

Writing Analytics: Conceptualization of a Multidisciplinary Field

Joe Moxley, *University of South Florida*

Norbert Elliot, *University of South Florida*

David Eubanks, *Furman University*

Meg Vezzu, *The Journal of Writing Analytics*

Sophie Elliot, *The Journal of Writing Analytics*

Will Allen, *Colorado State University*

Welcome to the inaugural issue of *The Journal of Writing Analytics*. *Analytics* operates in a long-standing multidisciplinary tradition of empirical educational research. Yet the subject of this journal, writing analytics, is new—and so a conceptual statement is warranted as we launch our publication.

Making sense of writing analytics is complex. The research target, written communication, has been in play since the second half of the fourth millennium (BCE) when Mesopotamians sliced signs onto wet clay. While its occurrence is relatively recent (tool use has been dated between 2.6–2.5 million years ago), writing is a significant technological advancement for humans. Perhaps, as Walter Ong (1982) has proposed, writing restructures consciousness, and perhaps as Jack Goody (1987) has argued, writing underpins civilization. Certainly, writing renders spoken language a material object. Whatever else writing may do, this sense of permanency prevents time from vanishing.

At its most basic level, writing is usefully defined as “a system of more or less permanent marks used to represent an utterance in such a way that it can be recovered more or less exactly without the intervention of the uttered” (Daniels, 1996, p. 3). While systems that capture ideas without reference to speech are termed semaisographic (mathematical formulas falls here), systems that meet the criteria of speech connection are classified as glottographic (Wood, 2010). Glottographic artifacts are our subjects. Taken together, those Mesopotamian script systems allow us to understand a civilization that would otherwise have vanished into the sands of present day Iraq. Taken together, the scripts you are about to read in *Analytics* tell us much about the cares and concerns of the scholars whose work appears in this issue. As readers, we are analysts forever involved in the reconstruction and construction of meaning.

In the five thousand years that separate present day writers and readers from the Mesopotamian literate culture, three human activities endure: the materialization of language in the written word; the development of technology that makes language visible; and the interpretation and use of information gained from analysis of that language. In our own time, twenty-first century digital affordances have made it possible to produce and analyze texts rapidly. As analysts, we can create corpora—bodies of texts—and perform linguistic analysis upon them in order to determine, at the level of word choice, the embodiment of ideas.

As this far too brief narrative illustrates, writing is a human project best understood through a variety of perspectives, from the historian who establishes contextualization to the linguist who interprets the function of language in these contexts. Mike Palmquist, the Founder of the Colorado State University Press and our publisher, notes in his introductory letter the range of professions contributing to this inaugural issue in the eleven studies to follow: computer science; creative writing; digital humanities; educational psychology; English language and literature/letters; health sciences; information science; language and literacy education; linguistics; mathematics; quantitative methodology; rhetoric and composition/writing studies; science education; second language learning; and writing program administration. As has always been the case, there is more to the analysis of writing than any one field can manage alone. We imagined a journal specifically dedicated to the language materialization, enabling technologies, and information interpretation beneficial to multidisciplinary researchers and those they hoped to influence. Because ours is a digital world in which information is produced and collected at stunning speed, a web-based journal with a turn-around time of one year from submission to print that reported state-of-the-art writing analytics research would allow form to follow function.

In this introduction to the first issue of *Analytics*, we therefore want to identify our present understanding of writing analytics, the process by which the journal functions, the taxonomy that threads together the scholarship in this issue, and ways that taxonomy has so far proven integrative.

1.0 Definition

While the *Oxford English Dictionary* holds no etymological entry for the term “writing analytics,” we propose that the origin is to be found in the 2016 publication of “Critical Perspectives on Writing Analytics.” In a workshop introduction for the Sixth International Conference on Learning Analytics and Knowledge (LAK), Simon Buckingham Shum, Simon Knight, Danielle McNamara, Laura Allen, Duygu Bektik, and Scott Crossley define the term

