JOURNAL OF LEARNING ANALYTICS

Volume 1 - Issue 2

SOCIETY for LEARNING ANALYTICS RESEARCH

(2014). Scientometrics as an Important Tool for the Growth of the Field of Learning Analytics. *Journal of Learning Analytics, 1*(2), 1-4.

Scientometrics as an Important Tool for the Growth of the Field of Learning Analytics

Negin Mirriahi, Dragan Gasevic,
Phil Long and Shane Dawson
Editors, Journal of Learning Analytics

ABSTRACT: This article introduces the special issue from SoLAR's Learning Analytics and Knowledge conference. Learning analytics is an emerging field incorporating theory and practice from numerous disciplines to investigate how learner interactions with digital artefacts can provide actionable data to the learner herself, and educators about the learning process. As the field continues to expand there is a timely opportunity to evaluate its ongoing maturation. This evaluation could be in part informed by regular scientometric analyses from both the Journal and Conference publications. These analyses can collectively provide insight into the development of learning analytics more broadly and assist with the allocation of resources to under-represented areas for example.

KEYWORDS: Special issue, learning analytics, research, practice, Society for Learning Analytics Research, SoLAR, LAK'13

1 EDITORIAL

We would like to dedicate this issue to our dear colleague and friend Erik Duval from Katholieke Universiteit Leuven, Belgium. Erik has been key figure in the development of the Society for Learning Analytics Research as both executive member and prominent researcher advancing the field. At LAK'14 Indianapolis, USA, Erik informed us that he was unable to attend the conference as he had been diagnosed with non-Hodgkin lymphoma[1]. Despite undergoing a difficult medical treatment, Erik has maintained his unfailing level of energy, sense of humour and passion – traits which capture the hearts of those he interacts with[2]. Erik continues to contribute to the field of learning analytics undertaking an active role in the organization of the 2nd Learning Analytics Summer Institute recently held at Harvard Graduate School of Education in Cambridge, MA. This issue is another endeavour which Erik has played a leading role as one of the guest editors of the special issue that was compiled based on the selected papers presented at LAK'13; the conference that Erik had organized and chaired in Leuven, Belgium. We are pleased to learn the news about the positive effects of his treatment and improvements in his health. We wish him a speedy recovery and send our collective love and support toward Leuven and his family.

1.1 Scientometrics for the Field of Learning Analytics

This special issue led by Ochoa, Suthers, Verbert and Duval includes an invited set of extended papers from the LAK 2013 conference held in Leuven, Belgium (Suthers, Verbert, Duval, & Ochoa, 2013). The

[1] http://erikduval.wordpress.com/2014/03/31/not-so-great-news/
[2] http://erikduval.wordpress.com/2014/07/20/good-week/

first paper, by Ochoa et al. (2014) provides a scientometric analysis of LAK2013 (Price, 1978; Van Raan, 1997), as part of their guest editorial. In so doing, the paper highlights the authorship and citation trends along with content analysis of the conference papers to illustrate the diversity of topics and research domains represented. For example, content and text analysis conducted by Ochoa et al. (2014) as part of the scientometric analysis of LAK2013 papers led to the emergence of six clusters or themes: visualizations, behaviour analysis, social learning analytics, applications, challenges, and reflections. This editorial picks up this thread to emphasise the need for continued systematic authorship, citation, and content analysis of publications to help guide and inform the development of the field of learning analytics – its research, theory, and practice.

Learning analytics research activity has rapidly progressed over the past five years. As the latest call for papers for the 5th International conference on Learning Analytics and Knowledge (LAK'15) has been recently released[3], there has been rapid growth in both conference attendance and submitted papers and workshops. From past conferences and special journals, we have seen diverse representation of academic disciplines and hence associated methodologies and practices, assumptions and theory (Dawson, Gašević, Siemens, & Joksimovic, 2014). The commingling of various academic disciplines can provide both friction and benefits. For instance, the convergence of disciplines can result in an unanticipated paradogma or push for hegemonic status (Kirschner, 2014). This perceived concern however is outweighed by the benefits derived from interdisciplinary collaborations. The concerted effort of researchers from a broad range of disciplines, such as but not limited to computer science, education, linguistics, neuroscience, and psychology, can provide a powerful impetus for moving towards understanding complex learning processes through the amalgamation of analytical methods previously confined to particular disciplines. However, to date, the majority of published learning analytics research, whether in journals or conference proceedings has largely been conducted by non-multidisciplinary research teams (Dawson et al., 2014). Hence, regular scientometric analysis of the state of learning analytics research can uncover whether there is a shift towards more interdisciplinary teams or if new strategies are required to facilitate greater collaboration. Further, with the progressive rise in scholarly contributions to the field, there is a need to extend conclusions drawn from various studies to practical applications to enhance learning and teaching and investigate scalability and impact of applied learning analytics.

Future publications in the Journal for Learning Analytics and the LAK conferences will continue to be fruitful resources to benchmark and evaluate the progression of this research drawing on analyses such as those presented in Ochoa et al. As the research advances we will look to integrate the associated data sets where possible as an added venue for validation and replication studies and to assist the development of the field. As an emerging analytics field we must use the tools of our trade as a means for reflection and insightful retrospection in order to advance the discipline.

1.2 This Issue

Many thanks for the outstanding work of Xavier Ochoa, Dan Suthers, Katrien Verbert and Erik Duval for developing this special issue. The papers selected by our guest editors for this special issue are

[3] http://lak15.solaresearch.org/call-for-papers

representative of the various topics discussed at LAK 2013 as revealed in their scientometric analysis presented in the first paper of this issue. Continuing on from the opening paper, in the second paper, Knight, Shum, & Littleton (2014) examine the relationships between pedagogy, assessment, and epistemology; challenging us to identify where we stand in our own learning analytic practice. In a more empirical approach, the next two papers analyze the knowledge construction process. Wise, Zhao, & Hausknecht (2014) draw our attention to the importance of “invisible behavior” such as student message reading as well as writing of messages in online forums. Halatchliyski, Hecking, Goehnert, & Hoppe (2014) apply main path analysis to the evolution of artifacts in a collaboratively edited Wikiversity wiki to analyze the introduction and evolution of ideas in the community. This special issue continues with two papers that focus on the evaluation of content quality. Monroy, Rangel, & Whitaker (2014) study how learning analytics could help to assess the effectiveness and impact of digital curricula. Gunnarson and Alterman (2014), on the other hand, analyze the quality of content produced by students based on peer-promotion information. Finally, the analysis of a less digital and more hands-on type of learning is presented when Worsley and Blikstein (2014) investigate how design skills manifest themselves during the construction of physical artifacts. All these papers, diverse in topic, approaches and techniques, provide a sample of the "middle-space" that is Learning Analytics and demonstrate the growth and multifariousness of this research domain and its potential to inform learning and teaching practice.

