# Coordinated Symposium: NCME 2018 Panel on Writing Analytics

Jill Burstein, *Educational Testing Service*

Writing analytics, which includes a number of computation methods for the study of texts, helps us to study genres and processes as they naturally occur in digital educational environments. Within the emerging field of writing analytics, automated writing evaluation (AWE), while its history is instantiated in scoring of essays, can also provide a computational analysis of construct-relevant linguistic characteristics (that is, features) in writing which can provide evidence of a writer's process and proficiency. AWE, thus as a subset of writing analytics approaches, provides a lens with which to study the relationship between aspects of the socio-cognitive writing achievement framework—writing and knowledge domain, and intra- and inter-personal skills—and broader success outcomes that may help us to understand complex educational processes such as college retention and completion (Allen, Snow, Crossley, Jackson & McNamara, 2014; Burstein, McCaffrey, Beigman Klebanov, & Ling, 2017).

At the 2018 Annual Meeting of the National Council for Educational Measurement (NCME) a Coordinated Symposium featured a panel on this topic: *What Writing Analytics Can Tell Us About Broader Success Outcomes* (Burstein & McCaffrey, 2018). The session description was as follows:

> Writing is a challenge, especially for at-risk students who may lack the prerequisite writing skills required to persist in U.S. 4-year postsecondary institutions. Educators teaching postsecondary courses that require writing could benefit from a better understanding of writing achievement as a socio-cognitive construct (including writing domain knowledge, general knowledge, and intra- and inter-personal factors). They would also benefit from understanding its role in

postsecondary success vis-a-vis retention and college completion. While there is a long tradition of research related to essay writing on standardized tests and college success, typically only the final overall essay score is used for decision-making. This session examines relationships between finer-grained writing analytics and outcomes. We define writing analytics as the study of processes, language use, and genres as they naturally occur in digital educational environments. The five presentations will address these research questions: (1) How do writing analytics derived from student writing samples relate to measures of broader outcomes?, and (2) How might these relationships between writing analytics and broader outcomes inform instruction and assessment to advance student learning? The presentations will address analytics from a variety of writing genres that reflect on-demand, authentic, and intervention-based writing assignments from 2- and 4-year college students.

To address these questions, five papers were presented at the session:

1. "Linking Writing Analytics and Broader Cognitive and Intrapersonal Outcomes," Jill Burstein, Educational Testing Service; Daniel McCaffrey, Educational Testing Service; Beata Beigman Klebanov, Educational Testing Service; Guangming Ling, Educational Testing Service

2. "*MyReviewers*: Understanding How Feedback Supports Writers in Higher Education," Val Ross, University of Pennsylvania; Joe Moxley, University of South Florida; Dave Eubanks, Furman University

3. "Writing Mentor™: Writing Progress Using Self-Regulated Writing Support," Beata Beigman Klebanov, Educational Testing Service; Jill Burstein, Educational Testing Service; Nitin Madnani, Educational Testing Service; Norbert Elliot, University of South Florida

4. "Writing Analytics and Formative Assessment," Charles MacArthur, University of Delaware; Zoi Phillipakos, University of North Carolina, Charlotte; Amanda Jennings, University of Delaware

5. "Utility Value and Writing Analytics," Stacy Priniski, University of Wisconsin, Madison; Beata Beigman Klebanov, Educational Testing Service; Jill Burstein, Educational Testing Service; Judith Harackiewicz, University of Wisconsin, Madison; Dustin Thoman, San Diego State University

With Daniel McCaffrey, I served as chair of the session, and Mya Poe served as discussant. Paper 1 and Paper 4 were pledged to other publications. Paper 2 is under development, but some of its content representing an on-going program of research may be found in Ross and LeGrande (2017). Paper 3 and Paper 5 are presented in this special section of *The Journal of Writing Analytics* that follows. Here, we bring to light the expanded value of AWE-generated writing analytics and implications for measuring broader outcomes.

The two papers that follow address analytics drawn from a variety of writing genres that reflect on-demand, authentic, and intervention-based writing assignments, representing writer populations from middle and high school through 2- and 4-year college settings. "Writing Mentor™: Writing Progress Using Self-Regulated Writing Support," authored by Burstein et al., describes a novel writing instruction application designed for struggling writers in postsecondary settings. The Writing Mentor (WM) app provides automated writing feedback using natural language processing methods and linguistic resources. Features support a breadth of writing constructs, including use of sources, claims, and evidence; coherence and topic development; organization; knowledge of written English conventions; and vocabulary choice. The app takes initial steps to integrate the reading and writing process by offering vocabulary support, intended to be used for help with unfamiliar vocabulary in course sources (texts). The app also elicits information on writer self-efficacy by asking users optional related questions, and questions also ask users about their English language status as part of an entry survey. The app maintains a user event log that captures de-identified user writing samples and feature use which can be used for research. The paper discusses an Amazon Mechanical Turk (AMT) task situated in the *Writing Mentor* in which 108 AMT workers responded to a writing task and used the app to revise their writing. As the app is accessible to the general population and event log data are collected for users-in-the-wild, the paper discusses real-world use of features and submission revision, as well as relationships between feature use, self-efficacy, and English language status. Present work suggests that WM holds the potential to meet the challenges of formative assessment and personalized learning in terms of improving student writing.

