JOURNAL OF LEARNING ANALYTICS

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

Keys to Detecting Writing Flexibility Over Time: Entropy and Natural Language Processing

Erica L. Snow,[1] Laura K. Allen,[1] Matthew E. Jacovina,[1] Scott A. Crossley,[2] Cecile A. Perret,[1] and Danielle S. McNamara[1]

[1] Learning Sciences Institute, Arizona State University, Tempe, AZ, USA
[2] Georgia State University, Atlanta, GA, USA
Erica.L.Snow@asu.edu

ABSTRACT: Writing researchers have suggested that students who are perceived as strong writers (i.e., those who generate texts rated as high quality) demonstrate flexibility in their writing style. While anecdotally this has been a commonly held belief among researchers and educators, there is little empirical research to support this claim. This study investigates this hypothesis by examining how students vary in their use of two linguistic features (i.e., narrativity and cohesion) across 16 prompt-based essays. Forty-five high school students wrote 16 essays across 8 sessions within an Automated Writing Evaluation (AWE) system. Natural language processing (NLP) techniques and Entropy analyses were used to calculate how rigid or flexible students were in their use of narrative and cohesive linguistic features over time and how this trait related to individual differences in literacy abilities (i.e., vocabulary knowledge and comprehension ability), prior world knowledge, and essay quality. For instance, through the unique combination of NLP and Entropy, we found that patterns of narrative flexibility (or rigidity) were related, significantly and reliably, to students' prior reading comprehension ability after 2 sessions (4 essays). Similarly, students' flexible (or rigid) use of cohesive features was reliably related to their prior reading comprehension ability after 5 sessions (10 essays). These exploratory methodologies are important for researchers and educators, as they indicate that writing flexibility is indeed a trait of strong writers and can be detected rather quickly using the combination of textual features and dynamic analyses.

Keywords: Writing flexibility, stealth assessment, natural language processing, Entropy analyses

1 INTRODUCTION

The ability to communicate effectively through written text is crucial for professional and academic success (Flower & Hayes, 1981; Sharma & Patterson, 1999). Strong writers effectively maneuver a variety of subtasks while generating text: coordinating information, setting goals, solving problems, and regulating their use of multiple sources of knowledge (Flower & Hayes, 1981). However, many students struggle with this complex cognitive task (Graham & Perin, 2007). Indeed, as recently as 2011, approximately only one quarter of American students in 8th through 12th grade scored at a proficient level on a nationwide computer-based writing assessment (National Center for Education Statistics, 2011). These findings demonstrate a need for researchers to understand the writing process and the skills related to its mastery.

To better understand how students' writing skills can be improved, researchers examine the characteristics of writing, particularly as they relate to its quality (McNamara, Crossley, & McCarthy,

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

2010; McNamara, Crossley, Roscoe, Allen, & Dai, 2015; Witte & Faigley, 1981). Recently, researchers have argued that differences in writing quality can be detected based on the linguistic features embedded within student essays. Studying the ways in which writers employ these features, and how they do so in different contexts, can provide insights about how writers succeed or struggle. Eventually, this understanding can lead to improvements in writing instruction. In the current paper, we investigate student writers within a computer-based system. We assess student flexibility in two linguistic features that may be critical to essay quality and important for understanding student writing skills: narrativity and cohesion. Specifically, we employ a form of dynamic analyses, Entropy, to calculate changes in students' use of these two linguistic features across 16 prompt-based essays generated within a computer-based environment. We also investigate how changes in the use of narrativity and cohesion relate to students' reading comprehension ability, vocabulary knowledge, and prior world knowledge.

1.1 Narrativity

Narrativity refers to the degree to which texts are more story-like, include events, and refer to characters as opposed to texts that describe things and provide information about those things. Narrative texts are easier to understand and remember than texts that contain few narrative elements (e.g., informational texts; Graesser, Olde, & Klettke, 2002; Haberlandt & Graesser, 1985). Readers' ease of processing narrative texts makes sense given that people's everyday conversations and other experiences with language are replete with narrative elements (Clark, 1996; Gerrig, 1993). Educators have thus made logical arguments that student writing (including non-fictional writing) should include narrative elements as a means to increase reader engagement (Newkirk, 2012).

Although some students, in some contexts, would likely benefit from including more narrative elements in their writing, research has not supported the idea that successful essays typically have higher levels of narrativity. One analysis, for example, found that in a corpus of prompt-based, argumentative essays written by college freshman, higher scores from expert graders were associated with lower narrativity (McNamara, Crossley, & Roscoe, 2013). McNamara and colleagues argued that although texts higher in narrativity are likely easier to read, they may lack the informational content crucial for composing a successful argumentative essay. However, not all successful essays follow the same formula. In a different analysis, four distinct linguistic profiles were identified for highly scored argumentative essays (Crossley, Roscoe, & McNamara, 2014). Essays in one profile (described as academic) seemed to be successful due to their structural components, syntactic complexity, and their relatively infrequent use of causal verbs and pronouns, often found in narrative texts. Essays in another profile (described as accessible), by contrast, contained relatively more causal connectives and verbs common to narrative writing. These two divergent profiles of successful essays highlight the difficulty of understanding what makes a successful writer.

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

1.2 Cohesion

Like narrativity, cohesion has also been identified as an important component of writing (Crossley & McNamara, 2010; 2011; Crossley, Roscoe, & McNamara, 2014; Allen, Snow, & McNamara, 2014). Cohesion generally refers to the presence or absence of explicit cues in the text that allow the reader to make connections between the ideas in the text. Such explicit cues may include overlapping words and concepts between sentences or the use of connectives (e.g., *and*, *because*, *therefore*, and *consequently*) that mark links between sentences, phrases, and words. In general, these explicit cues communicate to the reader that similar ideas are being referred to across consecutive sentences or point to relations between ideas (Halliday & Hasan, 1976). Such sentence level cohesion is known as local cohesion. Local cohesion can be contrasted with global cohesion, which links larger chunks of text together at the paragraph level (McNamara, Kintsch, Songer, & Kintsch, 1996). Previously, researchers have investigated how cohesion relates to essay quality. This work has revealed mixed results concerning the relation of cohesive elements and essay quality. For instance, McNamara, Crossley, and McCarthy (2010) examined the role linguistic features (including cohesive devices) play in predicting independent essay quality (i.e., writing that requires no specific background knowledge). Their results found no relation between text cohesion and essay quality.

A follow-up study (Crossley, Roscoe, Graesser, & McNamara, 2011) examined the relation between students' use of cohesive features within their essay and expert-rated scores on their essays. This study found that two indices of global cohesion and one index of overall text cohesion (text givenness) significantly correlated with essay quality. Finally, Crossley and McNamara (2012) examined the relation between indices of cohesion (e.g., content word overlap, positive logical connectives, aspect repetition, and semantic similarity between sentences) and essay quality for second language (L2) writers. Results from this analysis found a negative correlation between students' use of cohesive devices and essay quality. Taken together, these results suggest that the impact of cohesive elements on essay quality may differ based on the type of cohesion measured (local, global, or overall text cohesion) and may be situational (i.e., based on the population of writers or the expectations of the raters). Indeed, recently, researchers have argued that successful writing should not be defined through a single set of features; on the contrary, successful writing has multiple profiles that vary as a function of context (Crossley, Roscoe, & McNamara, 2014; Allen, Snow, & McNamara, in press, 2014; Snow, Allen, Jacovina, Perret, & McNamara, 2015).

