y*-axis (see #4 in Figure 1). These rules were used to generate a unique random walk for each of the 45 students, which represented the fluctuations in their use of narrativity across the 16 essays that were written in the AWE system.

Figures 2 and 3 illustrate two random walks that were generated using two students' actual training essays from the current study. These walks represent students' degree of "narrative flexibility" across the training essays.

Figure 2 illustrates the walk of a student who wrote highly narrative (above 75 narrativity percentile score) essays across each of the training essay assignments. In other words, regardless of the writing prompt, this student employed the same range of narrativity throughout all of her essays. On the other hand, the walk depicted in Figure 3 comes from a student who was highly flexible in the use of narrativity across the 16 essays. As the various factors varied from essay to essay (e.g., the essay prompt), this student employed varying degrees of narrativity to develop arguments and ideas.

Table 2
Writing Pal Essay Prompt Order

Session	Essay prompts
Session 2	Planning: Does every individual have an obligation to think seriously about important matters?
	Originality: Can people ever be truly original?
Session 3	Winning: Do people place too much emphasis on winning?
	Loyalty: Should people always maintain their loyalties, or is it sometimes necessary to switch sides?
Session 4	Patience: Is it better for people to act quickly and expect quick responses from others rather than to wait patiently for what they want?
	Memories: Do personal memories hinder or help people in their effort to learn from their past and succeed in the present?
Session 5	Heroes: Should we admire heroes but not celebrities?
	Choices: Does having a large number of options to choose from increase or decrease satisfaction with the choices people make?
Session 6	Perfection: Do people put too much importance on getting every detail right on a project or task?
	Optimism: Is it better for people to be realistic or optimistic?
Session 7	Uniformity: Is it more valuable for people to fit in than to be unique and different?
	Problems: Should individuals or the government be responsible for solving problems that affect our communities and the nation in general?
Session 8	Beliefs: Are widely held views often wrong, or are such views more likely to be correct?
	Happiness: Are people more likely to be happy if they focus on their personal goals or on the happiness of others?
Session 9	Fame: Are people motivated to achieve by personal satisfaction rather than by money or fame?
	Honesty: Do circumstances determine whether or not we should tell the truth?

Table 3
Narrativity Classification and Vector Assignment

Essay narrativity level	Axis direction assignment
Less than 25% narrativity	−1 on x-axis (move left)
Between 25% and 50% narrativity	+1 on y-axis (move up)
Between 50% and 75% narrativity	+1 on x-axis (move right)
Greater than 75% narrativity	−1 on y-axis (move down)

Euclidian distance measure. The random walks described in the Assessment of Narrative Flexibility section provide *visualizations* of the fluctuations in students' use of narrativity across time. To *quantify* these changes in students' writing patterns, distance time series were calculated for each student using Euclidian distance measures. This measure calculated the distances between the origin of the scatterplot (0, 0) and each step in the walk (see Equation 1). In this equation, y represents the current position of the particle (the end point of the walk) on the y-axis, x represents the particle's position on the y-axis *and* i represents the ith "step" in the walk.

$$\text{Distance} = \sqrt{(y_i - y_0)^2 + (x_i - x_0)^2} \quad (1)$$

After calculating the Euclidian distance of the steps in each walk, an average Euclidian distance score was calculated for each student's entire walk. Broadly, this measures how far each student "walked" from the origin of the scatterplot across the 16 essays. This resulting distance measure (i.e., a student's *narrative distance score*) was used to represent students' flexibility in their use of narrativity. If a student, for example, employed the same degree of narrativity across all 16 training essays, that student would travel further from the origin, resulting in a high narrativity distance score (see Figure 2 for a visualization of this type of student). Conversely, if a student varied considerably in the use of narrativity across the essays, the resulting narrative distance score would be lower, as the fluctuations would cause the walk to remain closer to the origin (see Figure 3 for a visualization of this type of student). Overall, students' distance scores provide information about whether they are varied in their writing style (i.e., lower distance scores and more flexible) or whether they tend to remain inflexible (i.e., consistent) across multiple essays (i.e., higher distance scores and inflexible). It is important to note that the directionality of students' random walks does not matter, as the Euclidian distance measure captures how far (in any direction) students' walks move away from the center point.

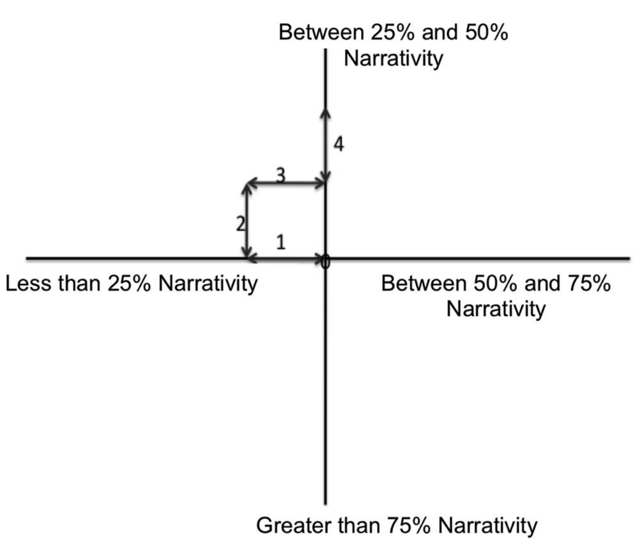

Figure 1. This is an example of a random walk as described in the text.