REFERENCES

Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014). Current State and Future Trends: A Citation Network Analysis of the Learning Analytics Field. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 231–240). New York, NY, USA: ACM. doi:10.1145/2567574.2567585

Gunnarsson, B. L., & Alterman, R. (2014). Peer promotions as a method to identify quality content. *Journal of Learning Analytics*, *1*(2), 126–150.

Halatchliyski, I., Hecking, T., Goehnert, T., & Hoppe, H. U. (2014). Analyzing the Path of Ideas and Activity of Contributors in an Open Learning Community. *Journal of Learning Analytics*, *1*(2), 72–93.

Kirschner, P. a. (2014). When a paradigm becomes a paradogma. *Journal of Computer Assisted Learning*, *30*(4), 297–299. doi:10.1111/jcal.12074

Knight, S., Shum, S. B., & Littleton, K. (2014). Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space. *Journal of Learning Analytics*, *1*(2), 23–47.

Monroy, C., Rangel, V. S., & Whitaker, R. (2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K-12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94–125.

Ochoa, X., Suthers, D., Verbert, K., & Duval, E. (2014). Analysis and reflections on the third learning analytics and knowledge conference. *Journal of Learning Analytics*, *1*(2), 4–21.

Price, D. de S. (1978). Editorial statements. *Scientometrics*, *1*(1), 3–8. doi:10.1007/BF02016836

Suthers, D., Verbert, K., Duval, E., & Ochoa, X. (Eds.). (2013). *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. New York, NY, USA: ACM.

Van Raan, A. F. (1997). Scientometrics: State-of-the-art. *Scientometrics*, *38*(1), 205–218.

Wise, A., Zhao, Y., & Hausknecht, S. (2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics*, *1*(2), 48–71.

(2014). Scientometrics as an Important Tool for the Growth of the Field of Learning Analytics. *Journal of Learning Analytics*, *1*(2), 1–4.

Worsley, M., & Blikstein, P. (2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151–186.

(2014). Analysis and Reflections on the Third Learning Analytics and Knowledge Conference (LAK 2013). *Journal of Learning Analytics*, *1*(2), 5-22.

Analysis and Reflections on the Third Learning Analytics and Knowledge Conference (LAK 2013)

Xavier Ochoa
Escuela Superior Politécnica del Litoral, Ecuador
xavier@cti.espol.edu.ec

Dan Suthers
University of Hawai'i at Manoa, USA

Katrien Verbert
Vrije Universiteit Brussel and University of Leuven, Belgium

Erik Duval
University of Leuven, Belgium

Abstract: Analyzing a conference, especially one as young and focused as LAK, provides the opportunity to observe the structure and contributions of the scientific community around it. This work will perform a Scientometric analysis, coupled with a more in-depth manual content analysis, to extract this insight from the proceedings and program of LAK 2013. Authorship analysis reveals an open and international community, while internal citation analysis provides evidence of the construction of a body of knowledge central to Learning Analytics. The analysis of the content of the papers found five main topics: visualization, behaviour analysis, social learning analytics, learning analytics for MOOCs, and learning analytics issues (ethical, scalability, etc.), as well as papers reflecting on the field itself. We discuss representative papers presented at the conference, highlighting trends and new developments. Learning analytics is a diverse multidisciplinary field with an emerging interdisciplinary core, well situated to benefit from productive dialogue concerning its scope and purpose and to reflect on the pedagogies and epistemologies implied by its methods.

Keywords: Scientometrics, conference review, learning analytics

1 INTRODUCTION

Learning Analytics is a new, expanding field that grows at the confluence of learning technologies, educational research, and data science. From early works, such as Tinto's (1993) study on factors affecting the persistence of college students, to field-defining expositions, such as Siemens and Long (2011) and Campbell et al.'s (2007) work on Academic Analytics, to the current papers in this special issue, researchers and practitioners with different backgrounds and methodologies have tried to solve two simple but challenging questions: How do we measure the important characteristics of the learning process? And how do we use those measurements to improve it?

To assemble this disparate group of researchers and practitioners, the first Learning Analytics and Knowledge (LAK) conference was organized in 2011 in Banff, Alberta, Canada. This first event, with only 27 presentations and 34 different authors, served as the nucleus for what is now the Learning Analytics community. The success of this event led to the organization of LAK 2012 in Vancouver, British Columbia, Canada, this time with 52 presentations and 140 authors. LAK 2012 recognized the existence of the community and set the basis for an annual conference in the field. The articles presented in this special issue belong to the third iteration of this community meeting, LAK 2013, organized in Leuven, Belgium. The focus of this third year was Learning Analytics as a "middle-space" where different disciplines could exchange visions, ideas, and methodologies to improve learning through a deep understanding of the processes in which students, instructors, and institutions are involved.

As a transition point from a fledgling to a consolidated community, LAK 2013 provides a snapshot of the history of Learning Analytics. The goal of this issue of the journal is to provide an overview of the composition and content of the LAK 2013 program. This overview will be conducted in two ways: 1) a Scientometric analysis of the articles published in LAK '13, and 2) a review of the content of representative conference papers. Our conclusions will highlight the main findings of these two approaches.

