

M-Write: A Large-Scale Laboratory for Writing Analytics Research

**Anne R. Gere, Ginger Shultz, Chris Teplovs, and Dave Harlan
University of Michigan**

In the scholarly work of higher education and educational psychology, writing is identified as one of a set of high impact learning activities (Kuh, 2008) that are positively correlated with student engagement (Kuh & Kuh, 2001; Light 2001; Carini, 2006) and thereby greater student gains in critical thinking, (Kuh & Kuh, 2001; Carini, 2006), problem solving, and effective communication. Engaged learning in the context of writing means that students apply psychological energy to meaningful processing of content in order to draft writing about discipline-specific subject matter (Schreiner, 2011). Engagement with writing also means that students more frequently connect to faculty and peers because the more frequently students write, the more they receive feedback, and as a consequence the more often they are likely to communicate with faculty and peers in the classroom (Kuh & Kuh, 2001).

The primary challenge of incorporating writing into instruction is that faculty need to read, assess, and provide feedback, which is problematic in large-enrollment classes. This practical constraint contributes to a paucity of writing in large-enrollment gateway courses, and these are the very courses where students are most likely to depart from their intended path (Seidman, 2005). The overarching goal of the M-Write project is to infuse writing into large gateway courses across campus. In doing so, M-Write provides a large-scale laboratory to support writing analytics. If this proposal is accepted we will bring to the workshop three members of the M-Write team: (1) a pedagogic and writing expert (Gere) who will focus on enhancing learning for diverse populations underrepresented in STEM, (2) a learning analytics expert (Teplovs) who will address aspects of the machine learning models that we have developed, and (3) a technology expert (Harlan) who will detail the analytic pipeline that underpins the M-Write project. Regrettably, one of our team members (Shultz) has a scheduling conflict and won't be able to attend the workshop. Specifically, we will describe progress on an analytics system that currently does two things: (1) provides automated support of peer review of student writing, and (2) uses tools of automated text analysis to give faculty actionable data (drawn from student writing) about student learning.

Automated peer review of student writing supports student learning with content-focused writing assignments, offers students individualized feedback, enables students to learn from one another's written responses, and provides faculty with information about the nature of student engagement in the peer review process. We will briefly describe the architecture we have built and highlight the importance of designing for analytics at the core of the project rather than as an add-on.

Automated Text Analysis (ATA), which includes Natural Language Processing and Latent Semantic Analysis, is used to analyze students' writing samples and produce actionable data that inform faculty about student learning. ATA, for the purposes of this proposal, refers to the collection of tools and techniques with which computers analyze writing produced by students. The ATA techniques we employ focus on applying algorithms and other computational tools to relatively large collections of texts (corpora) to identify specific features in the writing. For example, natural language processing (via the NLTK package in Python) is used to generate information about syntactic structures such as sentence complexity and coherence among sentences. Latent Semantic Analysis provides data about lexical items or the vocabulary used in selections of writing. Together these tools are used to analyze thousands of student responses to writing prompts to identify patterns of language that indicate whether or not students can describe in writing key concepts of a given course. When combined with supervised

machine learning techniques based on human experts, these techniques provide a solid foundation for the creation of a powerful writing analytics framework.

Our team members' diverse perspectives neatly map onto the theme of the workshop: bridging from research to practice and collectively can play a pivotal role in the emerging sub-community of writing analytics. We are, of course, open to a variety of roles in the workshop itself. One of the participants (Teplovs), is connected to the Software Carpentry initiative and can help with the development of a Learning Analytics Carpentry approach in SoLAR.

Brief Bios of Workshop Attendees

Anne Ruggles Gere holds an appointment in the Department of English as well as in the School of Education, she directs the Sweetland Center for Writing, and she is chair of the Joint PhD Program in English and Education. Her research interests include literacy, composition studies, and disciplinary writing. She regularly teaches courses on composition studies, assessment, and literacy practices. Her current projects include a study of disciplinary writing. She is incoming president of the Modern Language Association, has served as president of the National Council of Teachers of English, as chair of the Conference on College Composition and Communication, and as a member of the Executive Council of the Modern Language Association. She has been honored with the Regent's Award for Distinguished Public Service; the D'Arms Award for Distinguished Graduate Student Mentoring; the Distinguished Faculty Achievement Award; and in 2010, she was named an Arthur F. Thurnau Professor in recognition of her excellent teaching of undergraduates. She has received grants from the Spencer Foundation, the National Endowment for the Humanities, and the U.S. Department of Education. Her recent publications include *Writing on Demand for the Common Core State Standards*, and articles in *College Composition and Communication*, *Assessing Writing*, and *The Journal of Chemical Education*.