through its focus “on the measurement and analysis of written texts for the purpose of understanding writing processes and products, in their educational contexts, and improving the teaching and learning of writing” (p. 481). “The principal goal of writing analytics,” the authors write, “is to move beyond assessment of texts divorced from contexts, transitioning instead to a more nuanced investigation of how analytics may be effectively deployed in different writing contexts. Writing analytics thus aims to employ learning analytics to develop a deeper understanding of writing skills” (pp. 481-482). Metaphorically, the authors view writing as a window into the mind in which natural language processing (NLP)—techniques using computational techniques suitable to identify and extract linguistic features from stored electronic text or speech (Burstein, 2013)—can provide detailed information about students’ writing at multiple textual levels. Contextually, while such techniques are key, the algorithmic metrics driving them need to be “tuned by theories of how writing and learning shape each other, the scholarship of teaching writing, appropriate pedagogical practices and user interface design, and evidence from empirical evaluation of the total system” to be successful (p. 284). Socially, writing analytics promotes equity: While only a “privileged minority” has access to detailed feedback on a piece of writing, information from writing analytics can be available anywhere, anytime. Challenges are those common to all learning analytics applications—limits of conventional metrics, questions regarding socio-technical frameworks, impact of pedagogic and assessment contexts, and concern with ethical issues—and, as such, are best addressed as design problems.

One can imagine no better founding statement. The publication of *Analytics* seeks to provide an annual forum for research and multidisciplinary collaboration and to extend the definition of writing analytics offered by Shum and his colleagues. To achieve this extension, we sought an innovative process for our born-digital journal.

2.0 Process

To build credibility, we required the support of an international team of researchers who would serve on an annual basis as reviewers. To that end, colleagues agreed to serve on the Board of Reviewers, presented in Table 1 with the area of manuscripts each specialist reviewed.

Table 1

*Journal of Writing Analytics Board of Reviewers and Review Specialization,
Volume 1, 2017*

Reviewer	Affiliation	Review Specialization
Chris M. Anson	North Carolina State University	Rhetoric and Composition
Ian G. Anson	University of Maryland, Baltimore County	Electronic Forecasting
Laura Aull	Wake Forest University	Corpus Linguistics
Ryan Baker	University of Pennsylvania	Learning Analytics
Duncan Buell	University of South Carolina	Computer Science
Hugh Burns	Texas Woman's University and US Air Force Academy	Computational Rhetoric
Scott Crossley	Georgia State University	Applied Linguistics
Irvin R. Katz	Educational Testing Service	Cognitive Psychology
David Kaufer	Carnegie Mellon University	Digital Textual Analysis
Andrew Klobucar	New Jersey Institute of Technology	Digital Humanities
Suzanne Lane	Massachusetts Institute of Technology	Writing in the Disciplines
Djuddah A.J. Leijen	University of Tartu, Estonia	English Language Learning
Collin F. Lynch	North Carolina State University	Intelligent Tutoring Systems
Mya Poe	Northeastern University	Writing Assessment
Valerie Ross	University of Pennsylvania	Critical Writing
Alex Rudnay	Farleigh Dickinson University	Educational Data Mining
David Slomp	University of Lethbridge	Qualitative Research
Erica Snow	SRI International	Artificial Intelligence
Swapna Somasundaran	Educational Testing Service	Sentiment and Discourse Analysis
Jennifer Pei-Ling Tan	National Institute of Education, Singapore	Creativity & 21st Century Competencies

With the benefit of our Board (without whose brilliance the journal would not exist), the inaugural issue of the *Journal of Writing Analytics* was produced under a rigorous time line, with submissions undergoing rapid, detailed reviews at all stages of the process. Upon receiving a submission, the journal's Editor-in-Chief consults with the editors to ensure that the manuscript aligns with the aim of the journal. Once the Editor-in-Chief endorses the manuscript as appropriate, it is assigned to two members of the Board of Reviewers who are specialists in the submission's area of research. The reviewers then have four weeks to provide a publication recommendation and feedback on the manuscript, using a set of detailed reviewer guidelines. After both reviewers have completed their review and submitted their recommendation, the Editor-in-Chief makes a final decision as to whether or not to accept the manuscript for publication and notifies the author(s) of any requested revisions. Authors also receive the anonymous results of the feedback form submitted by the peer reviewers.

Following peer review, authors are given no more than eight weeks to make the requested revisions and re-submit their manuscript. The Editor-in-Chief reviews the revised manuscript to ensure that the reviewers' concerns have been addressed. Authors are then notified that their manuscript has been accepted for publication and are informed of the next steps in the publication process: copyediting and the creation of galley proofs. Manuscripts are edited for grammar, style, and clarity and returned to the authors within two weeks. The

authors then have 48 hours to review the edits and make any additional changes to the manuscript. Once all edits are complete, the production of galley proofs begins. This step consists of transferring the manuscript to the journal article template and formatting all tables and figures to produce the final PDF file. These proofs are created within two weeks, and authors are given 48 hours to review them for typographical and formatting issues. Once the authors review the proofs and all requested changes have been made, the manuscript is ready for publication in the upcoming journal issue. The first manuscript was submitted on December 10, 2016; the journal was published on September 7, 2017.