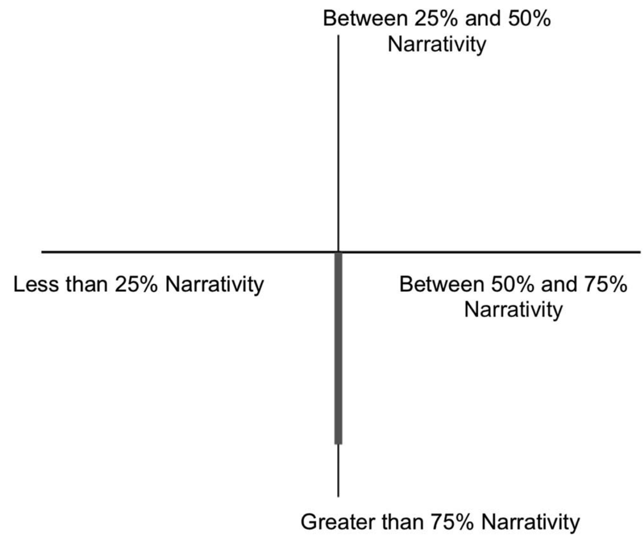

Figure 2. This is an example of a random walk for an inflexible writer.

The random walk and Euclidian distance analyses used in the current study afford researchers the ability to capture flexibility that would otherwise be missed by traditional (i.e., static) metrics. In particular, random walk analyses capture movements as they take place across time. In this sense, we can analogize the narrative flexibility examined in this study to the dancing of the waltz. In the waltz, dancers make multiple movements that result in rotations of the dancers around the floor. Importantly, in the waltz, skilled dancers do not travel across the room in a straight line. Although this

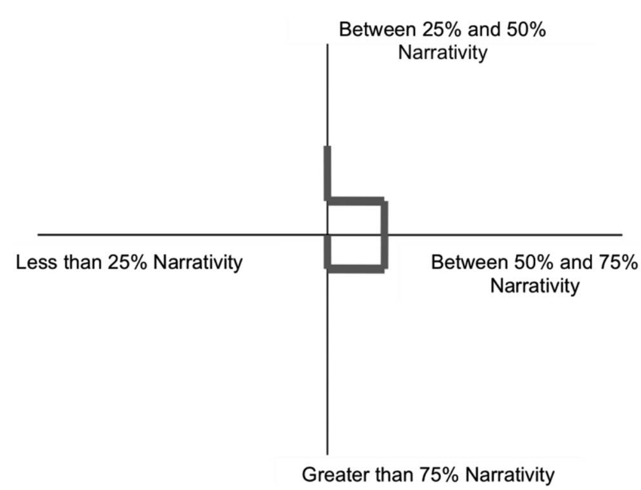

Figure 3. This is an example of a random walk for a flexible writer.

would result in more *efficient* travel, these dancers recognize that in order to perform the dance in the most graceful way, they must make small rotations that result in larger movements across the floor. Additionally, they must make adjustments to their behaviors based on their partner's behaviors, as well as the behaviors of the other dancers on the floor. Thus, in the waltz, the fine-grained steps and patterns of the dancers are important to its overall aesthetics and success. Similarly, we propose that skilled writers will demonstrate more *flexible* patterns of narrativity across their essays. Thus, rather than consistently producing essays of the same style, these writers will flexibly adapt their behaviors to the demands of the prompt (e.g., based on their own prior knowledge, the audience). Related to the random walk analyses, if a student generates essays that vary in their degree of narrativity, the student's random walk will hover around the center point of the *x, y* axis and contain more movements that change directions. In contrast, a student who is less flexible and consistently generates essays with similar levels of narrativity will demonstrate a random walk that moves in one direction and covers a greater distance along the *x*- or *y*-axis.

Statistical Analyses

To assess the degree to which writing quality is associated with students' flexible use of narrativity, we calculated random walks, Euclidian distances, Pearson correlations, and regression analyses. The random walk analyses allowed us to *visualize* students' use of narrativity across their 16 essays. Additionally, this random walk allowed us to calculate a Euclidian distance measure, which reveals students' consistency in their use of narrativity across their 16 essays. Pearson correlations were used to assess the relation between flexibility (as defined by the Euclidian distance measure) and essay quality, as well as individual differences in students' prior global knowledge, prior vocabulary knowledge, and prior reading comprehension ability (see Table 4 for descriptive statistics on these pretest and posttest materials). Finally, regression analyses were conducted to follow-up the correlation analyses in order to provide an indication of the variables that accounted for the most variability in the dependent variables (i.e., essay quality and flexibility).