2 SCIENTOMETRIC ANALYSIS

All published scientific endeavours leave traces susceptible to quantitative analysis in order to gain a better understanding of epistemological phenomena occurring in the scientific field. Co-authorship and bibliographical coupling networks could provide insight into the nature of the scientific community and how the individuals in it influence each other (Kessler, 1963). Text processing of the content of the scientific papers could provide evidence for the emergence or disappearance of research topics and concepts from the research field (Ding, Chowdhury, & Foo, 2001). Simple statistics about the authors and their papers can provide an indication of the reach, diversity, and "health" of the community of scientists and practitioners interested in the field. This type of analysis is commonly known as Bibliometrics (when related only to the documents produced by research) or more generally as Scientometrics (when the focus is the research community and the knowledge building process) (Hood & Wilson, 2001). In this section, we study the LAK 2013 program through the Scientometric lens. The rest of this section will provide a quantitative analysis of the status of Learning Analytics research as presented at LAK 2013.

2.1 Authorship Analysis

The first step in understanding the authorship of LAK 2013 is to obtain simple statistics regarding the author distribution of the papers presented at LAK 2013:

Basic Quantities: LAK 2013 had an approval rate of 28% for full papers (58 submitted, 16 accepted) and an approval rate of 22% for short papers (36 submitted, 8 accepted) without counting 14 full papers accepted as short. By comparison, the acceptance rate for full papers for LAK 2011 and LAK 2012 was 45% and 39% respectively. The LAK 2013 proceedings contained 47 documents (16 full papers, 22 short papers, 2 design briefs, 4 workshops, 2 panels, and 1 note from the editors). A total of 143 authors produced those documents.

New and Returning Authors: *At LAK 2011, there were 34 authors. Of those, 30 returned to LAK 2012, together with 110 new authors, expanding the community to 140 individuals. At LAK 2013, there were 41 returning authors from LAK 2011 (4), 2012 (17), or both (20). At LAK 2013, the community annexed 102* *new authors.*

Authorship distribution: How prolific an author is can be measured by how many papers he or she has been involved in. In the case of LAK '13, the most prolific authors are Simon Buckingham Shum (5 documents), Peter Reimann, Rebecca Ferguson, and Ryan Baker (3 documents each). If their contributions are weighted (the weight of a paper is the inverse of the number of authors), the list changes to Paulo Blikstein (1.5 papers), Simon Buckingham Shum (1.075 papers), Andreas Harrer, Doug Clow, and June Ahn (1 paper each). The distribution is very flat, with no author actually dominating the scene. This evidence, together with the fact that most of the authors were new to LAK, suggests that Learning Analytics is still an open community where interested researchers can be included without being part of a closed "academic club."

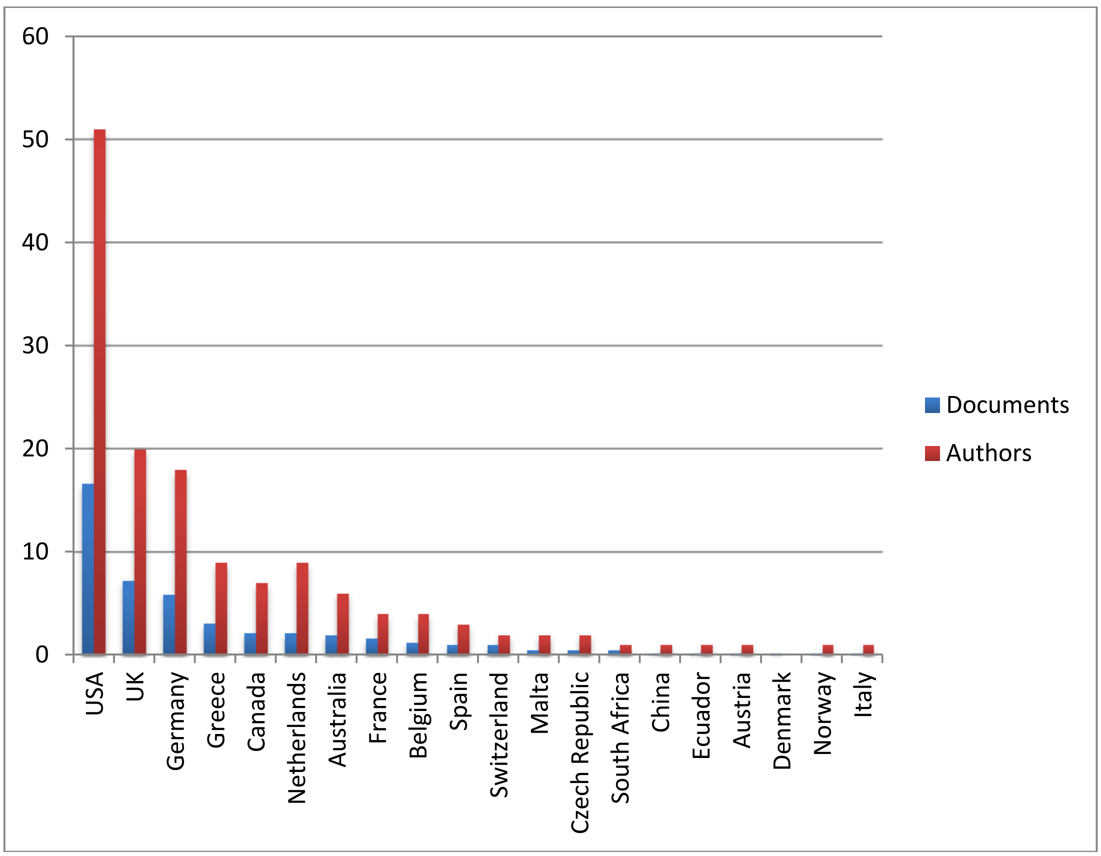

Figure 1. Country Distribution by Authors and Documents

Geographical Distribution: To calculate the geographical distribution, two measurements were considered: contribution by author and contribution by document. The first was calculated simply by the number of different authors affiliated with each country. The second was calculated by the weighted contribution of each author (each paper weight as the inverse of the number of authors). The top countries for the first measurement were USA (51 authors), UK (20 authors), Germany (18 authors), and Greece and Netherlands (9 authors each). For the second metric, the top countries do not vary much: USA (16.65 papers), UK (7.275 papers), Germany (5.9 papers), Greece (3.05 papers) and Canada (2.2 papers). A continental view shows the following results: Europe contributed with 76 authors, North America with 58, Oceania with 6 and Asia, South America, and Africa with 1 author each. A full graph of the distribution (Figure 1) presents a conference still dominated by the main research players worldwide: USA, West and Central Europe, Canada, and Australia. While small contributions are present from Latin America and Africa, there is a notable absence of Asian contributors.