This rapid process is achieved through the implementation of editor desk rejects and the board commitment to publication. When a manuscript is received and evaluated as unsuitable for publication by the editors, the Editor-in-Chief writes to the author within 48 hours identifying the reasons the manuscript may not move forward in its present form. If the author(s) decide to revise and resubmit, the Editor-in-Chief then makes a second evaluation. When a manuscript is sent to the Board of Reviewers, every attempt is made to suggest revisions that will result in publication. If one reviewer accepts a manuscript and another rejects it, the editors work with the author to ensure that the reservations have been addressed before an adjudicating reviewer is identified.

To lend breadth to the submissions, the journal adopted five genres: research articles, research notes, new quantitative techniques, advances in code, and book reviews. Because the journal has been born digital, readers will note that *Analytics* publishes substantially longer articles. Research articles, for example, may be up to 30,000 words (approximately 100 double spaced manuscript pages, including references). We believe this length is often necessary to provide the technical and contextual detail needed to justify the study claims. Detailed author guidelines, presented as constructed responses, were posted on the journal homepage. Aligned with the author guidelines for each genre of submission, detailed reviewer guidelines were also published. In this way, *Analytics* adhered to best practices of structured submission and peer review in writing studies as demonstrated by *Research in the Teaching of English* and *IEEE Transactions in Professional Communication (IEEE PCS)*. Readers will also notice that the structured abstract format used to provide key information regarding research design and results has been adapted from *IEEE PCS*.

With support from the University of South Florida, Colorado State University Open Press publishes *Analytics* on an annual basis. The press uses Open Journal Systems, a journal management and publishing system that has been developed by the Public Knowledge Project through its federally funded efforts to expand and improve access to research. The graphics for *Analytics* were designed to reflect traditional academia and modern technology as they integrate

seamlessly. Inspired by the support of the University of South Florida, the color scheme and classic font mirrors the mission of the university to generate knowledge, foster intellectual development, and ensure student success in a global environment.

3.0 Taxonomy

Taking NLP techniques as the key technology for writing analytics, Shum and colleagues propose that the LAK community should play a substantial role in guiding educators and students on the evidence of impact of writing analytics in educational settings. For instance, there is over a decade's history of workshops supporting innovative use of NLP for educational applications, co-organized yearly by Jill Burstein and her colleagues.

We acknowledge the centrality of NLP to writing studies research. To lend additional perspectives, as the inaugural issue demonstrates, the editors, board, and authors aim to broaden the field through additional technologies and frameworks to advance student learning. Supported by the annual International Conferences in Writing Analytics—the fifth will occur on January 11 and 12, 2018—colleagues have advanced a new perspective associated with the four interrelated programs of research shown in Figure 1.

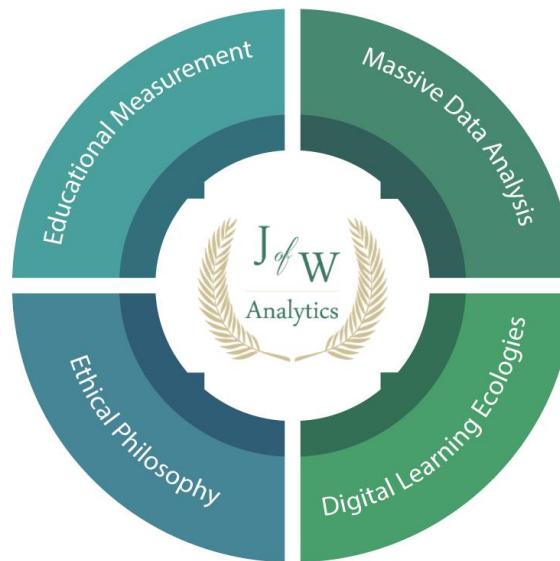


Figure 1. The Journal of Writing Analytics: Taxonomy.

