The Long and Winding Road: Investigating the Differential Writing Patterns of High and Low Skilled Writers

Laura K. Allen
Tempe, AZ, USA
Arizona State University
LauraKAllen@asu.edu

Erica L. Snow
Tempe, AZ, USA
Arizona State University
Erica.L.Snow@asu.edu

Danielle S. McNamara
Tempe, AZ, USA
Arizona State University
Danielle.McNamara@asu.edu

ABSTRACT
We investigate how writing proficiency relates to the flexible use of cohesion. Forty-five students wrote 16 essays across 8 sessions. Natural language processing techniques were used to calculate the cohesion of each essay. Random walk and Euclidian distance measures were then used to visualize and classify students' flexibility in cohesion across the essays. Results revealed that students who were more flexible in their cohesion also had greater literacy skills and prior knowledge. Further, cohesive flexibility was most strongly related to the *unity* of the pretest essays.

Keywords
Intelligent Tutoring Systems, dynamical analysis, writing, flexibility, cohesion, automated essay scoring

1. INTRODUCTION
Students' ability to effectively communicate via writing has been shown to be a critical skill for academic and professional achievement. Standardized tests, for instance, typically require students to complete a single assignment that is designed to tap into their proficiency at writing. This assessment has a profound impact on college acceptance and other opportunities, such as scholarships, honors organizations, and assistantships [1].

Unfortunately, teachers do not have the time to provide thorough feedback on every essay a student generates. In response to these needs, researchers have developed adaptive computerized systems designed to assess the quality of essays [2]. Automated essay scoring (AES) systems employ natural language processing (NLP) and statistical methods to evaluate the structure, content, and holistic quality of written text [2-3]. Although the validity of these scores has been questioned [4], AES systems tend to calculate automated scores that are comparable to human scores [5].

AES systems have been recently integrated into learning environments, such as automated writing evaluation (AWE) systems [6] and intelligent tutoring systems (ITSs) [7]. These environments emphasize the provision of instruction and formative feedback based on the quality and specific characteristics of students' writing. This has presented a number of problems regarding the ability of the algorithms to provide specific and formative feedback that is beneficial to students [8].

The above categories of systems (AES, AWE, and ITSs) tend to rely on text-level features of *individual* essays to assess writing ability. Although essay scores are generally comparable to those provided by humans, they rarely incorporate information about the students themselves (e.g., their skills, affective states, etc.) into the system feedback or scoring algorithms. Additionally, the systems place little to no focus on the writing style of individual students. In other words, they do not take into consideration the possibility that high-quality essays may exhibit different textual properties across multiple writers.

Researchers have identified a number of linguistic features that are associated with writing quality [9-13]. Through the use of NLP tools, these indices can be automatically calculated and combined to develop algorithms for essay scoring. Recently, Coh-Metrix was used to examine which linguistic features were capable of discriminating between high- and low-quality essays [11]. The results revealed that high-quality essays included more diverse and novel word choices and more complex syntax. Interestingly, no indices of cohesion were related to essay scores. Crossley and colleagues (2011) conducted an analysis using similar indices to predict essay scores. In contrast to the previous results, this study found that essay quality was *positively* associated with cohesion. These mixed findings indicate that writing proficiency is a more complex and dynamic construct than previously assumed. Specifically, this skill may not be adequately captured by a single writing sample, as linguistic properties associated with essay quality vary across multiple contexts, such as writer populations, time constraints, and prompts [9,14-15].

We hypothesize that writing proficiency is associated with a *flexible* use of linguistic properties, rather than a fixed set of features. For example, certain prompts and contexts may require different levels of cohesion to effectively convey the main idea. In this case, strong writers may have the ability to assess the context of their writing task and flexibly employ different cohesive devices that match the evidence and arguments presented in that specific essay, whereas less skilled writers might not have developed the strategies necessary to vary their style across different contexts. Researchers have cited flexibility as a characteristic of strong writers [2]. However, few studies (if any) have explicitly measured writing flexibility and examined its relation to writing skills and other individual differences. We address this gap by investigating how writing proficiency relates to students' flexible use of cohesion across various prompts, and examine how individual differences relate to this flexibility.

2. METHODS
The data was collected as part of a larger study, which compared students' use of a writing strategy ITS to an AWE component of the system. We focus on the participants who engaged with the

AWE component of the system (n = 45). Students completed a 10-session experiment. During the first session, students completed a pretest. Training occurred during the following eight sessions. Throughout each training session, students wrote two essays, each on a different prompt topic. Thus, 16 training essays were collected for each student. During session 10, students completed a posttest, which was similar to the pretest.