1.3 Individual Differences

Although linguistic features such as those related to narrativity and cohesion provide a glimpse into the writing process, researchers have also begun to investigate how individual differences may relate to writing quality (Delisle & Delisle, 2011; Graham & Harris, 2000; Graham & Perin, 2007; Shanahan & Tierney, 1990). Indeed, every student brings a unique set of strengths and weaknesses to the writing process. Thus, researchers have begun to include individual differences within their writing models as a

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

means to understand this complex process (Kellogg, 2008; McCutchen, 2000; Swanson & Berninger, 1996).

A vast range of student attributes can encompass individual differences, such as self-regulation ability, prior knowledge, goal orientation, and literacy skills (i.e., vocabulary knowledge and reading comprehension ability), many of which have been shown to influence writing quality. For instance, Shanahan and Tierney found that students' reading comprehension ability was related to their ability to write effectively. Thus, students who are better at comprehending text are also better at generating it (Shanahan & Tierney, 1990). Similarly, Graham and Perin (2007) found that students' vocabulary knowledge was an important factor in writing proficiency. Thus, students who possessed a higher vocabulary generated essays that revealed more diversity in the way ideas were presented. Finally, prior knowledge has also been hypothesized to influence writing quality. Allen and colleagues (in press) postulate that when students have more world knowledge they are better able to develop strong arguments supported by real-world examples and evidence.

1.4 Flexibility

Another proposed characteristic of strong writers is their flexibility in writing style (Graham & Perin, 2007). Skilled writers do not simply reuse language or linguistic features in multiple writing tasks; instead, they assess each writing situation and adapt their style accordingly (Graham & Perin, 2007). However, one problem that researchers have encountered when trying to examine the link between flexibility and writing quality (or any other individual differences) is that flexibility is hard to measure. Thus, few studies have explicitly investigated whether or not flexibility is an important component of writing quality.

Allen, Snow, and McNamara (2014; in press) recently completed two empirical investigations into writing flexibility. By using natural language processing (NLP) and random walk analyses to investigate SAT-style prompt-based essays, they found that students who had higher literacy abilities (i.e., higher vocabulary knowledge and reading comprehension skills) generally used narrative elements flexibly across a series of SAT-style prompt-based essays. They also showed that narrative flexibility was significantly related to essay quality (in press). Their first similar analysis investigated how literacy skills and essay quality were related to students' flexible use of cohesive elements across time (2014). Similar to their work on narrativity, they found that writers who produced high-quality essays were more flexible in their use of cohesion across time (2014). Students' flexibility in the use of cohesive elements was also associated with higher reading ability and greater prior knowledge, particularly knowledge of science.

Combined, the two analyses conducted by Allen, Snow, and McNamara (2014; in press) provide a starting point for researchers to begin to understand the construct of writing flexibility. However, an important question that has not yet been investigated is how writing flexibility manifests and evolves across time. Specifically, Allen, Snow, and McNamara examined aggregated flexibility across 16 essays

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

using a random walk distance score (see Snow, Likens, Jackson, & McNamara, 2013, for a detailed description of random walk analyses). Although this suggests that writing flexibility is important, it does not reveal *when* flexibility can be detected and *how* this skill evolves across time, both of which must be understood to guide instruction effectively for students who are less flexible in their writing.

One reason for the lack of studies that focus on the emergence of writing flexibility across time is that research has predominantly focused on summative assessments of writing, which lack the nuances required to investigate flexibility. Indeed, the previously mentioned studies conducted by Allen, Snow, and McNamara (2014; in press) are, to our knowledge, the only studies to measure writing flexibility across time using online metrics (i.e., linguistic features of essays). The current work builds upon that work by using additional covert measures (i.e., stealth assessments) to capture nuanced changes in writing style as they manifest across time.

1.5 Stealth Assessments

Traditional, summative assessments often take the form of tests or graded assignments intended to measure student learning. Formative assessments, however, are intended to gauge students' current knowledge or skills in order to guide instruction. Formative assessment, when well implemented, is beneficial to student learning (Black & William, 1998). Thus, developing meaningful formative assessments is an important, albeit challenging, objective for educators, and is essential for the success of computer-based educational environments that aim to guide instruction based on the needs of individual students (Gikandi, Morrow, & Davis, 2011). One way to capture student behaviours and knowledge without disrupting their experience within a computer-based system is through stealth assessment (Shute, 2011; Shute & Kim, 2013). Stealth assessments, like formative assessments, are intended to measure student characteristics (e.g., their current content knowledge or level of engagement) with the ultimate goal of adapting instruction and providing formative feedback based on those characteristics. Importantly, though, stealth assessments are measured through students' normal use of an educational system. That is, stealth measures do not require students to report on their perceptions of their own behaviour and knowledge, which are often inaccurate when compared to their actual behaviour and knowledge (McNamara, 2011).

Stealth assessments, used to measure numerous constructs, can take many forms. For example, in Zoo U, an educational game designed to assess young students' social skills, students' clicking behaviour in animated scenes is used to measure impulse control (DeRosier, Craig, & Sanchez, 2012). This stealth assessment, built directly into the normal interaction with the environment, does not require students to disengage from the activity to answer questions about how impulsive they feel. Other research has used stealth assessments to measure the amount of exerted agency (Snow, Jacovina, Allen, Dai, & McNamara, 2014) and gaming behaviours (Baker, Corbett, Roll, & Koedinger, 2008). In an intelligent tutor system that focuses on student essays, stealth assessment affords the assessment of writing that goes beyond traditional summative measures of overall essay quality.

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

1.6 Current Study

This study builds upon and expands the work by Allen, Snow, and McNamara (2014; in press) by proposing that a successful writer should flexibly adapt their use of narrative and cohesion features in text across time, depending on the context of the writing task and their literacy abilities. We hypothesize that these traits will become detectable relatively quickly. This study further examines the extent to which the relation between writing flexibility and students' literacy skills remains stable across time. To test our hypotheses, we employ a stealth measure of writing flexibility that emerges through the dynamic analysis of students' use of narrativity and cohesion across multiple essays. The current study uses NLP and Entropy analysis to investigate whether individual differences related to writing proficiency are associated with students' flexible use of linguistic features across multiple prompt-based essays. The overarching goal of this work is to shed light upon the complex interplay of flexibility, individual differences, and writing skills. By identifying when students become more or less flexible, instruction could be adapted for individual students, such as providing better instruction to students who rigidly stick to one writing approach or another. The primary research questions of this study are as follows:

1. At what point across multiple essays is writing flexibility associated with individual difference in prior literacy skill and essay quality?