Volumes can be written on each of the four research programs. In terms of *Analytics*, however, the topics of very recent studies suggest the taxonomy—the classification of principles—upon which the journal is founded:

- Advancements in educational measurement have allowed researchers to re-imagine ways we validate the interpretation and use of information (Kane, 2013) and the challenges to validity argumentation presented by psychology and technology (Mislevy, 2016).
- Innovation in massive data analysis, as the National Research Council has demonstrated in its 2013 report, has forced us to re-think our understanding of the collection of information in terms of three developments: a notable leap in the amount of data regarding human attitudes and activities that presently exists; the emergence of distributed computing systems that have become a reality with significant implications for the collection and processing of massive data; and the alarming fact that even if an individual time frame in which information is gathered does not involve a massive data set, information associated with that time frame can rapidly overwhelm storage and computing resources.
- Newly created platforms designed to be construct-specific digital ecologies such as *M-Write* at the University of Michigan and *MyReviewers* at the University of South Florida feature environments in which student writing can be examined in detail within naturalistic classroom settings.
- Focused attention in writing studies on ethical (moral) philosophy has recently launched two programs of research: reformulation of fairness (the identification of opportunity structures created through maximum construct representation under constrained conditions); and emphasis on justice (a form of fairness reliant on moral philosophy, resistance to Platonism, attention to context, and dedication to supporting the least advantaged (Kelly-Riley & Whithaus, 2016; Poe & Inoue, 2016).

As a set of guiding principles, the taxonomy had proven an integrative force for our first issue.

4.0 Integration

The range of research programs in this first issue is remarkable. These studies serve as a testament to the vibrancy of writing analytics as an academic discipline. In the 391 pages that follow, readers will find exciting study designs, innovative methodologies, important findings, evidence-based conclusions, and promising directions for future research.

There are seven research articles in Volume 1.

- In “Corpus Analysis of Argumentative Versus Explanatory Discourse in Writing Task Genres,” Laura Aull conducts a context-informed corpus analysis of lexical and grammatical keywords in part-of-speech tagged writing by first-year college students. Attention is given to genres of annotated bibliography, visual analysis, and argument. Based on key discourse patterns in tasks within the same course and in macro-genres across courses, her findings pose important questions regarding writing task design and the ability of students to adapt to different genres. (pp. 1-47)
- In “Applying Natural Language Processing Tools to a Student Academic Writing Corpus: How Large are Disciplinary Differences Across Science and Engineering Fields?” Scott A. Crossley, David R. Russell, Kristopher Kyle, and Ute Römer examine student-produced science and engineering texts from the Michigan Corpus of Upper-level Student Papers with regard to lexical sophistication and textual features of cohesion. Overall, the findings provide substantial evidence that macro-disciplinary and micro-disciplinary differences exist in student writing and that these differences are likely not related to student level. These findings have important implications for understanding disciplinary differences. (pp. 48-81)
- In “Statistical and Qualitative Analyses of Students’ Answers to a Constructed Response Test of Science Inquiry Knowledge,” Seohyun Kim, Minho Kwak, Lourdes Cardozo-Gaibisso, Cory Buxton, and Allan S. Cohen report on a comparative study of the language used by middle school students in their answers to a constructed response test of science inquiry knowledge. The researchers used two types of linguistic analysis: Latent Dirichlet allocation was used to extract topics from the texts of student responses; and systemic functional linguistic analysis was used to analyze the text of student responses on the same test. Results of the analyses demonstrate that students improved over time in their ability to use the discipline-specific and academic terminology of the language of scientific communication in the structured curriculum. (pp. 82-102)
- In “A Text Analytic Approach to Classifying Document Types,” Steven Walczak demonstrates the usefulness of a text mining and text analytics algorithm, STAR' (System for Text Analytics-based Ranking). The proof of concept study examines the power of the text mining platform to classify the research type of articles as theoretical or applied research. As Walczak finds, the STAR' method may be used by students and faculty to identify the likely source of research or discipline-specific information.

Additionally, STAR' classifications may be used by researchers to determine the most appropriate audience and publication for the reported information in their own manuscripts (pp. 103-146).