Results

Random Walks

To visualize and categorize how students varied the narrativity in their writing style, random walk analyses were calculated using the rules described in the previous section (see Table 3) for each student. These walks produced distance measures for each student, which is indicative of how flexible or inflexible the student's use of narrativity was across all 16 essays. Overall, these narrative distance measures suggested that students varied considerably in their narrative flexibility, ranging from a minimum narrative distance score of 2.03 to a maximum narrative distance score of 8.50 ($M = 6.11$, $SD = 1.73$). The narrative distance score for each student in this study is plotted in Figure 4 to provide a visualization of the degree to which students' varied in their flexible use of narrativity across the 16 training essays.

This variation in narrative flexibility was examined according to students' writing proficiency. To provide a coarse visualization of the flexibility differences between the less and more skilled writers, we created a visualization that compared the narrative distance scores for two groups of students (based on a median split on students' pretest essay scores): *less skilled writers* and *more skilled writers.* To confirm that the visualization was depicting two separate groups of students, a between-subjects ANOVA investigated the difference between these less skilled and more skilled writing ability students' narrative distance scores and revealed that more skilled writers had significantly lower narrative distance scores ($M = 5.29$, $SD = 1.47$) compared with less skilled writers ($M = 7.02$, $SD = 1.60$), $F(1, 42) = 14.06$, $p = .001$, $d = 1.13$.

Figure 4 provides an illustration of these differences between less and more skilled writers. In this figure, less skilled writers are represented as black dots and more skilled writers are represented by light gray dots. As shown in this image, the less skilled writers (black dots) traveled further from the origin of the scatterplot (0, 0) than the more skilled writers (light gray dots), who seem to cluster more frequently near the origin. This visualization indicates that the more skilled writers were also the students who were more varied in their use of narrativity across the training essays (i.e., they hovered more around the origin), whereas the less skilled writers traveled much further from the origin and were less flexible in their use of narrativity.

Writing Proficiency

Although the visualization analyses provided preliminary evidence that less and more skilled writers differed in their narrative flexibility, this analysis was based on a median split and, therefore, has potential statistical weaknesses. Median splits pose problems to statistical validity because they create a false dichotomous variable from a continuous variable. Therefore, we conducted further analyses to provide more statistically valid tests of our research questions. Specifically, Pearson correlations were calculated to further assess the validity of these analyses (i.e., to assess the degree to which students' flexible use of narrativity was related to their writing proficiency). We calculated the correlations between students' narrative distance scores and their pretest and posttest essay scores (assessed by the expert human raters), as well as their average scores across the 16 training essays (assessed by the AWE algorithm). Results from these analyses indicated that narrative distance scores were significantly negatively related to the quality of pretest essay scores, $r = -.45$, $p = .002$, and training essay scores, $r = -.47$, $p = .001$. Overall, these results reveal that skilled writers were more flexible in their use of narrativity across the training essays (i.e., they exhibited lower

Table 4
Descriptive Statistics for Pretest and Posttest Materials

Measure	Minimum	Maximum	Mean (*SD*)
Pretest essay score	2.00	4.00	2.80 (.57)
Posttest essay score	2.00	4.50	3.10 (.64)
Reading comprehension[a]	21.00	75.00	47.55 (17.12)
Vocabulary knowledge[a]	13.00	89.00	56.44 (20.20)
Prior knowledge (overall)[a]	27.00	77.00	51.70 (14.54)
Science prior knowledge[a]	20.00	90.00	52.67 (18.02)
History prior knowledge[a]	10.00	100.00	54.00 (22.60)
Literature prior knowledge[a]	10.00	70.00	48.44 (14.92)

[a] Score is based on percentage correct.

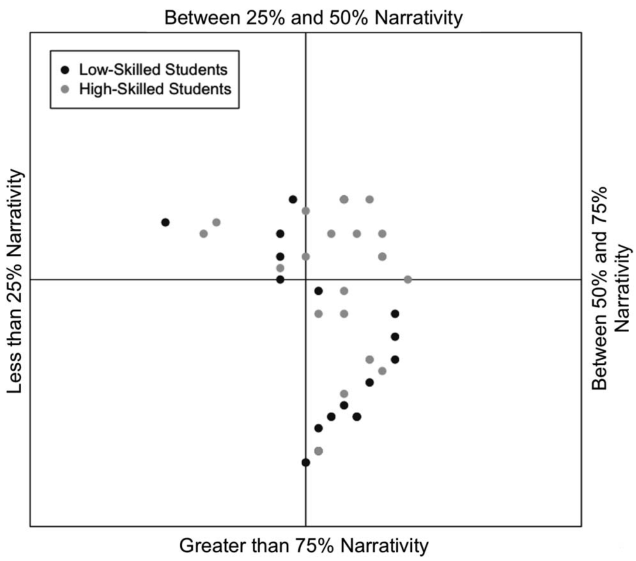

Figure 4. Visualization of less skilled and more skilled students' random walks end points.

narrative distance scores). However, the relation between narrative flexibility and essay scores was no longer present at posttest ($p = .08$). These findings suggest that over the course of persistent writing practice, the relation between flexibility in writing style and essay quality is reduced.