2. How many essays are required to reliably detect the relation between literacy skills and flexibility in writing style?

1.6.1 Automated Writing Evaluation Systems

To answer the above questions, we use an Automated Writing Evaluation (AWE) system. AWEs offer students the opportunity to practice writing and automatically receive scores and feedback on their essays (Grimes & Warschauer, 2010). This form of deliberate practice (i.e., practice with individualized feedback) is critically important in order to develop strong writing skills; AWE systems, therefore, can benefit students by providing them with relevant feedback, which provides more beneficial writing practice opportunities and simultaneously reduces the burden on instructors. The algorithms that drive the scores provided by these systems have demonstrated fairly high reliability and accuracy (e.g., Attali & Burstein, 2006; Warschauer & Ware, 2006). The scores provided by expert human raters and computers tend to correlate between r = 0.80 and 0.85, which is a similar range to the range found between human raters (Rudner, Garcia, & Welch, 2006; Warschauer & Ware, 2006). Similarly, they tend to report perfect agreement between 40–65% and adjacent agreement (within one point of each other) between 90–100%. Overall, AWE systems are beneficial for students because they can provide immediate summative feedback to students on their essays — all without any input from the instructor. Because of these features, a number of AWE systems have been developed and are now being used in classrooms, such as Criterion (Attali & Burstein, 2006), MyAccess (Rudner et al., 2006), WriteToLearn (Landauer, Laham, & Foltz, 2003), and WPP Online (Page, 2003).

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

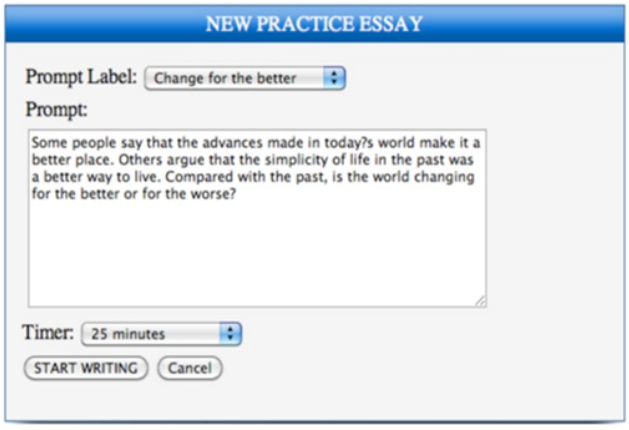

Figure 1: Screenshot of the W-Pal prompt

In this study, we use the AWE found in the Writing Pal (W-Pal) system. W-Pal is an intelligent tutoring system developed to provide high school and entering college students with writing strategy instruction and practice (Roscoe, Snow, Brandon, & McNamara, 2013). W-Pal contains an AWE component that allows students to write essays and receive both summative and formative feedback automatically. The system includes a basic word processor (see Figure 1), which allows students to write essays based on a number of pre-determined prompts (alternatively, teachers can add prompts). When a student finishes writing an essay, it is submitted to the W-Pal system and the student's essay is provided with a holistic score and formative feedback (see Figure 2 for a screenshot of the feedback screen). Finally, after reviewing this feedback, students have the opportunity to revise their essays.

An important aspect of the W-Pal system is the provision of high-level feedback. Specifically, the feedback in the W-Pal system focuses on strategies that students can enact when they revise their essays (Roscoe, Varner, Crossley, & McNamara, 2013). For example, if a student submits an essay to W-Pal that contains no structure (i.e., it only contains one paragraph), the feedback may focus on strategies that can improve essay organization, such as the use of flow charts or outlines to visualize the structure of an essay.

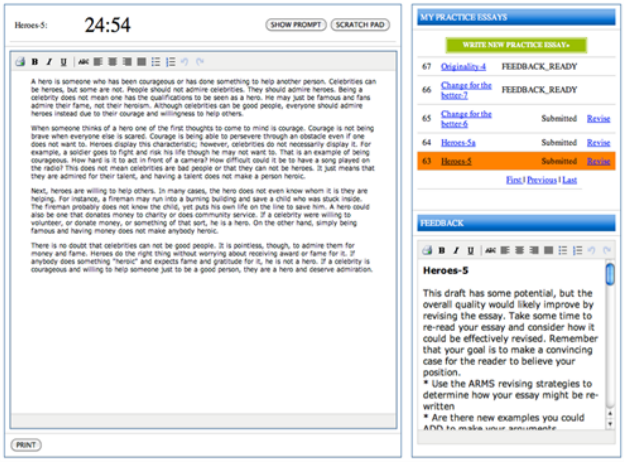

Figure 2: Screenshot of the W-Pal essay processor and feedback screen

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

2 METHOD

2.1 Participants

The current study is part of a larger study (n = 86) that compared a writing tutor (Writing Pal) to an Automated Writing Evaluation (AWE) system (Allen, Crossley, Snow, & McNamara, 2014). The analyses presented here focus solely on the participants who engaged with the AWE system (n = 45). All 45 participants were high school students recruited from a southwest urban environment located in 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. In this study, students' self-reported ethnicity breakdown was 62.2% Hispanic, 13.3% Asian, 6.7% Caucasian, 6.7% African-American, and 11.1% "other."

2.2 Study Procedure

The current study was comprised of a ten-session experiment. During the first session, students completed a pre-test that contained measures of prior writing ability, prior knowledge, reading ability, and vocabulary knowledge. During the following eight sessions, students wrote two essays per day (totalling 16) in the AWE system. Finally, during the last session, students completed a post-test that contained measures similar to the pre-test.

2.2.1 Pre-test

During session 1, all students completed a battery of individual difference measures. This battery of measures took 1 hour to complete and included demographics, prior knowledge, prior reading ability, prior vocabulary knowledge, and writing proficiency (25-minute prompt-based SAT-style essay).

2.2.2 Training

During sessions 2–9, students engaged in the training portion of the experiment where each student practiced writing 25-minute timed essays on SAT-style prompts. During each training session, students wrote two prompt-based essays (see Figure 1 for screenshot of W-Pal AWE prompt). This resulted in each student writing 16 essays across 8 training sessions (see Table 1 for prompt topic and order). After each essay, students received automated formative feedback about their essay from the W-Pal AWE system (see Figure 2 for a screenshot of the essay processor and feedback messages). Students could then use this feedback to revise their essay for approximately 10 minutes.

2.2.3 Post-test

During session 10, all participants completed a post-test comprised of measures similar to the pre-test, including a writing proficiency test (25-minute SAT-style essay).

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

Table 1: Essay Prompt Order

Training Session Day	Essay Prompts
Training Day 1	Planning & Originality
Training Day 2	Winning & Loyalty
Training Day 3	Patience & Memories
Training Day 4	Heroes & Choices
Training Day 5	Perfection & Optimism
Training Day 6	Uniformity & Problems
Training Day 7	Beliefs & Happiness
Training Day 8	Fame & Honesty

2.3 Materials and Measures

2.3. Prior Reading Ability

Students' reading ability was assessed using the Gates-MacGinitie (4th ed.) reading skill test (MacGinitie, MacGinitie, Maria, & Dreyer, 2000). The test included 48 multiple-choice questions that 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. All students were given 20 minutes to complete as many questions as they could.

