Writing Analytics, Data Mining, and Writing Studies

Norbert Elliot
New Jersey Institute of Technology
323 Dr. Martin Luther King Jr. Blvd
Newark, NJ 07102
1-856-952-7680
elliot@njit.edu

Katie Walkup
University of South Flordia
4202 E. Fowler Ave
Tampa, FL 33620
1-813-922-1492
kwalkup@mail.usf.edu

Joseph Moxley
University of South Florida
4202 E. Fowler Ave
Tampa, FL 33620
1-813-974-2421
mox@usf.edu

ABSTRACT
The primary goal of this workshop is to facilitate a research community around the topic of large-scale data analysis with a particular focus on writing studies, data mining, and analytics. The workshop aims hopes to generate cross-disciplinary research among writing program directors and faculty, computational linguists, and educational measurement specialists.

Keywords
Feedback; foundational measurement issues; *My Reviewers*; STEM, visual mapping

1. INTRODUCTION
The following passage from Victor Hugo's *Notre-Dame de Paris* is quoted in David Jay Bolter's *Writing Space: The Computer, Hypertext, and the History of Writing* [1].

Opening the window of his cell, he pointed to the immense church of Notre Dame, which, with its twin towers, stone walls, and monstrous cupola forming a black silhouette against the starry sky, resembled an enormous two-headed sphinx seated in the middle of the city.

The archdeacon pondered the giant edifice for a few moments in silence, then with a sigh he stretched his right hand toward the printed book that lay open on his table and his left hand toward Notre Dame and turned a sad eye from the book to the church.

"Alas!" he said, "This will destroy that."

In Bolter's seminal work *Writing Space: The Computer, Hypertext, and the History of Writing*, he begins with the above epigram about the book replacing the church [1]. Bolter uses this idea to parallel the replacement of the printed book with hypertext. As Bolter explains, "The idea and the ideal of the book will change: print will no longer define the organization and presentation of knowledge."

This workshop fully realizes Bolter's idea that "ceci tuera cela," or "this will destroy that." When applied to Writing Analytics (WA), researchers and practitioners stand at a pivotal point of change. Writing Analytics are going to redefine the teaching and learning space by replacing feedback as teachers and students have always delivered feedback. The affordances of digital tools mean that machines can process and present knowledge to an extent unimaginable by Bolter is a future that seems to have few limits.

Data-collection methods such as Latent Semantic Analysis (LSA) and Natural Language Processing (NLP) have enabled researchers in WA to present studies that portend a complex future for the discipline of Writing Studies—a discipline where humanities collaborative with mathematicians on predictive algorithms, corpus linguists on linguistic patterns discerned from big data, and computer sciences on intelligent tutoring systems. WA may eventually replace grading as we know it, but the research area is controversial, especially for researchers in the humanities.

To understand these concerns, it is important to recognize the history of providing machine feedback. While many humanities researchers reject the idea of WA and the use of corpus methods to refine feedback, Tackitt et al. point out that feedback and grading have always been controversial practices [2]. Many have investigated the reliability of instructor evaluation of writing within and without the disciplines [2]. Before these, however, Tackitt et al. write that "the evaluation of student learning through student writing is a modern model made possible through modern means and methods."

With this brief reminder that technologies replace technologies, the workshop leaders can look beyond the controversies of WA. This workshop will then seek to extend and surmount the current boundaries of WA [3] by attention to the following :

1. Structuring opportunities for students to learn
2. Understanding the cognitive, interpersonal, intrapersonal constructs, as they emerge within sociocognitive and sociocultural settings, that enable students to recognize and respond to feedback
3. Gaining actionable information about what practices will help students to become better writers in academic and workplace settings

When WA is reconfigured to embrace student learning, we can see that the efforts of researchers and practitioners change the learning space. With interdisciplinary collaboration, we can mediate the constructs that underlie WA as a field of research.

2. PRESENTATIONS

This workshop centers around mapping writing analytics from an interdisciplinary, student-centered perspective. As researchers point out, discussing WA from a disciplinary perspective can distract researchers and practitioners from completing actionable research.

The workshop begins with an activity in mapping writing analytics, led by Joseph Moxley (University of South Florida). The ensuing presentation extends foundational perspectives on the definition of Writing Analytics (WA) to further conceptualize the field. The authors use the metaphor of mapping to understand the tensions and successes navigated by researchers and practitioners and to chart new ways in which this field can benefit the domains of academia, business, and culture.