2.1 Measures

Students' writing proficiency was assessed at pretest and posttest through the use of timed (25-minute) and counterbalanced prompt-based essays. All essays were assessed on a scale of 1-6 by two expert raters. The holistic grading rubric was based on a standardized rubric typically used for the assessment of Scholastic Achievement Test (SAT) essays. The rubric contained subscale scores, which assessed the quality of various sections of the essay. These subscales related to the following aspects of the essay: effective lead, clear purpose, clear plan, topic sentences, paragraph transitions, organization, unity, perspective, conviction, and grammar, syntax and mechanics.

Reading comprehension ability and vocabulary knowledge were assessed using the Gates-MacGinitie reading skill test. Students' prior knowledge was assessed using a measure of prior knowledge that assessed knowledge of science, literature, and history.

Coh-Metrix was used to assess the cohesion of the students' 16 essays. Coh-Metrix [16] is a computational text analysis tool that was developed, in part, to provide stronger measures of text difficulty. This tool includes *Easability Components,* which were developed to account for the multiple dimensions of text difficulty. Referential cohesion is one of the Easability Components and reflects the degree to which words and ideas overlap across a text. For each of the 16 essays, a *referential cohesion percentile score* was computed on a scale from 0 to 100.

3. QUANTITATIVE METHODS

To assess the flexibility of students' use of cohesion, we employed NLP and dynamical methodologies. The unique combination of these methodologies affords researchers a new assessment technique that can visualize and capture the degree to which students exhibit a controlled or flexible writing style.

Table 1. Referential cohesion classification and vector assignment

Essay Cohesion Level	Axis Direction Assignment
Less than 25% Referential Cohesion	-1 on X-axis (move left)
Between 25% and 50% Referential Cohesion	+1 on Y-axis (move up)
Between 50% and 75% Referential Cohesion	+1 on X-Axis (move right)
Greater than 75% Referential Cohesion	-1 on Y-axis (move down)

Random walks are mathematical analyses that provide a spatial representation of patterns that form in categorical data across time [17]. Random walks were used to visualize patterns in students' use of cohesive devices across their 16 training essays. Each essay was classified as one of four orthogonal Cohesion Level groups (see Table 1) using the Coh-Metrix *referential cohesion percentile score* (ranging from 0-100). These orthogonal categories were then assigned to individual vectors along a scatter plot. For instance, an essay that received a referential cohesion score below 25 would be assigned to the vector (-1,0), along the left side of the X-axis. Each student's random walk began at the origin. For each essay that the student wrote, the walk would move in the direction that was consistent with its assigned vector. The resulting walk depicts each student's use of cohesion across the training essays.

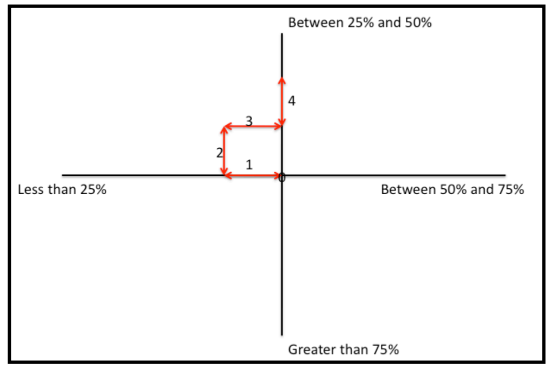

Figure 1. Example Random Walk

Figure 1 illustrates what a random walk might look like for a student who wrote 4 essays. All walks begin by placing a dot at the origin. In this example, the first essay written was low in referential cohesion (score < 25); thus, the dot moved one step left along the X-axis (#1). The next essay received a referential cohesion score between 25-50; thus, the dot moved up along the Y-axis (#2). The third essay had a referential cohesion score that ranged from 50-75, so the dot moved one step right along the X-axis (#3). The last essay received a cohesion score that ranged from 25-50; so, the dot moved one step up (#4). Using these rules, unique random walks were generated for each of the students.

To quantify the information in the random walk visualizations, distance time series were calculated for each student using Euclidian distance. Here, y represents the particle's position on the Y-axis, x represents the particle's position on the X-axis, and i represents the ith step in each student's walk.

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

A measure of Euclidian distance was calculated for each step in a student's walk. This produced a distance time series, which reflected the degree to which students were flexible in their use of cohesion. For example, if a student used the same degree of cohesion throughout all 16 essays, that student would move far away from the origin, resulting in a high Euclidian distance score. Conversely, if a student varied a great deal in the use of cohesion, the resulting Euclidian distance for their walk would be lower, as their changes would cause them to remain close to the origin.