2.3.2 Vocabulary Knowledge

Students' vocabulary knowledge was assessed using the Gates-MacGinitie (4th ed.) vocabulary test (MacGinitie, MacGinitie, Maria, & Dreyer, 2000). In this test, students were shown 45 simple sentences, each with an underlined vocabulary word. They were then asked to choose the word most closely related to the underlined word within the sentence from a list of five choices. All students were given 10 minutes to answer as many questions as they could.

2.3.3 Prior Knowledge

Prior science knowledge was assessed using a ten-item, four alternative multiple-choice test addressing knowledge of areas such as biology, scientific methods, earth science, and chemistry (O'Reilly, Best, & McNamara, 2004). These questions are designed to measure students' general science knowledge and are modified from a previously validated prior knowledge measure (Cronbach's alpha =α= .74; O'Reilly, Best, & McNamara, 2004).

2.3.4 Essay Quality

Students wrote 18 essays (1 at pre-test, 16 during training, and 1 at post-test). Pre-test and post-test essays were scored by two expert human raters. These raters used an SAT-style rubric that ranged from 1 to 6. Inter-rater reliability was reached at r=.70. These raters also reached 100% adjacent agreement. Students' essay scores during training were measured using a hierarchical classification approach

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

(McNamara, Crossley, & Roscoe, 2013) that provided automated assessment (range of 1 to 6) of all the training essays. This algorithm uses features computed through the automated tools, Coh-Metrix, the Writing Assessment Tool (WAT), and Linguistic Inquiry and Word Count (LIWC).

2.4 Coh-Metrix

To assess the linguistic features of students' essays (during training), Coh-Metrix was used. Coh-Metrix (McNamara & Graesser, 2012; McNamara, Graesser, McCarthy, & Cai, 2014) is an automated text analysis tool that calculates indices related to text properties at the word, sentence, and discourse levels. This tool was originally developed to provide more robust measures of text difficulty (Duran, Bellissens, Taylor, & McNamara, 2007; Graesser, McNamara, Louwerse, & Cai, 2004); its focus on the multiple levels of texts affords the opportunity for this tool to provide more specific information about the challenges and scaffolds contained within a given text. The indices in Coh-Metrix range from basic text properties to higher-level measures of cohesion and discourse. The basic text indices within Coh-Metrix report statistical information related to the length of specific discourse units within the text, such as the number of words and paragraphs, as well as the average length of words (average number of syllables per word), sentences (average number of words per sentence), and paragraphs (average number of sentences per paragraph). The lexical measures in Coh-Metrix relate to the types of words contained within a text, such as the overall level of word specificity (hypernymy), the average frequency of the individual words, and the diversity of words used throughout a given text. Syntactic indices describe the complexity and properties of the sentences within a text; this category of measure includes the number of modifiers per noun phrase within the text, the similarity of the syntax among the sentences, and the frequency of passive sentences in the text. Finally, Coh-Metrix reports on the cohesion of a given text using a number of measures; some of these indices include the incidence of connectives in a text as well as word overlap measures across sentences, which indicate how frequently words overlap amongst the sentences within a text.

Research with Coh-Metrix has suggested that there are multiple dimensions within texts that work together to affect comprehension (McNamara, Graesser, & Louwerse, 2012). Therefore, to account for these dimensions, Graesser and colleagues developed the Coh-Metrix Easability Components (Graesser, McNamara, & Kulikowich, 2011). Coh-Metrix reports five primary Easability components: Narrativity, Syntactic Simplicity, Word Concreteness, Referential Cohesion, and Deep Cohesion, which all provide differential information about aspects of texts that can influence comprehension (see Graesser, McNamara, & Kulikowich, 2011 for more specific information about these indices). Importantly, these Easability Components are aligned within an existing multilevel framework of comprehension (Graesser & McNamara, 2011). Relevant to the current study are the Narrativity and Referential Cohesion Components. These components are discussed in further detail below.

2.4.1 Narrativity

The narrativity component score provided by Coh-Metrix is informed by 17 Coh-Metrix indices that assess the extent to which a text contains narrative versus informational content. The overall narrativity

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

of a text is representative of how "story-like" it is — using characters, places, events, or other discourse elements that may be familiar to readers. One of the primary characteristics of highly narrative texts is that they contain more descriptions of actions and events (as opposed to informational texts, which contain less familiar words and a high incidence of nouns); therefore, the Coh-Metrix narrativity component includes information about the incidence of main verbs, adverbs, and intentional events, actions, and particles within a given text. Additionally, narrative texts tend more closely resemble oral language than informative texts (Biber, 1988), as they contain simpler sentence constructions and a higher frequency of familiar words and pronouns. Texts that are highly narrative are typically considered easier to read, recall, and comprehend than informative texts because they are engaging and more familiar to readers (Graesser & McNamara, 2011; Haberlandt & Graesser, 1985).

2.4.2 Cohesion

The referential cohesion component score provided by Coh-Metrix is informed by 13 indices related to referential cohesion, or the extent to which a text contains words and ideas that overlap across the sentences in a text, as well as the entire text (Graesser McNamara, & Kulikowich, 2011; McNamara et al., 2014). The indices collectively assess the degree to which the text contains words or concepts that overlap across sentences and the entire text.

2.5 Flexibility

Students' propensity to use narrativity and cohesion in a flexible way across the 16 training essays was calculated using Shannon Entropy (Shannon, 1951). Shannon Entropy analysis is a statistical measure to capture random, controlled, and ordered processes (Fasolo, Hertwig, Huber, & Ludwig, 2009; Grossman, 1953; Snow, Allen, Jacovina, & McNamara, 2014; Snow, Jackson, & McNamara, 2014). In the current study, Shannon Entropy is used to gain a deeper understanding of how students' use of linguistic features fluctuates across time. Previous work by Allen, Snow, and McNamara (in press) used random walk distance scores as a measure of flexibility across the set of 16 essays. While this measure successfully captures changes in students' overall use of linguistic features, it is not designed to capture fine-grained changes as they manifest over time. This is because the random walk distance score is calculated through the culmination of all prompt-based essays. Thus, it is unlike Entropy analyses, which do afford researchers the opportunity to examine how this skill manifests across time.

In the current study, students wrote 16 essays. Each essay was analyzed by Coh-Metrix and assigned a narrativity percentage score and a cohesive percentage score (ranging from 0 to 100), which represented the amount of narrativity and cohesion, respectively, used within the essay. This resulted in students having 16 narrativity scores (1 per essay) and 16 cohesion scores. These scores were then separated into orthogonal quartiles for each linguistic feature. This resulted in four narrative categories (High Narrativity being >75%, Medium-High Narrativity ranging between 50% and 75%, Medium-Low Narrativity ranging between 25% and 50%, and Low Narrativity being <25%) and four cohesion categories (High Cohesion being >75%, Medium-High Cohesion ranging between 50% and 75%, Medium-Low Cohesion ranging between 25% and 50%, and Low Cohesion being <25%).

