

Critical Perspectives on Writing Analytics

Simon Buckingham Shum, Simon Knight

Connected Intelligence Centre
University of Technology Sydney
PO Box 123, Broadway, Ultimo
NSW 2007, AUS
first.lastname@uts.edu.au

Duygu Bektik

Knowledge Media Institute
The Open University
Walton Hall
Milton Keynes, MK7 6AA, UK
duygu.bektik@open.ac.uk

Danielle McNamara, Laura Allen

Science of Learning and Educational Technology Lab
Arizona State University
Tempe, AZ, USA

Danielle.McNamara@asu.edu
LauraKAllen@asu.edu

Scott Crossley

Department of Applied Linguistics/ESL
34 Peachtree St. Suite 1200
One Park Tower Building
Georgia State University
Atlanta, GA 30303, USA
scrossley@gsu.edu

ABSTRACT

Writing Analytics focuses 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. This workshop adopts a critical, holistic perspective in which the definition of “the system” and “success” is not restricted to IR metrics such as precision and recall, but recognizes the many wider issues that aid or obstruct analytics adoption in educational settings, such as theoretical and pedagogical grounding, usability, user experience, stakeholder design engagement, practitioner development, organizational infrastructure, policy and ethics.

Categories and Subject Descriptors

K.3.1 [Computers and Education]: Computer Uses in Education

General Terms

Design, Experimentation, Human Factors, Measurement

Keywords

Education, Writing, Natural Language Processing

1. WORKSHOP INTRODUCTION

The focus of this workshop is on the topic of writing analytics. Broadly defined, writing analytics involves the measurement and analysis of written texts for the purpose of understanding writing processes and products, in their educational contexts. Writing analytics are ultimately aimed at improving the educational contexts in which writing is most prominent. The principal goal of writing analytics is to move beyond assessment of texts divorced from contexts, transitioning instead to a more nuanced

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s). *LAK '16*, April 25-29, 2016, Edinburgh, United Kingdom. ACM 978-1-4503-4190-5/16/04.
<http://dx.doi.org/10.1145/2883851.2883854>

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. Thus, workshop discussions will focus on writing from a number of different perspectives. In particular, we will discuss analytics that can help to better understand both the writing *process* as well as the final *product*, as well as their interactions with task demands, such as essay genre and voice.

An additional focus of this workshop will be on the pedagogical context in which writing analytics should take place. Our aim is not simply to focus on the automated scoring of written essays. Rather, we aim to discuss writing analytics that can be meaningfully embedded within a pedagogical context. These discussions can relate to a number of issues, such as the delivery of feedback and adaptive instruction.

1.1 Writing as a window onto the mind

Effective writing is not only central to education and the workplace, but also a lifelong citizenship skill necessary for effectively engaging with society. A large majority of academic disciplines focus on the development of learners' skills in critical review, conceptual synthesis, reasoning, and disciplinary/professional reflection. In these subjects, writing arguably serves as the primary window into the mind of the learner. Huge effort is invested in literacy from the earliest schooling, extending into higher education. Yet educators and employers alike recognize the challenge of cultivating this ability in graduates, with poor written communication skills a common cause of complaint.

Extending beyond scholarly academic writing, many educators also have a keen interest in disciplined, autobiographical reflective writing as a way for students to review and consolidate their learning, thus providing a means for assessing the deepest kinds of shifts that can occur in learner agency and epistemology. Such approaches are also common in the training and development of professional reflective practitioners.

Writing is, however, time consuming, labor-intensive to assess, difficult for students to learn, and not something that all educators can coach well, or even consider their job to coach. It is in

addressing these systemic limitations that Writing Analytics is attracting significant educational interest and commercial investment.

1.2 Harnessing NLP

Natural language processing (NLP) techniques have been proposed as one of the most effective methods for analyzing writing. In particular, NLP can provide detailed information about the properties of students' writing at multiple levels of the text. For instance, NLP tools have been developed to provide information about the words in the text, such as their imageability and frequency in the English language. Additionally, NLP tools have been developed to calculate various other aspects of text, such as the difficulty of the sentences, and the presence of cohesion at multiple different levels of the text (e.g., between sentences, paragraphs, etc.).

As NLP moves beyond research labs and into mainstream products, the Learning Analytics community has the opportunity and challenge of harnessing language technologies, and delivering them in effective ways that enhance learning.

NLP approaches are, of course, the key enabling capability of these technologies, but they are just one piece of the puzzle for an effective learning analytics solution: these approaches 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, not just algorithmic metrics.

The LAK community should be in a position to guide educators and students on the evidence of impact in this new space. *What questions should be asked before buying a new product or trialling a new research prototype? What are the options for evaluating such tools? What staff competencies are required to ensure that such tools have the maximum chances of success? Do students need orientation or training? What pedagogical contexts can this tool be applied to?* These are the often-ignored constraints around a potentially disruptive technology.