Career Length Distribution: The length of the research career of each author was estimated using the date of first publication as recorded by the ACM Digital Library. While this data under-represents the career length of researchers in the field of education, the values obtained offer a general indication of distribution of seniority. For the 143 authors, the distribution was as follows: 28 senior researchers (>=15 years), 45 junior researchers (<=14 years and >4 years) and 70 Ph.D. students or initial stage researchers (<=4 years). This distribution reflects the pyramidal structure of most scientific groups (see Figure 2).

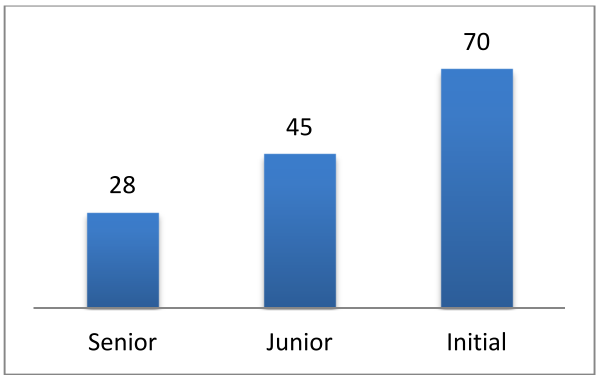

Figure 2. Distribution of Career Length

Sector Distribution: Authors were classified in three sectors (Academia, Research, and Industry) according to their affiliation. The results show that 128 authors belong to academic institutions, 8 to research centres, and 7 to companies. LAK ’13 can be clearly classified as a research-oriented conference.

Additional information about the Learning Analytics community can be extracted from how different individuals collaborate in order to produce research. The following is a series of collaboration indicators:

Authors per Paper: The average is 3.44 authors, which means that LAK 2013 publications usually needed a team of at least three people. This result maintains when the distribution of authors per paper is analyzed (Figure 3). Teams of two, three, four, and five individuals produced most of the papers rather than single authors.

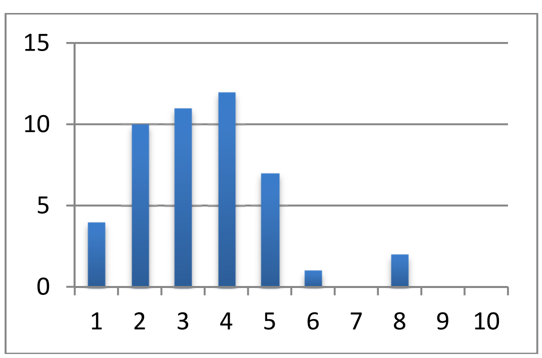

Figure 3. Distribution of Authors per Paper

Institutional Collaboration: We consider a paper inter-institutional if its authors are affiliated with more than one institution from the same country. We consider a paper international if its authors are affiliated with at least two institutions in different countries. Analysis of the collaboration shows that, of the 47 documents, 14 (30%) were international, 8 (17%) were inter-institutional, and 25 (53%) were produced by authors of a single institution. These values reflect the current percentage of collaboration in science (Leydesdorff & Wagner, 2008).