4. RESULTS

We calculated the average Euclidian distance of each student's walk (their *cohesion distance score*). Students varied considerably in their flexibility, ranging from a minimum cohesion distance score of 1.42 to a maximum score of 8.50 (M = 5.04, SD = 1.88). In Figure 2, each student's cohesion distance score is plotted to visualize the variation in the degree to which walks traveled.

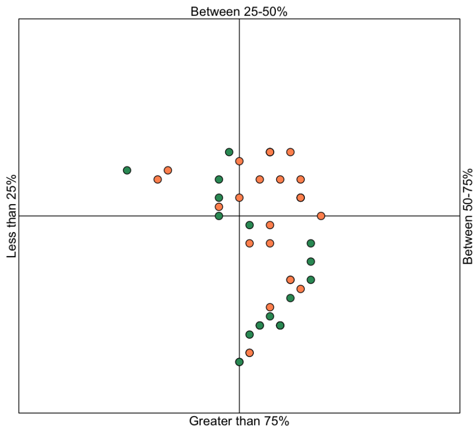

Figure 2. Visualization of Low-Skilled and High-Skilled Students' Random Walks

We next examined the degree to which cohesive flexibility varied according to students' writing proficiency. A median split was calculated on students' pretest essay scores to produce two groups: *low writing ability* and *high writing ability*. A between-subjects ANOVA revealed that high writing ability students had significantly lower cohesion distance scores (M = 4.49, SD = 1.30) compared to low writing ability students (M = 5.80, SD = 2.10), F (1, 42) = 6.28, p = .016. Figure 2 provides a visualization of these differences, with low writing ability students represented as green dots and high writing ability students represented by pink dots. As revealed in this visualization, low writing ability students (green dots) moved further from the origin than high writing ability students (pink dots) who clustered closer to the origin.

4.1 Essay Components

The correlation between cohesion distance scores and pretest holistic essay scores was marginally significant (r = -.30, p = .052), suggesting that students who were more flexible were more proficient writers. Additionally, distance scores were related to a number of the subscale scores on the writing rubric (see Table 2).

Table 2. Correlations between Cohesion Distance Scores and Rubric Subscales

Rubric Subscale	*r*
Lead	.02
Purpose	-.29 (M)
Plan	-.29 (M)
Use of Topic Sentences	-.27 (M)
Transitions	-.20
Organization	-.26 (M)
Unity	-.42**
Perspective	-.35*
Persuasion	-.40**
Accuracy	-.29 (M)

(M) = Marginal Significance; * = p < .05; ** = p < .01

To determine which rubric subscale scores were most predictive of writing flexibility, we conducted a stepwise regression analysis with the significantly correlated subscale variables as predictors of cohesion distance scores. One variable was retained in the final model and predicted 17% of the variance in distance scores [F (1, 42) = 8.85, p = .005; R^2 = .17]: Unity [B = -.417, t(1, 42) = -2.98, p = .005]. Overall, these results suggested that students who produced more *coherent* and unified ideas were the students who exhibited greater flexibility in their use of cohesion, or cohesive cues.

4.2 Individual Differences

Students' cohesion distance scores were significantly (or marginally significantly) related to a number of pretest measures (see Table 3).

Table 3. Correlations between Cohesion Distance Scores and Individual Difference Measures

Individual Difference Measure	*r*
Reading Comprehension	-.44**
Vocabulary Knowledge	-.19
Prior Knowledge (Overall)	-.31*
Science Prior Knowledge	-.50**
History Prior Knowledge	-.11
Literature Prior Knowledge	.16

(M) = Marginal Significance; p < .05*; p < .01**

To examine which of the individual difference measures were the most predictive of cohesive flexibility, we conducted a stepwise regression analysis including the significantly correlated variables as predictors of cohesion distance scores. One variable was retained and predicted 25% of the variance in cohesion distance scores [F (1, 43) = 14.59, p < .001; R^2 = .25]: Science Prior Knowledge [B = -.50, t(1, 42) = -3.82, p < .001]. Students who entered the writing task with greater knowledge about the world may have had an easier time adapting their writing style, as they could utilize various facts to develop their arguments.

5. DISCUSSION

One important consideration when assessing writing proficiency is the flexibility that students exert in their writing style across time. Although individual essay scores can provide valuable information about writing skills, they fail to consider the context of the writing assignments and consequently are not able to fully capture the construct of writing proficiency.