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

Entropy was calculated separately for both the cohesion and narrativity scores. Equation 1 represents the Shannon Entropy formula used within the current study. Within this equation, $P(x_i)$ represents the probability of a given state. For instance, the Entropy for student X is the additive inverse of the sum of products calculated by multiplying the probability of each category by the natural log of the probability of that category. This formula captures whether students are ordered or flexible in their use of narrativity across all 16 essays.

$$\mathrm{H(x)} = -\sum_{i=0}^{N} P(x_i)(\log_e P(x_i)) \qquad (1)$$

Shannon Entropy produces a number that captures the amount of certainty or uncertainty present in a specific time series or series of events. The differences between certain (rigid) and uncertain (flexible) states can be visualized using two jars of marbles. For example, in Figure 3, the jar on the left contains only black marbles. Thus, if you were to take a marble out of the jar (without looking), you would be very certain that the marble you choose would be black. This results in low Entropy, where the next choice in a series is highly predictable and certain (i.e., rigid). Conversely, the jar on the right contains both white and black marbles. Thus, if you were to pull blindly from the jar, you would be less certain of the colour of marble you would pick. Indeed, in this jar, the marbles are mixed and distributed unevenly. Thus the predictability of what comes next would be low, resulting in a high Entropy score (i.e., flexible).

Figure 3: **Visual representation of Low Entropy (Left) and High Entropy (Right)**

Within the context of the current study, high Entropy scores are suggestive of students' flexible use of the specific linguistic features (i.e., cohesion and narrativity), whereas low Entropy scores suggest students are demonstrating a highly rigid use of the specific linguistic feature. Indeed, this metric affords the opportunity to capture how much fluctuation or rigidity students exhibit in their use of narrativity and cohesion across all 16 essays. We can also use the Entropy formula to calculate cumulative Entropy scores for each day of training. These cumulative scores can then be used to measure how students' observed flexibility changes across time and how flexibility relates to their prior abilities, such as reading comprehension skill, vocabulary knowledge, and science knowledge.

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

3 RESULTS

The current work examines how students vary in their use of narrativity and cohesion across 16 prompt-based essays. First, correlation and regression analyses were conducted to examine the relations among student flexibility (Entropy scores) in writing style (narrative flexibility scores and cohesive flexibility), average essay quality, and three individual difference literacy ability measures (vocabulary, reading ability, and prior knowledge) across all essays. This analysis is intended to replicate previous findings from Allen, Snow, and McNamara (in press) that used a different measure of narrative flexibility and cohesive flexibility (i.e., random walk analysis) and found that it related to individual differences in literacy skill and essay quality. Descriptive statistics for this analysis are provided in Table 2.

Table 2: Descriptive Statistics for Narrative Flexibility, Cohesive Flexibility, Prior Abilities, and Essay Quality

Variable	M	SD	Range
Narrative Flexibility Score	0.94	0.40	0.00–1.59
Cohesive Flexibility Score	1.25	0.37	0.00–1.62
Reading Comprehension	23.31	7.77	10.00–36.00
Prior Knowledge	5.27	1.80	2.00–9.00
Vocabulary Knowledge	25.78	8.32	6.00–40.00
Average Essay Score	2.65	0.62	1.63–4.06

3.1 Narrative Flexibility

Pearson correlations were conducted to examine the relation between students' narrative flexibility score and pre-test measures of their reading ability, prior knowledge, vocabulary skills, and prior writing ability. Results from this analysis reveal that students' narrative flexibility scores were positively related to their reading ability, prior knowledge, vocabulary skill, and prior writing ability (see Table 3), thus suggesting that students with higher literacy skills are also more flexible in their use of narrativity across time. To examine these relations further, we conducted a forward stepwise regression analysis to examine the extent to which literacy skills were predictive of narrative flexibility. One variable was retained in the final model predicting 29% of the variance in students' narrative flexibility [F (1, 42) = 17.36 p < .001; R^2 = .29]; prior reading ability [B =.54, t(1, 43) = 4.17, p < .001].

Table 3: Correlations between Narrative Flexibility and Prior Abilities

Individual Difference Measure	Narrative Flexibility Score
Reading Comprehension	.390**
Prior Knowledge	.362*
Vocabulary Knowledge	.307*
Pre-test Essay Quality	.306*
p<.05*; p<.01**	

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

Additional correlations were conducted to examine the link between students' narrative flexibility score and essay quality during training and at post-test. Results revealed a significant relation between average training essay score (r=.319, p=.03) and flexibility. However, there was no relation to narrative flexibility and post-test essay quality (r=.11, p=.49). Thus, narrative flexibility seems to be an important factor during training but this relation may diminish over time. These findings replicate the results reported by Allen, Snow, and McNamara (in press) and reveal that Entropy and random walk distance measures are capturing similar constructs.

Table 4: Descriptive Statistics for Cumulative Narrative Flexibility Scores

Session Day(s)	M	Range
Day 1	0.25	(.00–0.69)
Days 1–2	0.56	(.00–1.32)
Days 1–3	0.95	(.00–1.39)
Days 1–4	0.97	(.00–1.33)
Days 1–5	1.00	(.00–1.67)
Days 1–6	0.97	(.00–1.64)
Days 1–7	0.95	(.00–1.60)
Days 1–8	0.94	(.00–1.59)
N=45		

These initial results support the hypothesis that writing flexibility is related to high literacy skills and essay quality. However, to investigate these results at a more fine-grained level and examine how prior skills impact flexibility across time, narrative flexibility scores were calculated for each day of training, thus providing a finite look at how flexibility manifests across time. For every student, a cumulative narrative flexibility score was calculated for each day of training. This score captures the cumulative amount of flexibility that the student has exhibited in their use of narrativity. For instance, after Day 1, students have completed two essays; therefore, their cumulative narrative flexibility score will be their narrative Entropy score on the Planning and Originality essays. Similarly, after Day 2, students' flexibility score will be the narrative Entropy score for the Planning, Originality, Winning, and Loyalty essays. Table 4 provides the descriptive statistics for each day's running Flexibility score.

Pearson correlations were conducted to examine the relation between students' cumulative narrative flexibility scores and pre-test measures of their vocabulary skills, reading ability, prior knowledge, and prior writing ability. Table 5 shows the relation between students' cumulative narrativity score and their pre-test vocabulary knowledge, reading ability, prior knowledge, and writing ability. Results from this analysis reveal that the relation between students' cumulative flexibility score and vocabulary skill, prior knowledge, and writing ability can be reliably detected around Day 5 (10 essays written). These results demonstrate that students' vocabulary skill, prior knowledge, and writing ability seem to be mostly related to their cumulative narrative flexibility score for the later sessions. Thus, the impact that these skills have on flexibility takes time to manifest.