Chris Teplovs leads the development on the design and development of digital educational applications housed within the Digital Innovation Greenhouse (DIG). He leverages his expertise in research-based educational ideas and best practices for development processes and applies them to scaling up digital innovations. Additionally, he provides support for funding proposals and continues to delve further into learning analytics and data visualization. Prior to joining DIG, Chris used his experience as a postdoctoral fellow at the Copenhagen Business School to create several startup ventures focused primarily on analytics. His doctoral work focused on the design and development of a tool for visualizing knowledge spaces. Chris's extensive background includes experience in subjects such as automated text analysis, machine learning, statistics and Big Data.

Dave Harlan works as an application developer with the Digital Innovation Greenhouse (DIG) in collaboration with Sweetland Center for Writing on M-Write, a digital toolkit supporting writing-to-learn pedagogies at scale in order to enhance student learning and provide individualized feedback for students in large-format gateway courses. After receiving his undergraduate degree, Dave performed a number of roles as a consultant for IBM Global Business Services in the finance and telecommunication industries. He then transitioned to a role as an application developer for U-M's Information and Technology Services, where he worked to extend and maintain hybrid and native mobile apps for U-M using a variety of technologies.

References

- Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of social and clinical psychology*, 4(3) 359-373.
- Blei, D. Topic modeling and digital humanities. *Journal of Digital Humanities*, 2(1), 2013.
- Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141-178.
- Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in higher education*, 47(1), 1-32.
- Carter, M., Ferzli, M., Wiee, E (2007). Writing to learn by learning to write in the disciplines. *Journal of Business and Technical Communication*. 21: 278-302.
- Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891.
- Collins-Thompson, K. and Callan. J. (2007). Automatic and human scoring of word definition responses. *Proceedings of the NAACL-HLT 2007 Conference*. Rochester, U.S.A. pp. 476-483.
- Crismond, D. (2001). Learning and using science ideas when doing investigate-and-redesign tasks: a study of naïve, novice and expert designers doing constrained and scaffolded design work. *Journal of Research in Science Teaching* 38.7: 791-820.
- Falchikov, N., and Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research* 70(3): 287-322.
- Gere, A.R., Aull, L., Lancaster, C. Perales, Escudero, M., & VanderLei, E. (2013). Local assessment: Using genre studies to validate directed self-placement. *College Composition and Communication* 64.4:605-633.
- Graesser, A., McNamara, D.S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36, 193-202.
- Hayes, R. Q., Whalen, S. K. and Canon, G. (2009). CSR D STEM retention report. Center for Institutional Data Exchange and Analysis, University of Oklahoma, Norman.
- Henderson, C., Cole, R., Froyd, J., & Khatri, R. (2012). Five Claims about Effective Propagation. White paper accessed from <http://stem.vcsu.edu/blog/five-claims-about-effective-propagation/> on Aug 30, 2015.
- Hmelo-Silver, C.E, Pfeffer, M.G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science* 28.1: 127-138.
- Kozma, R. B. (1997). Multimedia and understanding: Expert and novice responses to 12 different representations of chemical phenomena. *Journal of Research in Science Teaching* 34(9): 949-968.

Kuh, G. D. (2001). The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties. Bloomington, IN: Indiana University Center for Postsecondary Research, 1-26.

Light, R. (2001). Making the Most of College: Students Speak Their Minds. Cambridge: Harvard.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In Proc. of 52nd Annual Meeting of the Assoc. for Computational Linguistics: System Demonstrations, pp. 55-60.

Schreiner, L. A., & Louis, M. C. (2011). The engaged learning index: Implications for faculty development. *Journal on Excellence in College Teaching*, 22(1), 5-28.

Seidman, A. (2005). College student retention: Formula for student success. Greenwood Publishing Group.

Shultz, G., & Gere, A.R. (2014) Meaningful learning about Lewis structures through reading and writing about historical science literature. Presentation at annual meeting of the American Chemical Society, San Francisco.

Shultz, G, & Gere, A.R. (2015) Reading and writing to learn the Lewis structure model. *Journal of Chemical Education*, 92(8), 1325-1329.