We conducted a stepwise regression analysis with the significant variables as predictors to determine which writing proficiency measures were the most predictive of narrative flexibility, as well as to assess the amount of variance accounted for by these assessments. This analysis yielded a significant model, $F(1, 42) = 11.66$, $p = .001$, $R^2 = .22$, with one variable retained in the final analysis: training essay scores, $\beta = -.47$, $t(42) = -3.41$, $p = .001$. Results of this analysis suggested that students' flexible use of narrativity was most strongly predicted by the quality of the essays that they wrote across the 8 days of writing practice. Thus, students who consistently demonstrated strong writing proficiency were more flexible in their use of narrativity throughout essay writing practice.

Individual Differences

To further investigate the role of narrativity flexibility in the writing process, we examined its relationship with individual differences known to relate to writing proficiency. Specifically, we calculated Pearson correlations and regression analyses between narrative distance scores and students' pretest scores on assessments of prior world knowledge, vocabulary knowledge, and reading comprehension ability. Results of the correlation analyses suggested that the narrative distance scores were significantly related to all of the pretest measures except for prior knowledge in history and literature (see Table 5). These results suggest that narrative flexibility is related to other literacy skills and knowledge sources, rather than solely related to writing proficiency, as it is strongly associated with performance on assessments of prior science knowledge as well as literacy skills.

We conducted a stepwise regression analysis with the significant variables as predictors to determine which individual difference measures were the most predictive of narrative flexibility, as well as to assess the amount of variance accounted for by these assessments. This analysis yielded a significant model, $F(1, 43) = 22.47$, $p < .001$, $R^2 = .34$, with one variable retained in the final analysis: reading comprehension, $\beta = -.59$, $t(43) = -4.74$, $p < .001$. Results of this analysis suggested that students' flexible use of narrativity was most strongly predicted by ability to read and comprehend texts. Thus, students who entered the writing task with more strategies and knowledge about how to comprehend texts may have had a simpler time adapting their writing styles to various prompts, as they were potentially more aware of the processes engaged by their readers, and thus more strategic in their actions (McNamara, 2013).

Discussion

Evidence from the field of writing research largely supports the notion that the linguistic properties of texts are generally indicative of the holistic quality of those texts. Indeed, results from a number of studies have pointed toward specific characteristics that predict human judgments of writing quality (Crossley, Roscoe, & McNamara, 2013; McNamara et al., 2010; Witte & Faigley, 1981). The accuracy of these results, however, often varies along with various factors associated with the writing assignment, such as the individual rater or the writing prompt (Crossley et al., 2014; Crossley, Varner, et al., 2013; Varner et al., 2013). In this study, we empirically examined these assumptions through a computational linguistic analysis of students' essays. We leveraged both natural language processing and dynamic methodologies to capture variability in students' use of narrative style and to relate that variability to individual differences in writing proficiency, as well as prior science knowledge and reading comprehension skills.

The results from the current study support our hypotheses that writing proficiency can be characterized (at least in part) by students' flexibility across multiple essay prompts. Namely, students who are more flexible in their use of narrativity tend to receive higher scores on their essays, whereas less flexible writers tend to produce lower quality essays. Using random walk analyses, we were able to visualize students' flexible or inflexible use or narrativity across the 16 training essays. These analyses revealed the differential patterns exhibited by the less and more skilled writers, with the skilled writers remaining near the origin of the scatterplot and the less skilled writers straying further from the

Table 5
Correlations Between Narrative Distance Scores and Individual Difference Measures

Individual difference measure	*r*
Reading comprehension	$-.59^{**}$
Vocabulary knowledge	$-.41^{*}$
Prior knowledge (overall)	$-.39^{*}$
Science prior knowledge	$-.44^{*}$
History prior knowledge	$-.27$
Literature prior knowledge	$-.20$

* $p < .05$. ** $p < .01$.

origin. To quantify the findings from this random walk analysis, Euclidian distance measures were calculated. The resulting narrativity distance scores provided confirmatory empirical support for the random walk analyses. In particular, the results demonstrated that less skilled students tended to be more consistent (i.e., inflexible) in the degree to which they used narrative properties (i.e., higher narrative distance scores), whereas more skilled students demonstrated more flexibility in their use of narrativity across the 16 essays (i.e., lower narrative distance scores).