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

Table 5: Correlations between Cumulative Narrative Flexibility Scores and Prior Abilities

Session Day	Reading Ability	Prior Knowledge	Vocabulary Knowledge	Writing Ability
Day 1	.199	.203	.117	.106
Days 1–2	.464**	.306*	.416**	.431**
Days 1–3	.417**	.175	.141	.291 (M)
Days 1–4	.446**	.317*	.178	.260 (M)
Days 1–5	.517**	.456**	.280(M)	.310*
Days 1–6	.543**	.454**	.301*	.340*
Days 1–7	.537**	.466**	.326*	.300*
Days 1–8	.543**	.466**	.326*	.306*
$p < .10$ (M) Marginal; $p < .05$*; $p < .01$**				

Interestingly, however, the relation between students' cumulative narrative flexibility score and their reading ability was detectable after Day 2 (with four essays written so far). Thus, reading skill seems to be a reliable predictor of flexibility, after only a few writing prompts. This suggests that students' reading ability plays a critical role (more than the other literacy skills) in their ability to demonstrate narrative flexibility in their writing style.

3.2 Cohesive Flexibility

Pearson correlations were also conducted to examine the relation between students' overall cohesive flexibility score and pre-test measures of their reading ability, prior knowledge, vocabulary skills, and writing ability (see Table 6). Results from this analysis reveal students' cohesive flexibility scores were marginally related to pre-test essay quality and moderately related to their reading ability and prior knowledge. However, there was a weak, non-significant relation between students' cohesive flexibility and their vocabulary skill. These results suggest that students with higher reading ability and prior knowledge are more flexible in their use of cohesion across time.

Table 6: Correlations between Cohesive Flexibility scores and Prior Abilities

Individual Difference Measure	Cohesive Flexibility Score
Reading Comprehension	.368*
Prior Knowledge	.400**
Vocabulary Knowledge	.208
Pre-test Essay Quality	.249(M)
$p < .10$ (M) Marginal; $p < .05$*; $p < .01$**	

A forward stepwise regression analysis was conducted using only the significant variables from the correlation analyses (i.e., reading comprehension and prior knowledge) to examine the extent to which reading ability and prior science knowledge were predictive of cohesive flexibility as measured using Entropy. One variable was retained in the final model predicting 16% of the variance in students' cohesive flexibility [F (1, 43) = 2.22 p = .006; R^2 = .16]; prior knowledge [B =.40, t(1, 43) = 2.86, p = .006].

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

Additional correlation analyses were conducted to examine the link between students' overall cohesive flexibility score and essay quality during training and at post-test. Results revealed no significant relations between cohesive flexibility and average training essay score (r=.155, p=.31) or post-test essay score (r=.167, p=.28). Thus, though cohesive flexibility was marginally related to essay quality at pre-test (see Table 6), cohesive flexibility as measured by Entropy was not related to overall essay quality during training or at post-test.

These initial results reveal that students' overall cohesive flexibility is related to reading ability and prior science knowledge. To investigate further how reading ability and prior knowledge impact flexibility across time, cumulative cohesive flexibility scores were calculated after each training day. This results in a fine-grained quantification of how cohesive flexibility manifests across time. For example, on Training Day 1, students completed two essays; therefore, their cumulative cohesive flexibility score is the resulting cohesion Entropy score for the Planning and Originality essays. The descriptive statistics for students' cumulative cohesive Flexibility score are provided in Table 7.

Table 7: Descriptive Statistics for Cumulative Cohesive Flexibility Scores

Training Day	M	SD	Range
Day 1	0.27	0.34	(.00–0.69)
Days 1–2	0.64	0.48	(.00–1.32)
Days 1–3	0.95	0.53	(.00–1.39)
Days 1–4	1.01	0.45	(.00–1.33)
Days 1–5	1.18	0.48	(.00–1.67)
Days 1–6	1.23	0.40	(.00–1.66)
Days 1–7	1.21	0.40	(.00–1.64)
Days 1–8	1.25	0.37	(.00–1.62)
N=45			

To examine the relation between students' cumulative cohesive flexibility scores and pre-test measures of their reading ability and prior knowledge, Pearson correlations were calculated for each period (see Table 8). These results show that the relation between students' cumulative cohesive flexibility score and their prior knowledge and reading ability are first detected on day 2 (after 4 essays) and can be reliably detected around day 5 (10 essays written). It is important to note that there was no significant correlation between students' cumulative cohesive flexibility scores and their essay quality or vocabulary knowledge. These results demonstrate that students' reading ability and prior knowledge seem to be mostly related to their cumulative cohesive flexibility score for the later sessions. Thus, the impact that these skills have on flexibility takes time to manifest.

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

Table 8: Correlations between Cumulative Cohesive Flexibility and Reading Ability and Prior Knowledge

Training Day	Reading Ability	Prior Knowledge
Day 1	.124	.056
Days 1–2	.341*	.417**
Days 1–3	.224	.208
Days 1–4	.274(M)	.235
Days 1–5	.381**	.445**
Days 1–6	.376*	.421**
Days 1–7	.388**	.436**
Days 1–8	.368*	.400**

$p < .10$ (M) Marginal; $p < .05$*; $p < .01$**

4 DISCUSSION

A characteristic of strong writers is their ability to demonstrate flexibility in their writing style (Graham & Perin, 2007). These students are said to adapt their writing style to fit the context of the situation. While this is a predominant assumption made by many educators and researchers, very few empirical investigations have examined this claim. Recently, work by Allen, Snow, and McNamara demonstrated that writing flexibility could be measured through the combination of random walk analyses and NLP techniques (2014; in press). The current study replicates and expands upon these initial investigations by using a fine-grained pattern analysis technique, Entropy analysis, to investigate the manifestation of students' narrative flexibility and cohesive flexibility across 16 prompt-based essays. Correlation analyses were used to examine how essay quality, prior ability levels, and writing flexibility were related. Results from this study replicated the work by Allen and colleagues by indicating that students who are more flexible in their use of narrativity have more science knowledge, are better readers, and have a higher vocabulary (Allen, Snow, & McNamara, in press). Similarly, our results also replicate previous work by revealing a positive relation between prior reading ability, prior science knowledge, and cohesive flexibility. Thus, our Entropy analyses and the random walks analyses used by Allen, Snow, and McNamara (2014) seem to be capturing similar variations in students' flexible use of narrativity and cohesion.

Random walks generate a spatial representation of a path (Snow, Likens, Jackson, & McNamara, 2013), providing a visualization and quantification of changes in patterns over time. Thus, random walks analyses result in an aggregated measure of an entire time series (in this case all 16 prompt-based essays). While this method was able to capture *overall fluctuations* in students' use of narrativity and cohesion across time, it does not reveal *when* flexibility could be reliably detected at a more fine-grained level (i.e., after every essay or set of essays). To combat this issue, the current work used a combination of Entropy analyses and NLP techniques to aid in the detection of *when* the relationship between prior skills and writing flexibility (both narrative and cohesive flexibility) becomes evident. The results presented here suggest that Entropy analyses can aid researchers in the measurement of writing

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

flexibility across time (and not just in the aggregate). Indeed, Entropy affords researchers with a nuanced view of how writing flexibility emerges and relates to individual differences that would have otherwise been missed using random walk analyses. The results presented here suggest that Entropy may provide a new metric for analyzing flexibility across time when a limited number of data points are available.