We were able to capture cohesive flexibility through the use of two novel techniques: random walks and Euclidian distances. Random walk analyses allowed us to visualize students' rigid or flexible use of cohesion across the essay assignments; additionally, it allowed us to visualize the differential patterns exhibited by high- and low-ability writers. Euclidian distance scores were then used to calculate *cohesion distance scores*. These scores confirmed the results of the random walk visualizations. In particular, they revealed that that students varied considerably in their cohesive flexibility, with low-ability students showing more consistency in their use of cohesion than high-ability students.

The results support our hypotheses and provide evidence for assumptions that have been only anecdotally raised in the writing literature [2]. Namely, they suggest that students who are more flexible in their writing style are also better writers, and vice versa. Additionally, these students outperform less flexible students on measures of literacy skills and prior knowledge. The results also suggest that cohesive flexibility was most strongly related to the *unity* (i.e., the coherence) of the pretest essay. Thus, coherence was not directly related to the presence or absence of cohesive features. Rather, students who produced more coherent essays were more flexible in their use of cohesive devices. Taken together, these results indicate that the link between textual features and writing quality may be inconsistent from assessment to assessment. Thus, more sophisticated writing assessments are needed to capture students' proficiency within various contexts.

This study extends previous work suggesting that the link between textual features and essay scores can vary across a number of contexts [9,14-15] by considering students' performance across time. Although these analyses make strong progress towards developing our understanding of writing flexibility, a number of questions remain to be answered. For instance, how does flexibility of other linguistic features (e.g., narrativity) relate to writing proficiency? Can students be trained to exhibit greater flexibility in their writing? Such analyses will help shed light on the role that flexibility plays in the development of writing proficiency. Overall, this study provides a critical insight into the complexity of automated writing evaluation and provides a novel method for providing essay scores and feedback that are more sensitive to the surrounding context of the writing assessment.

6. ACKNOWLEDGEMENTS

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080589 to Arizona State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

7. REFERENCES

[1] Graham, S., and Perin, D. 2007. A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, (2007), 445-476.

[2] Dikli, S. 2006. An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5, (2006).

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

[4] Wang, J., and Brown, M. S. 2007. Automated essay scoring versus human scoring: A comparative study. *Journal of Technology, Learning, and Assessment, 6,* (2007), 46.

[5] Warschauer, M., and Ware, P. 2006. Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research, 10*, (2006), 1-24.

[6] Crossley, S. A., Varner, L. K., Roscoe, R. D., and McNamara, D. S. 2013. Using automated indices of cohesion to evaluate an intelligent tutoring system and an automated writing evaluation system. In K. Yacef et al (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED).* Springer, Heidelberg, Berlin, 269-278.

[7] Roscoe, R. D., Snow, E. L., Varner, L. K., and McNamara, D. S. in press. Automated detection of essay revising patterns: Application for intelligent feedback in a writing tutor. *Technology, Instruction, Cognition, and Learning.*

[8] Roscoe, R. D., Varner, L. K., Cai, Z., Weston, J. L., Crossley, S. A., and McNamara, D. S. 2011. Internal usability testing of automated essay feedback in an intelligent writing tutor. In R. C. Murray & P. M. McCarthy (Eds.), *Proceedings of the 24th International Florida Artificial Intelligence Research Society (FLAIRS) Conference*. AAAI Press, Menlo Park, CA, 543-548.

[9] Crossley, S. A., Allen, L. K., & McNamara, D. S. in press. 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 *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber.*

[10] Witte, S. P., and Faigley, L. 1981. Coherence, cohesion and writing quality. *College Composition and Communication, 22,* (1981), 189-204.

[11] McNamara, D. S., Crossley, S. A., and McCarthy, P. M. 2010. The linguistic features of quality writing. *Written Communication, 27*, (2010), 57-86.

[12] Varner, L. K., Roscoe, R. D., and 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,* (2013), 35-59.

[13] Crossley, S. A., Roscoe, R. D., McNamara, D. S., and Graesser, A. 2011. Predicting human scores of essay quality using computational indices of linguistic and textual features. In G. Biswas, S. Bull, J. Kay, and A. Mitrovic (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education*. Springer, New York, 438-440.

[14] Crossley, S., Weston, J., Sullivan, S., and McNamara, D. 2011. The development of writing proficiency as a function of grade level: a linguistic analysis. *Written Communication. 28*, (2011), 282-311.

[15] Crossley, S. A., Varner, L. K., and 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*. The AAAI Press, Menlo Park, CA, 202-207.

[16] McNamara, D. S., Graesser, A. C., McCarthy, P., and Cai, Z. 2014. Automated evaluation of text and discourse with Coh-Metrix. Cambridge: Cambridge University Press.

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