"Utility-Value Score: A Case Study in System Generalization for Writing Analytics," authored by Beigman Klebanov et al., discusses utility-value (UV) writing interventions intended to keep college students interested in science courses as a crucial means to retaining them in STEM majors and keeping them on track for STEM careers (Durik & Harackiewicz, 2007; Hidi & Harackiewicz, 2000). Natural language processing and other linguistic resources (Pennebaker et al., 2015) have been used to extract linguistic features and successfully build models that predict the presence of language associated with UV in writing samples (Beigman Klebanov et al., 2016; Beigman Klebanov et al., 2017). Building on previous work, the article describes and explores the generalizability of UV models that are built using AWE resources and machine learning. The research leverages a new writing intervention data collection from two independent UV writing intervention studies conducted at California State University, Long Beach (CSULB), University of Wisconsin, Madison (UW), and at several two-year campuses of the University of Wisconsin Colleges system. At CSULB, writing data were collected from biology, chemistry, and physics courses; at UW, writing data were collected across biology and related STEM courses from seven 2-year institutions in Wisconsin. The paper discusses generalizability of the models and how UV analytics may be identified and used to provide relevant scaffolding for students showing lower or no UV language in their responses. As the authors note, while it is tempting to think about an automatically produced utility-value score as a meaningful analytic on a large collection of essays, this kind of score creation is not without challenges.

In her role as discussant in her reflective paper that follows, Poe observes that answers to the two questions that informed the session—concern with the relationship between writing analytics and broader outcomes, and the ways that these relationships inform instruction and assessment to advance student learning—emerged in unexpected ways. If automated methods of evaluation are framed as narrowly-conceived technical analyses, as has been done in previous iterations of AWE for scoring purposes, then they offer little advancement of the pedagogical value of AWE. Given the panel's integration of socio-cognitive frameworks that integrate technical and social concerns, Poe notes, newer iterations of AWE reflect the possibilities of what sustained engagement with varied frameworks for understanding and teaching writing can yield.

## Author Biography

**Jill Burstein** is a Director of Research for the Natural Language Processing Group in the Research Division at Educational Testing Service in Princeton, New Jersey. She has led the NLP teams that invented e-rater®, an automated essay scoring system used in large-scale assessment, and the Writing Mentor™ application—a Google Docs Add-on. She co-led the development of the _Language Muse_® _Activity Palette_—an NLP-driven activity-generation tool for classroom teachers. Motivated by an interest to make a social contribution through her work, she has been working in educational data mining to explore NLP- based writing analytics. This work facilitates generation and evaluation of analytics to build a deeper understanding of student writing, stakeholder (educators and policy-makers) understanding of writing achievement, and relationships between writing achievement and success predictors. She works collaboratively with researchers in writing studies and has produced research relevant to that community through long-term collaborations.

## References

Allen, L. Snow, E., Crossley, S., Jackson, G.T., & McNamara, D. (2014). Reading comprehension components and their relation to writing. _L'Année Psychologique 114_(4), 663–91.

Beigman Klebanov, B., Burstein, J., Harackiewicz, J., Prinski, S. & Mullholland, M. (2016). Enhancing STEM motivation through personal and communal values: NLP for assessment of utility value in student writing. In _Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)_, NAACL 2016, San Diego, CA, USA.

Beigman Klebanov, B., Burstein, J., Harackiewicz, J., Prinski, S. & Mullholland, M. (2017). Reflective writing as a tool for increasing STEM motivation and retention—Can AI help scale it up? _Special Issue for the International Journal of Artificial Intelligence in Education: MARWIDE: Multidisciplinary Approaches to Reading and Writing Integrated with Disciplinary Education_ (Eds. D. McNamara, S. Muresan, R. J. Passonneau, & D. Perin), _27_(4), 791–818.

Burstein, J., & McCaffrey, D. (2018, April). _What writing analytics can tell us about broader success outcomes._ National Council for Measurement in Education, Coordinated Session A4. New York, NY.

Burstein, J., McCaffrey, D., Beigman Klebanov, B., & Ling, G. (2017). Exploring relationships between writing and broader outcomes with automated writing evaluation. In _Proceeding of the 12th_

*Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*, EMNLP 2017, Copenhagen, Denmark.

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How personal interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology, 99*, 597–610.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70,* 151–179.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC 2015*. Austin, TX: Pennebaker Conglomerates.

Ross, V., & LeGrande, R. (2017). Assessing writing constructs: Toward an expanded view of inter-reader reliability. *Journal of Writing Analytics*, *1*, 227–275. Retrieved from https://journals.colostate.edu/analytics/article/view/141