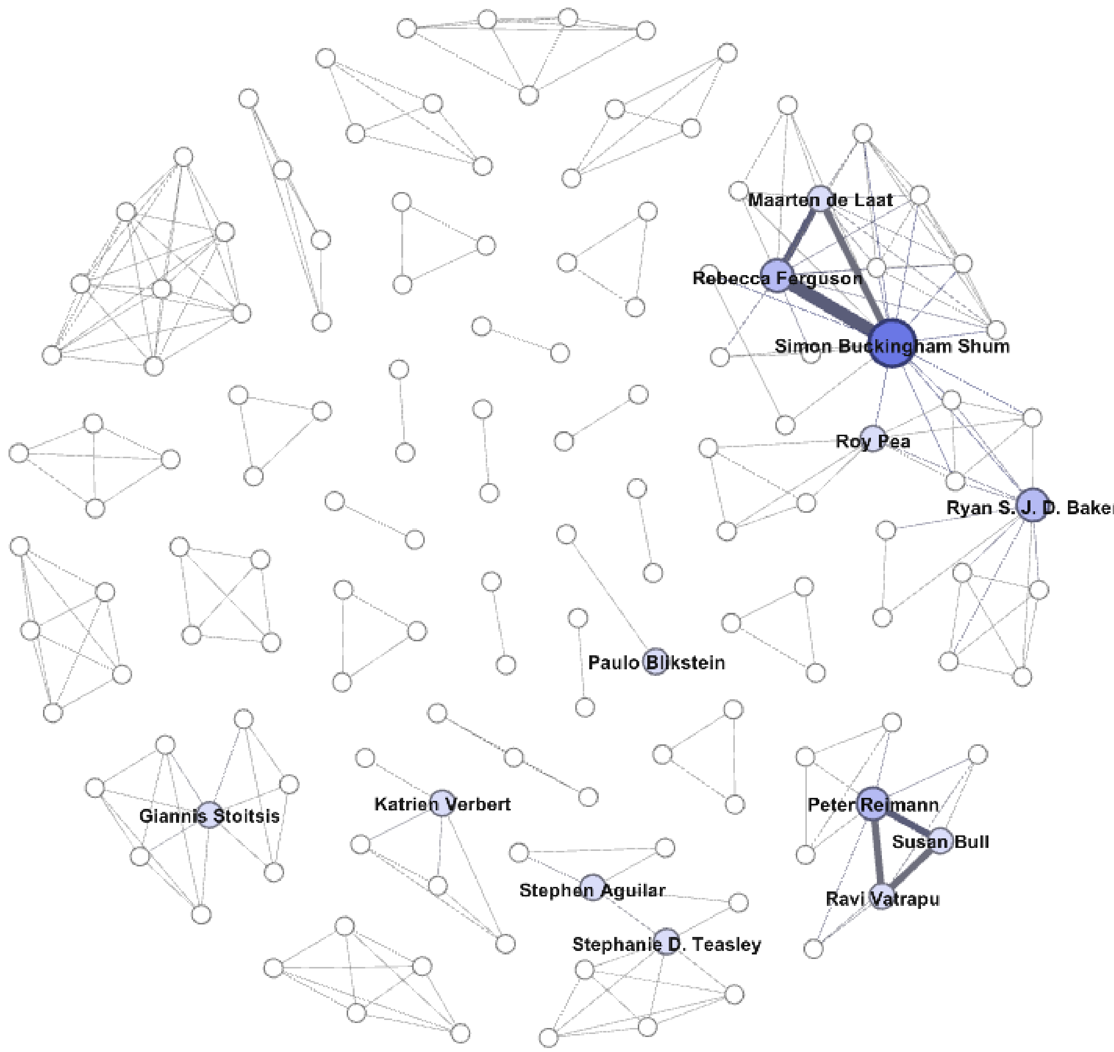

Figure 4. Co-Authorship Network: Size and colour represent the number of published papers from LAK 2013

Co-authorship Network: A useful way to visualize the author community is to draw a co-authorship network. In this network, each node is an LAK ’13 author and each line is a link between co-authors. The strength of the line depends on the number of LAK ’13 papers that the two authors have co-authored (see Figure 4). Most authors, the authors of individual papers, congregate in small cliques. It is interesting to note, however, that a more intricate collaboration network interconnects several important LAK ’13 authors. Simon Buckingham Shum, Rebecca Ferguson, Roy Pea, Maarten de Laat, and

Ryan Baker are part of this sub-network. Other small nuclei of collaboration surround Stephanie Teasley and Peter Reimann. A parallel analysis of all Learning Analytics literature conducted by Dawson, Gašević, Siemens, and Joksimovic (2014) found a similar network configuration.

2.2 Internal Citation Analysis

Scientific papers are not isolated entities; each new paper is linked to previous papers through citations. It is possible to gain insight on the main sources of knowledge in a field through the analysis of which papers and authors are the most referenced and how different papers may be similar based on their number of references in common. The following provides a basic quantitative analysis for the references contained in the LAK ’13 papers:

Basic Quantities: LAK 2013 documents had a total of 929 references, with an average of 21 bibliographic entries per paper, suggesting that the LAK papers draw knowledge from a large number of sources.

Shared References: The references present in LAK 2013 documents were directed to 819 different papers. Only 8% of the referenced papers (68 papers) are shared between at least two publications.

Most Referenced Papers: The most referenced articles of LAK 2013 are presented in Table 1. As can be seen, all of the top-12 papers are specifically Learning Analytics papers, published in previous editions of LAK or in related Learning Technology venues that existed before LAK. This result complements the one found in the previous analysis. Even though small, there is a body of knowledge central to Learning Analytics and new papers in the field use it. The maturity of the field can be judged again by the nature of the most referenced papers. Most are definitions of the concept of Learning Analytics or early examples of its use. This focus of citation on conceptual papers was also found in Dawson et al. (2014). The number of references to article follows a traditional long-tailed distribution.

Most Referenced Authors: The references found in LAK 2013 papers also allow us to find the more relevant scientists in the field, at least from the perspective of this particular conference. The top nine authors can be seen in Table 2. The results, not surprisingly, relate to the general perception of leaders in the field. The distribution of references to authors also follows the long-tailed distribution common to scientific citation.

Bibliographic Coupling Network: Two papers are bibliographically coupled if they both reference the same third paper. The strength of the coupling depends on the number of papers referenced in common. Figure 5 shows a network where each node is a paper published in LAK 2013 and the edges between them represent the strength of their bibliographical coupling. The size of each node represents the number of other papers to which that node is coupled. As Figure 5 shows, most of the papers are coupled to at least one other paper. While still nascent, the strength of the core network (papers 3, 26, 28, 31, 32, 33)[1] seems to validate a common body of knowledge that Learning Analytics considers its own. These papers seem to lay out the bounds of the field. Of the ten disconnected nodes, five are papers (7 is a long paper and 15, 30, 38, and 41 are short papers) and five are workshops and panels, which is not surprising since such documents contain few or no references.

[1] Paper 3: Santos, Verbert, Govaerts, and Duval (2013); paper 26: Clow (2013); paper 28: Dimopoulos, Petropoulou, and Retalis (2013); paper 31: Monroy, Snodgrass Rangel, and Whitaker (2013); paper 32: Dyckhoff, Lukarov, Muslim, Chatti, and Schroeder (2013); paper 33: Camilleri, de Freitas, Montebello, and McDonagh-Smith (2013).

(2014). Analysis and Reflections on the Third Learning Analytics and Knowledge Conference (LAK 2013). *Journal of Learning Analytics*, *1*(2), 5-22.

Table 1. References most cited by LAK 2013 papers

Paper Title	No. of References
Ferguson, R., & Buckingham Shum, S. (2012, April). Social learning analytics: Five approaches. *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (pp. 23–33). ACM.	7
Ferguson, R. (2012, March). The state of learning analytics in 2012: A review and future challenges. Knowledge Media Institute.	6
Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. US Department of Education, Office of Educational Technology, 1–57.	6
Buckingham Shum, S., & Ferguson, R. (2012). Social Learning Analytics. *Educational Technology & Society, 15*(3), 3–26.	5
Arnold, K. E., & Pistilli, M. D. (2012, April). Course Signals at Purdue: Using learning analytics to increase student success. *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (pp. 267–270).	5
Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012, May). The student activity meter for awareness and self-reflection. CHI ’12: Extended Abstracts on Human Factors in Computing Systems (pp. 869–884). ACM.	5
Arnold, K. E. (2010). Signals: Applying academic analytics. *Educause Quarterly, 33*(1), n1.	5
Clow, D. (2012, April). The learning analytics cycle: Closing the loop effectively. *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (pp. 134–138). ACM.	4
Siemens, G. (2012, April). Learning analytics: Envisioning a research discipline and a domain of practice. *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (pp. 4–8). ACM.	4
Oblinger, D., & Campbell, J. (2007). Academic analytics, Educause white paper. Retrieved 20 October 2011.	4
Elias, T. (2011). Learning analytics: Definitions, processes and potential. *Learning, 23*, 134–148.	4
Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 30–32.	4