Previous work has shown that reading ability and prior knowledge are related to writing quality, but the current work is the first study to examine how these skills relate to the way a student adapts to various writing contexts across time. The analyses presented here are among the first to use Entropy analyses and NLP to examine this link at a more nuanced level. Our results demonstrate that the relation between students' flexibility in writing style and their prior literacy skills can be detected reliably after only a few essays (depending on the literacy skill and linguistic feature). These results thus inform researchers about individual differences with the potential to influence writing flexibility (both narrative and cohesive flexibility) and when that influence may be detectable. For instance, the current results reveal that the relation between reading comprehension skills and narrative flexibility was detectable after day 2 (four essays). However, the relation between prior knowledge, vocabulary skills, and writing quality with students' narrative flexibility took a bit longer to detect reliably (around day 5). Similarly, results presented here reveal that the relation between cohesive flexibility, reading comprehension skills, and prior knowledge was reliably detectable around day 5 (10 essays written). These finding suggests that flexibility is related to reading comprehension, vocabulary, prior knowledge, and writing ability and that these relations can be detected reliably after only a short time (depending on the literacy skill and linguistic feature investigated) when using an AWE system. This finding is important because it sheds light upon how individual differences relate to (and potentially influence) writing flexibility. Thus, it is not just that students with higher reading ability and more prior knowledge are better at generating text (as shown in previous work), but also that these students adapt their style across time. Indeed, these results support the notion that strong writers are better able to adapt their style based on the writing task and that the unique combination of NLP and Entropy analysis are able to capture *when* these nuanced variations in writing style become evident.

The ultimate goal of this work is develop practical ways to unobtrusively assess students' writing skills and strategies in real-time. Indeed, the results presented here reveal that the combination of linguistic features and dynamic methodologies, such as Entropy, can be used to provide stealth assessments of students' flexibility in writing style. For instance, the stealth assessment of writing flexibility provided here may eventually be able to act as a proxy for more traditional methods of behaviour assessment (i.e., self-reports) thereby improving the student model of our systems and allowing for individualized feedback based on a student's demonstrated level of flexibility. This individualized instruction may inspire students to be more metacognitive about the writing process and aid them in the adaptation of their generated text based on the task and system requirements. It might also point them toward and provide training in specific strategies that they might use related to cohesion and narrativity (e.g., Roscoe & McNamara, 2013).

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

This study is only a first step in the investigation of flexibility in writing, and points toward complex relations between these individual differences. Prior research has examined the characteristics of strong writers. Some of this work has pointed toward individual differences in self-regulation and reading comprehension ability being predictive of writing ability. The current study adds to this literature by revealing a relation among individual differences in literacy skills, essay quality, and writing flexibility (as measured through narrative and cohesive flexibility). Interestingly, narrative flexibility significantly related to overall training performance, whereas cohesive flexibility did not. Perhaps in these particular essay topics, narrative flexibility was more important because students had differing amounts of prior knowledge about the topics as compared to personal experiences or anecdotes related to the topics. The most successful writers may have thus changed their narrativity level depending on how well they were able to provide information and evidence as compared to story-like elements. Future research can examine other genres of writing (e.g., informational) to examine the extent to which narrative flexibility is predictive of writing quality as compared to cohesive flexibility, or if there are genre-specific differences (McNamara, 2013). Another possibility for future research is to explore the degree to which the link between flexibility and essay quality diminishes over time, perhaps due to extended writing practice. Thus, as students gain more writing practice, the quality of their essays may improve regardless of their level of writing flexibility.

These exploratory methodologies have the potential to afford educational researchers a means to quantify students' ability to exert flexibility in their writing style across time. Indeed, the combination of Entropy and NLP provided us with a nuanced examination of writing flexibility that would otherwise be missed using summative measures alone. While these findings are promising, they also serve as only a starting point into the investigation of flexibility in writing. Indeed, a great deal of work focuses on how individual differences can influence writing quality or skills. The current work adds to the literature by showing that, using these dynamic methodologies, we are able to detect how flexibility relates to individual differences and writing skill. Previously, researchers have assumed that flexibility is a trait of successful writers (Graham & Perin, 2007); however, few methodologies afford researchers the opportunity to capture flexibility. Thus, the work and methodologies presented here can inform future studies into the empirical measurement of writing flexibility. In the future, these analyses will be built into the W-Pal AWE system as a means of monitoring writing flexibility as students generate essays. These analyses can also be used to guide formative feedback that instructs students on strategies that may help them adapt their writing style based on the context in which they find themselves.

In conclusion, the current paper is a starting point for researchers interested in the fine-grained assessment of writing flexibility across time. The results presented here reveal that the combination of Entropy analyses and NLP has the potential to capture and quantify changes in students' writing style across time. In the future, these analyses can be implemented within the W-Pal AWE system in real-time as a stealth assessment of the emergence of writing flexibility. Our work serves as a novel methodology to examine the link between linguistic and non-linguistic features in real-time. Indeed, our ultimate goal is to develop and test practical methodologies that can assess skills and abilities related to writing

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

quality in real-time. Thus, eventually the use of these methodologies to assess writing flexibility will have the potential to serve as a proxy for more traditional methods of behaviour and skill assessment.

REFERENCES

Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014).L2 writing practice: Game enjoyment as a key to engagement. *Language Learning and Technology, 18*, 124–150. Retrieved from http://llt.msu.edu/issues/june2014/allenetal.pdf

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, S. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 304–307). Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/short_papers/304_EDM-2014-Short.pdf

Allen, L. K., Snow, E. L., & McNamara, D. S. (in press). The narrative waltz: The role of flexible style on writing performance. *Journal of Educational Psychology*.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® v.2.0. *The Journal of Technology, Learning and Assessment, 4*(3). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1650

Baker, R. S. J. d., Corbett, A. T., Roll, I., & Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction*, *18*, 287–314. http://dx.doi.org/10.1007/s11257-007-9045-6

Biber, D. (1988). *Variation across speech and writing*. Cambridge, UK: Cambridge University Press.

Black, P., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139–148.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 984–989). Austin, TX: Cognitive Science Society.

Crossley, S. A., & McNamara, D. S. (2011). Text coherence and judgments of essay quality: Models of quality and coherence. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1236–1241). Austin, TX: Cognitive Science Society.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading, 35*(2), 115–135. http://dx.doi.org/10.1111/j.1467-9817.2010.01449.x

Crossley, S. A., Roscoe, R., Graesser, A., & McNamara, D. S. (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). http://dx.doi.org/10.1007/978-3-642-21869-9_62

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

Crossley, S. A., Roscoe, R., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication*, *31*(2), 184–214. http://dx.doi.org/10.1177/0741088314526354

Delisle, D., & Delisle, J. (2011). *Building strong writers in middle school: Classroom-ready activities that inspire creativity and support core standards*. Golden Valley, MN: Free Spirit Publishing.