Importantly, the relationship between flexibility and narrativity was no longer apparent at posttest. Our interpretation of this result is that the quality of the students' essays had substantially improved by the time they wrote the posttest essay, and, therefore, the individual differences in flexibility were no longer a factor in their posttest essay quality. In other words, the feedback generated by the AWE system was effective. Results from a previous analysis of the larger study (i.e., the comparison between the Writing Pal ITS condition and the AWE condition; Allen, Crossley, et al., 2014, 2015, under review; Crossley, Roscoe, et al., 2013; Roscoe & McNamara, 2013) revealed that students' essay scores substantially improved across the training sessions (Allen, Crossley et al., 2015). Additionally, the accuracy of the students' self-assessments of essay quality (compared with the W-Pal algorithm) increased in accuracy over time. This is important, because it potentially indicates that, with practice and feedback, students can become more aware of the quality and specific characteristics of their own writing and, therefore, produce essays that more effectively address the prompt question.

Additionally, results from the current study revealed important information about individual differences associated with students' flexible use of narrativity. In particular, flexible writers outperformed the inflexible writers on more general assessments of literacy and prior knowledge. Reading comprehension skills were most strongly linked to this flexibility, accounting for 34% of the variance in students' narrative distance scores. This finding suggests that students who were more skilled at comprehending texts and potentially more aware of readers' strategies and cognitive processes (e.g., O'Reilly & McNamara, 2007) were also more easily able to adapt their writing style to match certain contexts.

The results from this study are important for writing researchers and educators, as they indicate that the link between textual properties and writing quality may fluctuate according to the context of a given writing assignment. Accordingly, writing proficiency not only relates to the sophistication of the words and sentences a student produces in a given essay—but also is intimately related to the writer's ability to adapt style, narrative language, and other rhetorical content to individual writing assignments and different audiences. These results may be explained, in part, by the fact that narrativity tends to be an easier writing style to employ for high school students. Thus, when they are faced with multiple difficult writing assignments, they may resort to this easier writing style as a default. Additionally, the results of the individual difference analyses suggest that this flexibility is not exclusively related to writing proficiency; rather, high school students who are more skilled and knowledgeable are better able to adapt the style of their writing according to situational variations.

Although this ability to flexibly adapt to various contexts has been anecdotally cited as an important component of writing proficiency (Graham & Perin, 2007), to date, little to no research has been conducted to empirically test this assumption. The scarcity of research on this topic may be related to the difficulties associated with assessing writing flexibility. First, it requires a longitudinal data set, such as the one presented here, wherein students are asked to compose multiple essays over time and in response to different prompts. To our knowledge, other such data sets have not been reported in the literature. Second, flexibility is a complex construct to measure. This is particularly true for ill-defined domains, such as writing, which rely on human subjectivity to render judgments about quality and style. Standardized writing assessments typically only measure high school students' writing ability in one particular context and, therefore, cannot be sensitive to fluctuations in style, or in an individual's adaptation to different contexts. If researchers and educators aim to develop assessments that can truly capture students' writing proficiency, it is important to remain sensitive to their ability to adapt their style and language choices according to different assignments and contexts.

The findings and methodologies presented here have important implications for the assessment of students' writing proficiency. In particular, our study indicates that the linguistic properties that interact to predict writing quality may be inconsistent from assessment to assessment. Unfortunately, in their current state, standardized assessments of writing proficiency typically only collect a single writing sample from students. Thus, they are unable to take the construct of writing flexibility into account when making judgments about proficiency. This may constitute a critical oversight. Standardized assessments of writing have a strong influence on students' ability to enter college, as well as their receipt of scholarships and other such opportunities. This study suggests that standardized test developers should aim to develop more sophisticated assessments that can capture students' writing skills across a number of different contexts. Additionally, in the future, the techniques used in the current study may be integrated into a number of educational environments to better assess and improve students' writing skills. For instance, ITSs are computer-based educational environments that provide adaptive instruction and feedback to students based on their skills and performance. Writing-based ITSs might take advantage of this technique to provide feedback that not only looks at students' individual essays but also captures their flexibility across multiple time points (Allen, Jacovina, & McNamara, in press).

Notably, the results reported here call for replications across different populations and skill levels of writers and different writing genres. To our knowledge, there are currently no other data sets that would support replications of the current work. Thus, one goal of our future research will be to develop a corpus that contains multiple essays from different genres written by students from varying populations and skill levels. The achievement of this goal will help us to investigate a number of unanswered questions and concerns. Successful authors of persuasive essays, for example, may flexibility adapt their narrativity; however, in other genres, this flexibility may not be a positive writing characteristic. Future research will aim to answer this question as well as a number of other questions that currently remain unanswered. For example, is it the case that flexibility for *all* linguistic properties is positively related to essay quality? Or, are certain properties more consistently important across a number of different assignments? Further, this study points to the importance of feedback in promoting writing flexibility. This finding prompts the following questions: Can students be trained to be more flexible in their writing style? What is the role of *feedback* in the promotion of increased writing flexibility? Finally, what cognitive processes relate to students' flexible use of writing styles? Is this driven by some executive component skill, or is this driven more broadly by students' prior knowledge and use of strategies? Studies aimed at answering these (and other) questions have the potential to provide crucial information about the role of flexibility in students' ability to produce high-quality text.