Table 2. Top Nine Referenced Authors in LAK 2013

Author	Times Referenced	No. Papers in which Referenced
G. Siemens	22	16
R. Ferguson	27	15
E. Duval	27	13
S. Buckingham Shum	20	11
K. Verbert	22	10
R. S. J. d. Baker	21	8
S. Dawson	14	8
A. Pardo	11	8
K. Koedinger	11	8

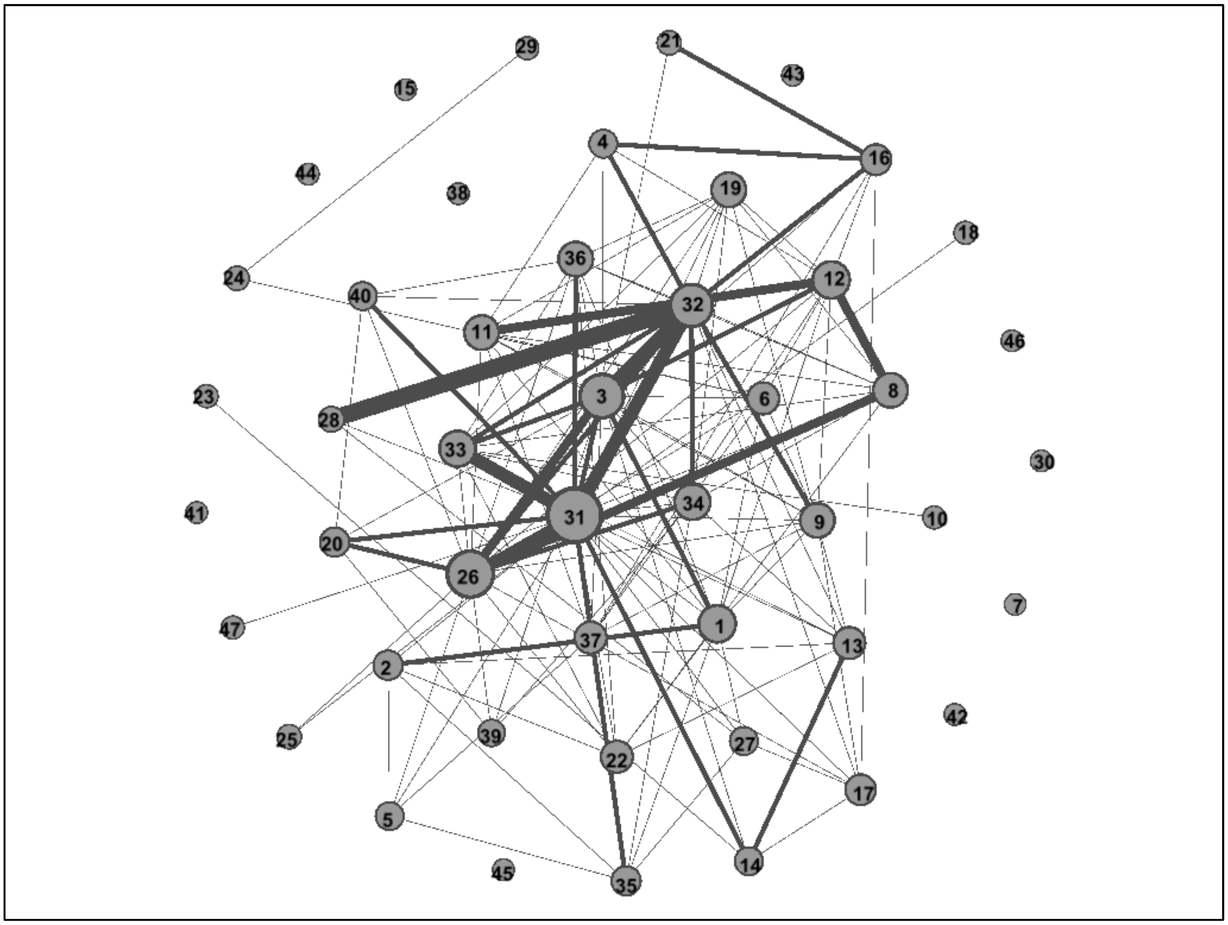

Figure 5. Bibliographic Coupling Network: Each node represents a paper published in LAK 2013

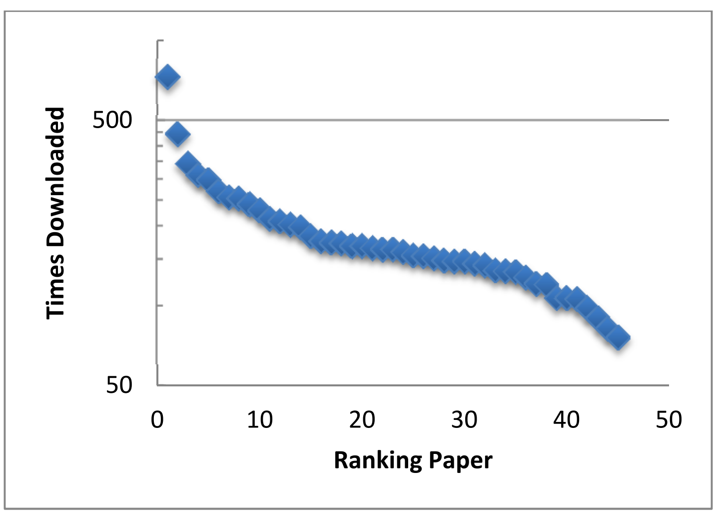

Figure 6. Paper downloads from ACM Digital Library

2.3 External Citation Analysis

While one year on is still too early to determine the relevance of the papers published in LAK 2013 by the number of citations, the number of times they have been accessed or downloaded from the ACM Digital Library could be used as a proxy indication of the impact of the different papers in the field (see Figure 6). The most downloaded paper — "MOOCs and the funnel of participation" (Clow, 2013) — had been downloaded 732 times by the time of this study. The second most downloaded paper — "Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses" (Kizilcec, Piech, & Schneider, 2013) — has been downloaded 444 times. These two papers deal with a trending topic: Massive online open courses (MOOCs). This result suggests that the early visibility of the papers and by proxy, their early impact on the field, is strongly connected to the popularity of the topics they discuss.