1.3 Promises and pitfalls

Ultimately, for the tools to be successful, educators and students must trust them, and the effort of learning these new tools must pay back. Computational assessment of writing elicits strong reactions from many educators. For skeptics, handing the task of assigning feedback or grading essays to a machine crosses a boundary line marking the limits of artificial intelligence (AI). An important research question is whether or not such skepticism is justified.

Writing Analytics have in common similar potential and pitfalls to other learning analytics applications. At the most optimistic level, the promise of writing analytics is the kind of 24/7 personalized feedback that is currently only available to a privileged minority via detailed, timely feedback from educators as they draft texts. However, this workshop will take a systemic perspective, problematizing the contexts that Writing Analytics are deployed within, and partially constituted by. Evaluation of Writing Analytics is thus framed as a design problem, raising questions about conventional metrics (such as precision, recall), alongside: socio-technical concerns; pedagogic and assessment contexts; and ethical issues.

1.4 Critical, systemic perspectives

Learning Analytics as a field sits at the confluence of existing research tributaries. The LAK *Discourse-Centric Learning Analytics (DCLA)* workshops forged connections with CSCL

discourse researchers to ensure that DCLA built on existing work. DCLA workshops have had a couple of papers on extended student writing, which we now argue merits its own workshop.

This workshop thus seeks to build similar bridges to existing research communities. There is a decade's tradition of *Workshops on Innovative Use of NLP for Building Educational Applications*, operating within the computational linguistics research paradigm, with evaluation based on information retrieval (IR) metrics, but applied specifically to educational texts. The *Computer Assisted Assessment* community has a tradition of research into student writing, and has a strong educational researcher presence. Research in *Computer-Supported Collaborative Learning* has a primary focus on student discourse.

The workshop aims to reflect the distinctive contribution that SoLAR and LAK bring, namely, a holistic perspective in which the definition of "the system" and "success" is not restricted to IR metrics such as precision and recall, but recognizes the many wider issues that aid or obstruct analytics adoption in educational settings, such as theoretical and pedagogical grounding, usability, user experience, stakeholder design engagement, practitioner development, organizational infrastructure, policy and ethics.

1.5 Submissions & workshop format

Thus, in this first Writing Analytics workshop, we aim to bring together active researchers and reflective practitioners in the field from both academia and industry, spanning K-12, higher education, and the workplace. This forum will provide the chance for newcomers to see the state of the art in a range of approaches, as well as appreciate the issues that arise around writing.

The workshop problematizes writing analytics providing space for critical reflection from a range of communities on the development and application of such techniques. We welcome contributions from technical and educational perspectives, inviting theoretical and empirical evidence, alongside critical perspectives. Issues include:

- Pedagogically-grounded requirements for language technologies to support a specific genre of writing (even if these are extraordinarily challenging)
- Design and validation of analytics for different genres of academic writing (e.g., literature review; debate analysis; personal reflection)
- The relationship between assessment regimes and choice of writing analytics (e.g., summative grading for high stakes tests; formative feedback on open-ended reflection; individual versus collaborative peer review)
- Writing analytics in support of wider pedagogic context (e.g., using writing to recommend readings; writing analytics to support peer assessment)
- Arguments for the potential benefits (or damage) of engaging with writing analytics (e.g., Might rapid feedback disrupt critical reflection processes? Is automated feedback perceived differently by students than human feedback?)
- Compelling (even fun?) user interfaces for engaging with automated writing feedback (e.g., annotations; visualizations of content and structure, games)
- Empirical evaluations of research prototypes and commercial products
- Principles for embedding software tools into practice (e.g., student and staff orientation; common misconceptions)
- Organizational adoption case studies
- Ethical issues specific to writing analytics (e.g., given the range of ideas and emotions that can be expressed)

Following an innovative format we have used in past workshops, brief position statements and presentations will be invited to address the above themes, with participants in each thematic session assigned different roles to provoke multidisciplinary dialogue, such as *analytics researcher, tool developer, data provider, writing researcher, writing educator, and commentator*.

2. AUDIENCE AND PARTICIPANTS

This workshop will be of interest to a wide range of LAK delegates including: students and researchers actively engaged in writing research, text analytics or writing analytics specifically; educators in schools, universities and businesses; leaders and policymakers; and companies active or potentially active in the field. As the first workshop devoted to this topic at LAK, it will also serve as a community-building event. Participants will be expected to leave with a clearer understanding of, and critical perspectives on, the range of purposes for which Writing Analytics may be deployed, the current state of the art, criteria and methods for evaluation, and organizational adoption issues.

3. WORKSHOP CHAIRS

Simon Buckingham Shum is Professor of Learning Informatics at the University of Technology Sydney, where he directs the Connected Intelligence Centre. His research focuses on learning analytics for higher order competencies such as academic writing, argumentation and collaboration. He served as LAK12 Program Co-Chair, and co-chaired the LAK13/14 workshops on Discourse-Centred Learning Analytics.