References

Allen, L. K., Crossley, S. A., Snow, E. L., Jacovina, M. E., Perret, C. A., & McNamara, D. S. (2015). Am I wrong or am I right? Gains in monitoring accuracy in an intelligent tutoring system for writing. In A. Mitrovic, F. Verdejo, C. Conati, & N. Heffernan (Eds.), *Proceedings of the 17th International Conference on Artificial Intelligence in Education* (pp. 533–536). Madrid, Spain.

Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014). Game-based writing strategy tutoring for second language learners: Game enjoyment as a key to engagement. *Language Learning & Technology, 18,* 124–150.

Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Computer-based writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 316–329). New York, NY: Guilford Press.

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

Allen, L. K., Snow, E. L., & McNamara, D. S. (2014). The long and winding road: Investigating the differential writing patterns of high and low skilled writers. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 304–307). London, UK.

Applebee, A. N., Langer, J. A., Jenkins, L. B., Mullis, I., & Foertsch, M. A. (1990). *Learning to write in our nation's schools.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Arneodo, M., Arvidson, A., Badełek, B., Ballintijn, M., Baum, G., Beaufays, J., & Prytz, K. (1995). Measurement of the proton and the deuteron structure functions. *Physics Letters B, 364,* 107–115.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater V. 2. *The Journal of Technology, Learning, and Assessment, 4*(3). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1650/1492

Barab, S. A., Gresalfi, M. S., Dodge, T., & Ingram-Goble, A. (2010). Narratizing disciplines and disciplinizing narratives: Games as 21st century curriculum. *International Journal of Gaming and Computer-Mediated Simulations, 2,* 17–30. http://dx.doi.org/10.4018/jgcms.2010010102

Benhamou, S., & Bovet, P. (1989). How animals use their environment: A new look at kinesis. *Animal Behaviour, 38,* 375–383. http://dx.doi.org/10.1016/S0003-3472(89)80030-2

Biber, D. (1988). *Variation across speech and writing*. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511621024

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Cheong, Y., & Young, R. M. (2006). A computational model of narrative generation for suspense. *Proceedings of AAAI 2006 Workshop on Computational Aesthetics*.

Collins, J. J., & De Luca, C. J. (1993). Open-loop and closed-loop control of posture: A random-walk analysis of centure-of-pressure trajectories. *Experimental Brain Research, 95,* 308–318.

Crossley, S. A., Allen, L. K., & McNamara, D. S. (2014). A multidimensional analysis of essay writing: What linguistic features tell us about situational parameters and the effects of language functions on judgments of quality. In T. B. Sardinha & M. V. Pinto (Eds.), *Multidimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. 197–238). Amsterdam, Netherlands: John Benjamins. http://dx.doi.org/10.1075/scl.60.07cro

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

Crossley, S. A., Roscoe, R. D., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write high quality essays. *Written Communication, 31,* 181–214.

Crossley, S. A., Roscoe, R. D., McNamara, D. S., & Graesser, A. C. (2011). Predicting human scores of essay quality using computational indices of linguistic and textual features. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education* (pp. 438–440). Auckland, New Zealand: AIED.

Crossley, S. A., Varner, L. K., & McNamara, D. S. (2013). Cohesion-based prompt effects in argumentative writing. In C. Boonthum-Denecke & G. M. Youngblood (Eds.), *Proceedings of the 26th International Florida Artificial Intelligence Research Society (FLAIRS) conference* (pp. 202–207). Menlo Park, CA: AAAI Press.

Crossley, S. A., Varner, L. K., Roscoe, R. D., & McNamara, D. S. (2013). Using automated cohesion indices as a measure of writing growth in intelligent tutoring systems and automated essay writing systems. In K. Yacef et al. (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED)* (pp. 269–278). Heidelberg, Germany: Springer.

Crossley, S. A., Weston, J. L., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication, 28,* 282–311. http://dx.doi.org/10.1177/0741088311410188

Crowhurst, M. (1990). Reading/writing relationships: An intervention study. *Canadian Journal of Education, 15,* 155–172. http://dx.doi.org/10.2307/1495373

Davison, A. (1984). Readability formulas and comprehension. In G. G. Duffy, L. R. Roehler, R. Mason, & J. Mason (Eds.), *Comprehension instruction: Perspectives and suggestions* (pp. 128–143). New York, NY: Longman.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing, 18,* 7–24.

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.