DeRosier, M. E., Craig, A. B., & Sanchez, R. P. (2012). Zoo U: A stealth approach to social skills assessment in schools. *Advances in Human–Computer Interaction*, 2012. http://dx.doi.org/10.1155/2012/654791

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

Fasolo, B., Hertwig, R., Huber, M., & Ludwig, M. (2009). Size, entropy, and density: What is the difference that makes the difference between small and large real-world assortments? *Psychology & Marketing, 26*(3), 254–279. http://dx.doi.org/10.1002/mar.20272

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 & Associates.

Gerrig, R. J. (1993). *Experiencing narrative worlds*. New Haven: Yale University Press.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, *57*, 2333–2351. http://dx.doi.org/10.1016/j.compedu.2011.06.004

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science, 3*(2*)*, 371–398. http://dx.doi.org/10.1111/j.1756-8765.2010.01081.x

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40*(5), 223–234. http://dx.doi.org/10.3102/0013189X11413260

Graesser, A. C., McNamara, D.S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36*(2), 193–202. http://dx.doi.org/10.3758/BF03195564

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. 229–262). Mahwah, NJ: Lawrence Erlbaum Associates.

Graham, S., & Harris, K. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*(1), 3–12. http://dx.doi.org/10.1207/S15326985EP3501_2

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.

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing *evaluation*. *Journal of Technology, Learning and Assessment, 8*(6). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1625/1469

Grossman, E. R. F. W. (1953). Entropy and choice time: The effect of frequency unbalance on choice-response. *Quarterly Journal of Experimental Psychology, 5*(2), 41–51. http://dx.doi.org/10.1080/17470215308416625

Haberlandt, K., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology, 114*(3), 357–374.

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

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, *1*(1), 1–26. Retrieved from http://www.jowr.org/articles/vol1_1/JoWR_2008_vol1_nr1_Kellogg.pdf

Landauer, T. K., Laham, R. D., & Foltz, P. W. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In M. Shermis & J. Bernstein (Eds*.), Automated Essay Scoring: A cross-disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum Publishers.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). Gates-MacGinitie Reading Test (4th ed.). Itasca, IL: The Riverside Publishing Company.

McCutchen, D. (2000). Knowledge acquisition, processing efficiency, and working memory: Implications for a theory of writing. *Educational Psychologist*, *35*(1), 13–23. http://dx.doi.org/10.1207/S15326985EP3501_3

McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: Challenges and successes. *Metacognition and Learning*, *6*(2), 195–203. http://dx.doi.org/10.1007/s11409-011-9082-8

McNamara, D. S. (2013). The epistemic stance between the author and the reader: A driving force in the cohesion of text and writing. *Discourse Studies, 15*(5)*,* 575–592. 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*(1), 57–86. http://dx.doi.org/10.1177/0741088309351547

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

McNamara, D. S., Crossley, S. A., Roscoe, R., Allen, L., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing, 23*(1), 35–59. 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. http://dx.doi.org/10.4018/978-1-60960-741-8.ch011

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

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.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1–43. http://dx.doi.org/10.1207/s1532690xci1401_1

National Center for Education Statistics. (2011). The nation's report card: Trial urban district assessment science 2009 (NCES 2011–452). Washington, D.C.: U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/dst2009/2011452.pdf

Newkirk, T. (2012). How we really comprehend nonfiction. *Educational Leadership*, *69*(6), 29–32.

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.

Page, E. B. (2003). Project essay grade: PEG. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 43–54). Mahwah, NJ: Lawrence Erlbaum Associates.

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*(4), 1010–1025. http://dx.doi.org/10.1037/a0032340

Roscoe, R. D., Snow, E. L., Brandon, R. D., & McNamara, D. S. (2013). Educational game enjoyment, perceptions, and features in an intelligent writing tutor. In C. Boonthum-Denecke & G. M. Youngblood (Eds.), *Proceedings of the 26th Annual Florida Artificial Intelligence Research Society Conference* (FLAIRS-13) (pp. 515–520). Menlo Park, CA: The AAAI Press.

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*(4), 362–381. http://dx.doi.org/10.1504/IJLT.2013.059131

Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of the IntelliMetricTM essay scoring system. *Journal of Technology, Learning, and Assessment*, *4*(4). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1651

Shanahan, T., & Tierney, R. J. (1990). Reading–writing connections: The relations among three perspectives. In J. Zutell & S. McCormick (Eds.), *Literacy theory and research: Analyses from multiple paradigms* (39th Yearbook of the National Reading Conference) (pp. 13–34). Chicago: National Reading Conference.

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell system technical journal, 30*(1), 50–64. http://dx.doi.org/10.1002/j.1538-7305.1951.tb01366.x

Sharma, N., & Patterson, P. G. (1999). The impact of communication effectiveness and service quality on relationship commitment in consumer, professional services. *Journal of Services Marketing, 13*(2), 151–170. http://dx.doi.org/10.1108/08876049910266059

(2015). Keys to detecting writing flexibility over time: Entropy and natural language processing. *Journal of Learning Analytics*, *2*(3), 40–54. http://dx.doi.org/10.18608/jla.2015.23.4

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte, NC: Information Age Publishers.

Shute, V. J., & Kim, Y. J. (2013). Formative and stealth assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed.), (pp. 311–323). New York: Lawrence Erlbaum Associates/Taylor & Francis Group.

Snow, E. L., Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2014). Does agency matter? Exploring the impact of controlled behaviors within a game-based environment. *Computers & Education, 82*, 378–392. http://dx.doi.org/10.1016/j.compedu.2014.12.011

Snow, E. L., Allen, L. K., Jacovina, M. E., Perret, C. A., & McNamara, D. S. (2015). You've got style! Detecting writing flexibility across time. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 194–202. http://dx.doi.org/10.1145/2723576.2723592

Snow, E. L., Jackson, G. T., & McNamara, D. S. (2014). Emergent behaviors in computer-based learning environments: Computational signals of catching up. *Computers in Human Behavior, 41*, 62–70. http://dx.doi.org/10.1016/j.chb.2014.09.011

Snow, E. L., Jacovina, M. E., Allen, L. K., Dai. J., & McNamara, D. S. (2014). *Entropy: A stealth assessment of agency in learning environments.* In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 241–244). Retrieved from http://www.educationaldatamining.org/proceedings

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). Retrieved from http://www.educationaldatamining.org/proceedings

Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology*, *63*(2), 358–385. http://dx.doi.org/10.1006/jecp.1996.0054

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, *10*(2), 157–180. http://dx.doi.org/10.1191/1362168806lr190oa

Witte, S., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication, 32*(2), 189–204. http://dx.doi.org/ 10.2307/356693