- In “Measuring the Written Language Disorder among Students with Attention Deficit Hyperactivity Disorder,” Diane Mitchnick, Clayton Clemens, Jim Kagereki, Vivekanandan Kumar, Kinshuk, and Shawn Fraser investigate Written Language Disorder (WLD) through a novel computational model. The model integrates the outcomes of common screening methods for WLD with common screening methods for Attention Deficit Hyperactivity Disorder (ADHD). By examining ADHD students against the WLD criteria, Mitchnick and her colleagues identify a relationship between WLD and ADHD. These results can be used to advance pedagogical techniques in education for ADHD and/or WLD learners (pp. 147-175).
- In “Discovering the Predictive Power of Five Baseline Writing Competences,” Vivekanandan Kumar, Shawn N. Fraser, and David Boulanger shed light on how to build a relatively simple Automated Essay Scoring (AES) system based on five baseline writing features. Focusing on spelling accuracy, grammatical accuracy, semantic similarity, connectivity, and lexical diversity, the authors show that the development process of a simple AES system compares well to state-of-the-art systems. In terms of design and evaluation of AES systems, the study provides important information in the call for transparency in the development of automated scoring systems (pp. 176-226)
- In “Assessing Writing Constructs: Toward an Expanded View of Inter-Reader Reliability,” Valerie Ross and Rodger LeGrand investigate construct representation and inter-reader agreement and reliability in ePortfolio assessment. This study contributes to the emergent field of empirical writing portfolio assessment that calls into question the prevailing, narrow standard of reliability evidence (built upon timed essay measurement rather than complex writing performances in naturalistic settings.) The study also contributes to recent research on multi-trait and discipline-based portfolio assessment used to establish validity evidence. In terms of the use of such information, Ross and LeGrand demonstrate how reliability and validity measures can significantly contribute to collaborative assessment, teacher training, and curricular improvement. (pp. 227-275).

There is one article focusing on a new quantitative technique in the issue:

- In “(Re)Visualizing Rater Agreement: Beyond Single-Parameter

Measures,” David Eubanks presents a new graphical technique for visualizing and assessing inter-rater agreement in discrete ordinal or categorical data, such as rubric ratings. Two applications are given—wine tasting and writing assessment—both involving social construction of meaning. As Eubanks reveals, such examples show that statistical reliability can emerge organically from the purpose of the project—a tradition involving factors ranging from social consequence to bureaucratic feasibility. The more we describe these purposes and their empirical consequences, Eubanks proposes, the more we can ensure a fascinating scholarly conversation holding benefits to assessment stakeholders and the students they serve. (pp. 276-310).

And there are three research notes:

- In “Transforming Text: Four Valences of a Digital Humanities Informed Writing Analytics.” Gregory J. Palermo examines existing and continuing potential crossover between digital humanities and writing studies. He draws from a history of meta-academic literature in digital humanities and writing studies to review shared methodological commitments, and he reviews current research in both fields involving computational techniques. While a review of past and current research in digital humanities and writing studies reveals shared attention to techniques, Palermo finds that future efforts to identify crossover between the two fields will benefit by identifying common values. He concludes with a set of heuristics beneficial to writing analytics in future research that draws upon, and contributes to, both digital humanities and writing studies. (pp. 311-343).
- In “Doing Big Data: Considering the Consequences of Writing Analytics,” Eric James Stephens focuses on consequences of big data as an emerging methodology used to make inferences in a wide variety of contexts, including educational settings. Providing a literature review of the development of new methods associated with massive data analysis, Stephens identifies critical questions researchers should consider in designing research and interpreting results. As we begin to adopt big data methodologies, he reminds us, we must continue to challenge and question their applications, implementations, and implications. (pp. 344-355).
- In “I Hear What You’re Saying: The Power of Screencasts in Peer-to-Peer Review,” Allison S. Walker reports on the screencast—an analytics tool that enables simultaneous recording of audio and video feedback on any digital document, image, or website—that may be used to enhance

feedback systems in educational settings. Using survey responses and screencasts exchanged among peer-to-peer interactions, the author provides in-depth analysis of students' experiences, perceptions, and attitudes toward giving and receiving screencast feedback. (pp. 356-391).

Based on eleven articles in this inaugural issue of *Analytics*, even the most conservative reviewer will have to agree that the field of writing analytics is an uncorked champagne bottle. It is probably true that each author, in one way or another, addresses issues involving educational measurement, big data, digitally based methodologies, and ethical use of information. It is certainly our hope that the 2018 issue will continue a tradition of multidisciplinary research in which building community is the first impulse as we go about our work.

And so, we arrive where we began: Making sense of this new area of study is complex—as complex as those whose work herein appears.