Ferrari, M., Bouffard, T., & Rainville, L. (1998). What makes a good writer? Differences in good and poor writers' self-regulation of writing. *Instructional Science, 26,* 473–488. http://dx.doi.org/10.1023/A:1003202412203

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist, 35,* 39–50. http://dx.doi.org/10.1207/S15326985EP3501_5

Flower, L., & Hayes, J. (1981). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Erlbaum.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science, 3,* 371–398. http://dx.doi.org/10.1111/j.1756-8765.2010.01081.x

Graesser, A. C., McNamara, D. S., & Kulikowich, J. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40,* 223–234. http://dx.doi.org/10.3102/0013189X11413260

Graesser, A. C., Olde, B., & Klettke, B. (2002). How does the mind construct and represent stories? In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 231–263). Mahwah, NJ: Erlbaum.

Graham, S., Bollinger, A., Olson, C. B., D'Aoust, C., MacArthur, C., McCutchen, D., & Ollinghouse, N. (2012). *Teaching elementary school students to be effective writers—An educator's practice guide for the Institute of Education Sciences*. Washington, DC: United States Department of Education.

Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools—A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

Haberlandt, K., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General, 114,* 357–374. http://dx.doi.org/10.1037/0096-3445.114.3.357

Haswell, R. H. (2000). Documenting improvement in college writing: A longitudinal approach. *Written Communication, 17,* 307–352.

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

Hiebert, E. H. (2002). Standards, assessments and text difficulty. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed., pp. 337–369). Newark, DE: International Reading Association.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research, 60,* 237–263. http://dx.doi.org/10.3102/00346543060002237

Huot, B. (1996). Towards a new theory of writing assessment. *College Composition and Communication, 47,* 549–566. http://dx.doi.org/10.2307/358601

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research, 1,* 1–26. http://dx.doi.org/10.17239/jowr-2008.01.01.1

Kincaid, J., Fishburne, R., Rogers, R., & Chissom, B. (1975). *Derivation of new readability formulas for navy enlisted personnel*. Branch Report 8–75. Millington, TN: Chief of Naval Training.

Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biology and Evolution, 13,* 660–665. http://dx.doi.org/10.1093/oxfordjournals.molbev.a025626

MacGinitie, W. H., & MacGinitie, R. K. (1989). *Gates MacGinitie reading tests*. Chicago, IL: Riverside.

McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. *Journal of Memory and Language, 25,* 431–444. http://dx.doi.org/10.1016/0749-596X(86)90036-7

McCutchen, D. (2000). Knowledge, processing, and working memory: Implication for a theory of writing. *Educational Psychologist, 35,* 13–23. http://dx.doi.org/10.1207/S15326985EP3501_3

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. (2013). The epistemic stance between the author and reader: A driving force in the cohesion of text and writing. *Discourse Studies, 15,* 579–592. http://dx.doi.org/10.1177/1461445613501446

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). The linguistic features of writing quality. *Written Communication, 27,* 57–86. http://dx.doi.org/10.1177/0741088309351547

McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods, 45,* 499–515. http://dx.doi.org/10.3758/s13428-012-0258-1

McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). Hierarchical classification approach to automated essay scoring. *Assessing Writing, 23,* 35–59. http://dx.doi.org/10.1016/j.asw.2014.09.002

McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 188–205). Hershey, PA: IGI Global.

McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Sources of text difficulty: Across genres and grades. In J. P. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 89–116). Lanham, MD: R&L Education.

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511894664

McNamara, D. S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research, 34,* 147–171. http://dx.doi.org/10.2190/1RU5-HDTJ-A5C8-JVWE

Meadows, M., & Billington, L. (2005). *Review of the literature on marking reliability*. Report for the Qualifications and Curriculum Authority. London, UK: National Assessment Agency.

Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics, 10,* 139–162. http://dx.doi.org/10.1016/0304-3932(82)90012-5

Newkirk, T. (1997). *The performance of self in student writing*. Portsmouth, NH: Heinemann-Boynton/Cook.

Newkirk, T. (2012). How we really comprehend nonfiction. *Educational Leadership, 69,* 29–32.

Olinghouse, N. G., Graham, S., & Gillespie, A. (2015). The relationship of discourse and topic knowledge to fifth graders' writing performance. *Journal of Educational Psychology, 107,* 391–406. http://dx.doi.org/10.1037/a0037549

O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-explanation reading training: Effects for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1053–1058). Mahwah, NJ: Erlbaum.

O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "High-Stakes" measures of high school students' science achievement. *American Educational Research Journal, 44,* 161–196. http://dx.doi.org/10.3102/0002831206298171

O'Reilly, T., Taylor, R. S., & McNamara, D. S. (2006). Classroom based reading strategy training: Self-explanation vs. reading control. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1887–1892). Mahwah, NJ: Erlbaum.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word Count: LIWC* [Computer software]. Austin, TX.

Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology, 94,* 3–13.