Simon Knight is a Research Fellow in Writing Analytics at the Connected Intelligence Centre, University of Technology Sydney. His research focuses on the relationship of analytics to epistemology, pedagogy and assessment, discourse analytics, and epistemic cognition, particularly around information seeking, work which has been presented at LAK and ICLS. He co-chaired the ICLS14 Workshop on Learning Analytics for Learning and Becoming in Practice and LAK15 Workshop on Temporal Analyses of Learning Data.

Danielle McNamara is a Professor of Psychology at Arizona State University, where she directs the Science of Learning and Educational Technology Lab. Her research focuses on discovering new methods to improve students' ability to understand text, learn new information, and convey their thoughts in writing. Her work integrates various approaches and methodologies including the development of intelligent tutoring systems and the development of natural language processing tools.

Laura Allen is a Doctoral Student in the Psychology Department at Arizona State University. The overarching aim of her research is to better understand the cognitive processes involved in language comprehension, writing, knowledge acquisition, and conceptual change, and to apply that understanding to educational practice by developing and testing educational technologies. Her research has been presented at LAK15 and other conferences related to writing analytics.

Duygu Bektik is a Doctoral Student at the Knowledge Media Institute, Open University UK. Her research investigates whether computational techniques can automatically identify the attributes of good academic writing in undergraduate student essays within different disciplines; and if this proves possible, how best to feedback actionable analytics to support educators in their essay assessment processes, which has been presented at LAK14/15.

Scott Crossley is Associate Professor of Applied Linguistics at Georgia State University. His primary research focus is on natural

language processing and the application of computational tools and machine learning algorithms in language learning, writing, and text comprehensibility. His main interest area is the development and use of natural language processing tools in assessing writing quality and text difficulty. Professor Crossley works as a senior researcher on Writing Pal, an intelligent tutoring system under development at Arizona State University.

4. REFERENCES

The following represent the chairs' research in this topic.

- [1] Allen, L., Jacovina, M. and McNamara, D. in press. Computer-based writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research*. The Guilford Press: New York, NY.
- [2] Allen, L., Snow, E. and McNamara, D. S. 2015. Are you reading my mind? Modeling students' reading comprehension skills with Natural Language Processing techniques. In: *5th International Learning Analytics & Knowledge Conference* (Poughkeepsie, NY, USA, March 16 - 20, 2015). LAK15. ACM, New York, NY, 246-254. DOI=<http://dx.doi.org/10.1145/2723576.2723617>
- [3] Buckingham Shum, S., Á. Sándor, R. Goldsmith, X. Wang, R. Bass and M. McWilliams (2016). Reflecting on reflective writing analytics: assessment challenges and iterative evaluation of a prototype tool. *6th International Learning Analytics & Knowledge Conference*, (Edinburgh, UK, April 25 - 29 2016). LAK16. ACM, New York, NY. DOI=<http://dx.doi.org/10.1145/2883851.2883955>
- [4] Knight, S. and K. Littleton (2015). Discourse-centric learning analytics: mapping the terrain. *Journal of Learning Analytics*, 2 1, 185-209.
- [5] Knight, S., S. Buckingham Shum and K. Littleton (2014). Epistemology, assessment, pedagogy: where learning meets analytics in the middle space. *Journal of Learning Analytics*, 1, 2, 23-47.
- [6] McNamara, D., Crossley, S., Roscoe, R., Allen, L. and Dai, J. 2015. A hierarchical classification approach to automated essay scoring. *Assessing Writing*, 23 (Jan. 2015), 35-59. DOI=<http://dx.doi.org/10.1016/j.asw.2014.09.002>
- [7] McNamara, D., Graesser, A., McCarthy, P. and Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Matrix*. Cambridge: Cambridge University Press.
- [8] Roscoe, R.D., Varner, L.K., Crossley, S.A. and McNamara, D.S. (2013). Developing pedagogically-guided algorithms for intelligent writing feedback. *International Journal of Learning Technology*, 8, 4, 362-381. DOI=<http://dx.doi.org/10.1504/IJLT.2013.059131>
- [9] Roscoe, R., Varner, L., Weston, J., Crossley, S. and McNamara, D. (2014). The Writing Pal Intelligent Tutoring System: Usability Testing and Development. *Computers and Composition*, 34 (Dec. 2014), 39-59. DOI=<http://dx.doi.org/10.1016/j.compcom.2014.09.002>
- [10] Simsek, D., Buckingham Shum, S, Sándor, Á, De Liddo, A., and Ferguson, R. (2013) XIP Dashboard: Visual Analytics from Automated Rhetorical Parsing of Scientific Metadiscourse. *1st Int. Workshop on Discourse-Centric Learning Analytics*, LAK13. Leuven (Apr. 8-12, 2013).
- [11] Simsek, D., Sandor, Á., Buckingham Shum, S., Ferguson, R., De Liddo, A. and Whitelock, D. (2015). Correlations between automated rhetorical analysis and tutors' grades on student essays. In: *5th International Learning Analytics & Knowledge Conference* (Poughkeepsie, NY, USA, March 16 - 20, 2015). LAK15. ACM, New York, NY, 355-359. DOI=<http://dx.doi.org/10.1145/2723576.2723603>