Author Biographies

Joe Moxley is professor of English at the University of South Florida. He presently serves as principal investigator on NSF Award 1544239, “Collaborative Research: The Role of Instructor and Peer Feedback in Improving the Cognitive, Interpersonal, and Intrapersonal Competencies of Student Writers in STEM Courses.” He co-authored, most recently, *Agency in the Age of Peer Production* (Studies in Writing & Rhetoric Series, 2012). Joe is Founding Editor of *The Journal of Writing Analytics*.

Norbert Elliot is research professor at the University of South Florida. He presently serves as principal investigator in NSF 1721749, “SBIR Phase I: Artificial Intelligence, Scientific Reasoning, and Formative Feedback: Structuring Success for STEM Students.” With Asao B. Inoue and Mya Poe, he is co-editor, most recently, of *Writing Assessment, Social Justice, and the Advancement of Opportunity* (forthcoming, The WAC Clearinghouse and University Press of Colorado). Norbert is Editor-in-Chief of the journal.

David Eubanks is Assistant Vice President for Assessment and Institutional Effectiveness at Furman University. His research focuses on assessing student learning, including writing assessment, predictive analysis, survey research, and the development of research methods and innovative software tools. These include rapid data-mining applications for structured data, inter-rater agreement visualization, and visualization of language usage of a corpus using vector mappings. He is co-author of “On Keeping Score: Instructors’ vs. Students’

Rubric Ratings of 46,689 Essays" (*WPA: Writing Program Administration*, (2016) 39(2), 53-80). David is Executive Editor of the journal.

Meg Vezzu holds a MS degree in Professional and Technical Communication from New Jersey Institute of Technology. Prior to joining the *Analytics* team, she was a Research Associate at Educational Testing Service. There, she worked on the development of interactive computer-based assessments, classroom materials, and score reports. Meg serves as the Managing Editor for the journal. For more about her work, visit <http://www.megvezzu.com>.

Sophie Elliot is a Philadelphia-based designer. While working as a Visual Merchandiser for Target Corporation, she currently runs her own graphic design and custom invitation business. Her designs range from whimsy and playful to sophisticated and modern. Sophie serves as Graphic Designer for the journal. Find her work at sophieelliot.carbonmade.com.

Will Allen holds a M.Ed. in Educational Foundations from the University of Oklahoma and currently serves as the Open Educational Resources Coordinator for Colorado State University Online. He has been a public educator for thirty years, serving both as a classroom teacher and as a building principal at the elementary and high school levels. He has also served as an instructor and a teaching assistant coordinator for the World Languages Department at the University of Oklahoma. Will serves as layout editor for the journal.

References

- Burstein, J. Automated essay evaluation and scoring. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-7). NJ: Wiley-Blackwell. DOI: 10.1002/9781405198431.wbeal1046
- Daniels, P. T. (1996). The study of writing systems. In P. T. Daniels and W. Bright (Eds.), *The world's writing systems* (pp. 3-17). Oxford, UK: Oxford University Press.
- Goody, J. (1987). *The interface between the written and the oral*. Cambridge, UK: University Press
- Kane, M. T. (2013). Validating the interpretation and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73. DOI: 10.1111/jedm.12000

- Kelly Riley, D., & Whithaus, C. (2016). A theory of ethics for writing assessment. [Special issue]. *Journal of Writing Assessment*, 9(1). Retrieved from <http://journalofwritingassessment.org/article.php?article=99>
- Mislevy, R. J. (2016). How developments in psychology and technology challenge validity argumentation. *Journal of Educational Measurement*, 53(3), 265-292. DOI: 10.1111/jedm.12117
- National Research Council (2013). *Frontiers in massive data analysis*. Washington, D.C.: The National Academies Press.
- Ong, W. J. (1982). *Orality and literacy: The technologizing of the word*. UK & NY: Methuen.
- Poe, M., & Inoue, A. B. (2016). Writing assessment as social justice [Special issue]. *College English*, 79(2).
- Shum, S. B., Knight, S., McNamara, D., Allen, L., Bektik, D., & Crossley, S. (2016). Critical perspectives on writing analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16)*. Association for Computing Machinery, vol. 25-29, pp. 481-483. DOI: <https://doi.org/10.1145/2883851.2883854>
- Woods, C. (2010). The earliest Mesopotamian writing. In C. Woods (Ed.), *Visible language: Inventions of writing in the ancient Middle East and beyond* (pp. 35-84). Chicago, IL: Oriental Institute of the University of Chicago.