This interdisciplinary approach allows the audience to reconceptualize the field. From there, Alex Rudniy (Fairleigh Dickinson University) and Norbert Elliot (New Jersey Institute of Technology) explore the use of n-grams in analyzing student and instructor comments within My Reviewers[1], a web-based learning environment. Shown to be informative in a wide variety of applications, n-gram analysis is of interest in determining concept proliferation in topics, purposes, terminologies, and rubrics used in writing courses. As the present study demonstrates, unigram, bigram, trigram, fourgram, and fivegram analytic methods reveal important information about instructor and student use of concepts. This analysis holds the potential to lead to precise and actionable revision behaviors.

David Kaufer and Sugura Ishizaki (Carnegie Mellon University) introduce the concept of textual visualization to enhance learning in core writing courses. These authors use corpus methods to show that writing tasks require countless composing decisions that are typically beyond the conscious grasp of writers. Much of the skill of being "text aware" is to understand that texts produced from classroom assignments are not just composed of words and sentences but of highly structured and often highly predictive composing decisions. However, the decision-making underlying writing is an extremely abstract idea that is hard to make tangible for students. Although a significant number of pedagogical approaches has been investigated in the past three decades, the means to help students acquire more tangible understanding and control of their composing decisions has not been addressed. The authors propose to address this gap by developing a corpus-based learning tool to help students notice and reflect on composition decisions in their writing and to become more self-aware, reflective writers.

Valerie Ross, Mark Liberman, Lan Ngo, Rodger LeGrand (University of Pennsylvania) address another kind of reflective writing: peer feedback. The Critical Writing Program at Penn began working with My Reviewers in the Fall of 2013, working collaboratively with the My Reviewers team at the University of South Florida to develop a portfolio solution. Since then, students evaluate peer's portfolios. In turn, instructors use eportfolio tools to evaluate middle and end-of-semesters portfolios. As a result, Penn has developed a large corpus of peer reviews and epotfolio reviews. In this study, Ross et al. use a weighted log-odds-ratio, informative Dirichlet prior method ("bag of words" approach) to analyze student comments and scores posed to My Reviewers, which is designed to collect student writing as well as their peers' comments and scores on those drafts. This preliminary study suggests that the use of this methods shows lower-performing writers might be receiving kinds of feedback generally viewed as counterproductive in the field of writing studies.

From examining the effectiveness of feedback on revision, attitudes toward writing, student and instructor training and motivation, participants in this workshop will then begin to understand how big data researchers approach corpuses of student revisions. Chris Holcomb and Duncan Buell (University of South Carolina) approach First Year Composition as a big data phenomenon by prototyping software to study revision in a large corpus of student papers. The authors address a question central to scholarship in Composition and Rhetoric: What role does revision play in students' writing processes?

Denise Comer (Duke University) closes the day's presentations by recasting the framework for big data and WA research. Comer uses the frame of writing transfer to explore how researchers can transfer strategies, approaches, and knowledge about writing gained from big-data writing analtyics to other writing pedagogy contexts. The author will share methods and results from four big-data research projects, which stem from research gained in a writing-based Massive Open Online Course. Comer will present findings on big data and writing assessment, writing and peer-to-peer interactions, writing and negativity, and peer-review and transfer.

This workshop closes on a final collaborative activity, as the participants are asked once again to return to a mind-map of Writing Analytics. Using lessons learned from corpus methods and big data techniques, participants will reconceptualize the field from an interdisciplinary, actionable perspective.

[1] Dr. Joseph Moxley wishes to disclose a potential conflict of interest: while the My Reviewers software is not commercially available, it may become commercially available in the future. Because the data collection methods used in this study demonstrate the viability of My Reviewers, this research study may enhance the commercial value of My Reviewers. Ultimately, USF owns My Reviewers; however, Moxley possesses the rights to license My Reviewers. Given this potential conflict, Professor Moxley has filed the necessary USF conflict of interest paperwork. The Conflict of Interest Committee at USF has developed a management plan with which Dr. Moxley has complied prior to submitting this and similar research.

3. REFERENCES

[1] Bolter, D. J. (1991). *Writing space: The computer, hypertext, and the history of writing*. Lawrence Erlbaum: Hillsdale, NJ.

[2] Tackitt, A., Moxley, J., and Eubanks, D. (2015). Signifying scores: Instructor rating as an assessment measure. Manuscript submitted for publication.

[3] Buckingham Shum, S., Knight, S., McNamara, D., Allen, L., Bektik, D., and Crossley, S. 2016. Critical perspectives on writing analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (LAK '16). ACM, New York, NY,USA,481-483.
DOI=http://dx.doi.org/10.1145/2883851.2883854