Rafoth, B., & Rubin, D. L. (1984). The impact of content and mechanics on judgments of writing quality. *Written Communication, 1,* 446–458. http://dx.doi.org/10.1177/0741088384001004004

Roscoe, R. D., Crossley, S. A., Snow, E. L., Varner, L. K., & McNamara, D. S. (2014). Writing quality, knowledge, and comprehension correlates of human and automated essay scoring. In W. Eberle & C. Boonthum-Denecke (Eds.), *Proceedings of the 27th International Florida Artificial Intelligence Research Society (FLAIRS) Conference* (pp. 393–398). Palo Alto, CA: AAAI Press.

Roscoe, R. D., & McNamara, D. S. (2013). Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of*

Educational Psychology, 105, 1010–1025. http://dx.doi.org/10.1037/a0032340

Roscoe, R. D., Varner, L. K., Crossley, S. A., & McNamara, D. S. (2013). Developing pedagogically-guided threshold algorithms for intelligent automated essay feedback. *International Journal of Learning Technology, 8,* 362–381. http://dx.doi.org/10.1504/IJLT.2013.059131

Saddler, B., & Graham, S. (2007). The relationship between writing knowledge and writing performance among more and less skilled writers. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 23,* 231–247. http://dx.doi.org/10.1080/10573560701277575

Schank, R. C., & Abelson, R. P. (1995). Knowledge and memory: The real story. In R. S. Wyer (Ed.), *Knowledge and memory: The real story* (pp. 1–85). Hillsdale, NJ: Erlbaum.

Shanahan, T., & Tierney, R. J. (1990). Reading-writing relationships: Three perspectives. In J. Zutell & S. McCormick (Eds.), *Literacy theory and research: Analyses from multiple paradigms* (pp. 13–34). Chicago, IL: National Reading Conference.

Shermis, M., & Burstein, J. (Eds.). (2003). *Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Erlbaum.

Snow, E. L., Allen, L. K., Russell, D. G., & McNamara, D. S. (2014). Who's in control? Categorizing nuanced patterns of behaviors within a game-based intelligent tutoring system. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 185–192). London, UK.

Snow, E. L., Likens, A., Jackson, G. T., & McNamara, D. S. (2013). Students' walk through tutoring: Using a random walk analysis to profile students. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 276-279). Heidelberg, Germany: Springer.

Soller, A., & Lesgold, A. (2003). A computational approach to analyzing online knowledge sharing interaction. In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Proceedings of the 11th Annual Meeting of Artificial Intelligence in Education* (pp. 253–260). Sydney, Australia: Springer.

Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology, 63,* 358–385. http://dx.doi.org/10.1006/jecp.1996.0054

Tierney, R. J., & Shanahan, T. (1991). Research on the reading-writing relationship: Interactions, transactions, and outcomes. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *The handbook of reading research* (Vol. 2, pp. 246–280). New York, NY: Longman.

Varner, L. K., Roscoe, R. D., & McNamara, D. S. (2013). Evaluative misalignment of 10th-grade student and teacher criteria for essay quality: An automated textual analysis. *Journal of Writing Research, 5,* 35–59. http://dx.doi.org/10.17239/jowr-2013.05.01.2

Vorderer, P., Wulff, H. J., & Friedrichsen, M. (1996). *Suspense: Conceptualizations, theoretical analyses, and empirical explorations*. Mahwah, NJ: Erlbaum.

Witte, S., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication, 32,* 189–204. http://dx.doi.org/10.2307/356693

Zhou, M. (2013). Using traces to investigate self-regulatory activities: A study of self-regulation and achievement goal profiles in the context of web search for academic tasks. *Journal of Cognitive Education and Psychology, 12,* 287–305. http://dx.doi.org/10.1891/1945-8959.12.3.287

Appendix

Pretest and Posttest Essay Prompts

Essay Prompt 1. You will now have 25 minutes to write an essay on the prompt below.

The essay gives you an opportunity to show how effectively you can develop and express ideas. You should, therefore, take care to develop your point of view, present your ideas logically and clearly, and use language precisely.

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

Whereas some people promote competition as the only way to achieve success, others emphasize the power of cooperation. Intense rivalry at work or play or engaging in competition involving ideas or skills may indeed drive people either to avoid failure or to achieve important victories. In a complex world, however, cooperation is much more likely to produce significant, lasting accomplishments.

Do people achieve more success by cooperation or by competition?

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

Essay Prompt 2. You will now have 25 minutes to write an essay on the prompt below.

The essay gives you an opportunity to show how effectively you can develop and express ideas. You should, therefore, take care to develop your point of view, present your ideas logically and clearly, and use language precisely.

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

All around us appearances are mistaken for reality. Clever advertisements create favorable impressions but say little or nothing about the products they promote. In stores, colorful packages are often better than their contents. In the media, how certain entertainers, politicians, and other public figures appear is sometimes considered more important than their abilities. All too often, what we think we see becomes far more important than what really is.

Do images and impressions have a positive or negative effect on people?

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

Received August 25, 2014
Revision received November 30, 2015
Accepted December 4, 2015 ■

See page 942 for a correction to this article.