2.4 Word Analysis

Counting the number of times a word appears in a set of documents usually correlates to the importance of the underlying concept to the topic. The Carrot[2] clustering engine (Stefanowski & Weiss, 2003) was used to detect the main words used in LAK 2013 papers and to find related words that appear frequently with those important words. This analysis provides an idea of the main concepts discussed in Learning Analytics papers. Figure 7 presents the seven most used, non-stop words in LAK 2013 documents. The size of each word relates to its relative frequency (number of documents in which the word appears divided by the number of documents) in the document set. Seven words have a relative frequency of larger than 0.1 (10%): students, data, analytics, learning, use, activity, and education. These words can be considered the ad-hoc definition of Learning Analytics: use of student and activity data to improve learning and educational processes. A more detailed view of the word analysis can be seen in Figure 8, which shows the most common words that appear together with the most frequent words. Here concepts used in Learning Analytics can be found: student performance, student interaction, datasets, linked data, process analytics, learning environments, learning process, collaborative learning, learning activities, activity theory, activity streams, higher education, educational research, etc. Comparing these words to similar analyses for LAK 2011 and LAK 2012 shows some consolidation in concepts (a lower number of frequent terms) and an increase in the richness of concepts (more related words for each frequent word). In summary, the word analysis suggests a conference with increased focus and depth.

3 CONTENT ANALYSIS

3.1 Identified Topics

The LAK '13 program grouped papers into 14 sessions (Suthers & Verbert, 2013): reflections on learning analytics, visualization for reflection and awareness, social learning analysis and visualization, communication and collaboration, discourse analytics, behaviour analysis, affect analytics, predictive analytics, sequence analysis, MOOCs, assessment, supporting teachers, challenges, and design briefings.

If the classification had been done without restrictions regarding timetable or session length, these papers could have been clustered into six categories:

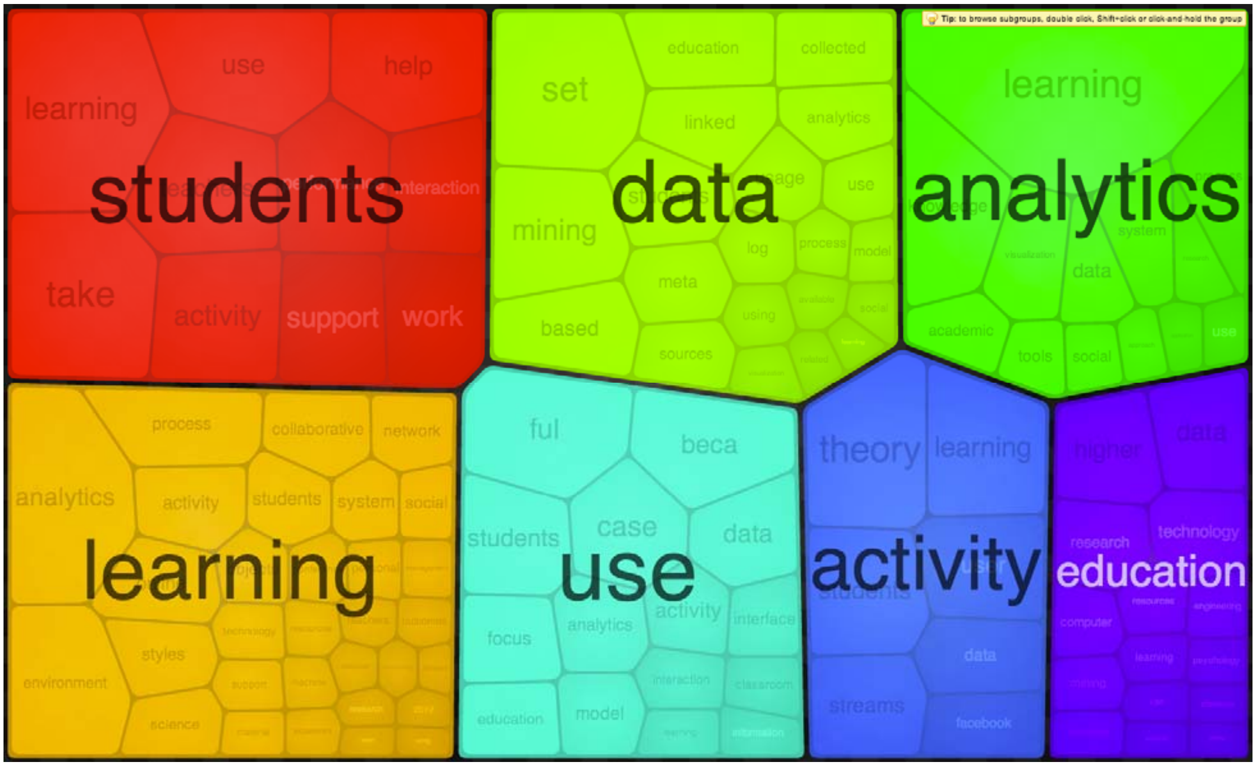

Figure 7. Most frequent words in LAK 2013 papers

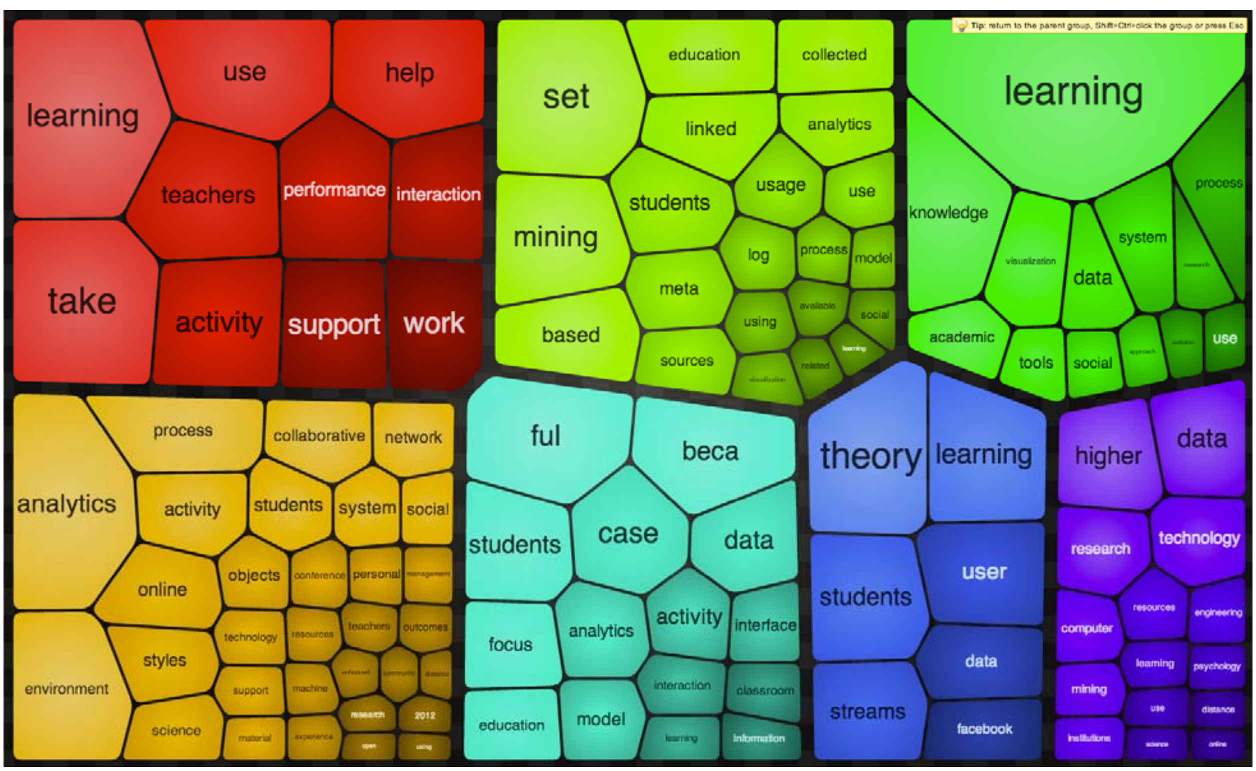

Figure 8. Words related to most frequent words in LAK 2013 papers

1. **Visualization**. Papers in this category focus on increasing reflection and awareness through visualization techniques. Dashboard applications are presented as well as evaluation studies that assess the impact of such dashboards on student learning. Visualization methods appear throughout the other categories, reflecting their relevance to multiple stakeholders.

2. **Behaviour analysis**. The first of four categories of these papers offers advances in making higher-level inferences from low-level actions by capturing and analyzing different data sources (gesture and object manipulation, eye tracking data, data captured from Kahn academy). A second category focuses specifically on effects; for example, to estimate confidence or level of expertise from eye tracking data. A third, predictive analytics, tries to predict at-risk students and to show the development of new media learning skills from participation in Facebook. A final category focuses on analysis of the sequential structure of student actions; for example, to categorize student trajectories for exploring a game environment on fractions.

3. **Social learning analytics**. This area combines both analysis and visualization techniques to gain insight into social behaviour. In the first of three sections, research focuses on analysis and visualization of social interactions to make people aware of their social context. In the second collection, communication and collaboration is analyzed, focusing on both reading and writing activities, including analysis of paragraph level revisions and how collaboration forms around topics in collaborative writing. The last section focuses on discourse analytics from diverse perspectives: scientometric path analysis (citation analysis that includes temporality) to model knowledge evolution; a comparative evaluation classification approach for identifying exploratory dialogue; and theoretical reflections on how learning analytics embody assumptions about both our own and learners' epistemologies, with applications for the analysis of user trace data.

4. **Applications**. Beginning with learning analytics applied to Massive Open Online Courses (MOOCs) in which large numbers of people participate with an open enrollment model, applications of learning analytics form an important topic. Assessment and teacher support — such as by visualizing data collected in real applications for monitoring and direct content development — are both important objectives of several applications.

5. **Challenges**. This group of papers discusses challenges such as scalable learning analytics, ethics and institutional policies for using student data, and collection and sharing of datasets for research purposes. These challenges include the design and development of architectures for organizing the analytic enterprise across micro and macro levels of activity, and across the various tools used in a collaborative learning environment.

6. **Definitional**. A few papers reflect on learning analytics itself, including its theoretical and pedagogical assumptions and examination of the identity and interdisciplinary nature of the field.

3.2 Representative Papers

In this section, we highlight representative papers that address the identified topics. The papers chosen include those nominated for the best paper award and others representative of their categories.

(2014). Analysis and Reflections on the Third Learning Analytics and Knowledge Conference (LAK 2013). *Journal of Learning Analytics*, *1*(2), 5-22.

3.2.1 Visualization

Santos, Verbert, Govaerts, and Duval (2013) present a study that assesses the use of visualizations to support learning analytics. The overall objective is to use learning analytics dashboards that incorporate relevant visualizations to support awareness, self-reflection, and sense-making for learners. Although several learning analytics dashboards have been discussed in recent literature (Verbert et al., 2013), little research has been done to evaluate the usefulness of these dashboards for learners. Santos et al. (2013) present the results of an evaluation study of the StepUp! dashboard, and the extent to which it addresses the issues and needs of learners as collected in brainstorming sessions. The authors then evaluate to what extent the dashboard addresses these issues and needs.

Another example of visualization is the paper by Southavilay, Yacef, Reimann, and Calvo (2013), who visualize document revisions, topic evolution, and collaboration networks based on feedback concerning the individual and collective agency of collaborative writers and their teachers, which leads to the next category of behaviour analysis.

3.2.2 Behaviour Analysis

Some papers in this category utilize sequential analysis at granularities ranging from the individual actions of learners to community learning processes over time. For example, main path analysis is a method from sociometrics for identifying key publications in the development of a scientific field by tracing paths in the directed acyclic graph of citations. Halatchliyski, Hecking, Göhnert, and Hoppe (2013), a best paper nominee, apply main path analysis to the evolution of artifacts in a collaboratively edited Wikiversity to analyze the introduction and evolution of ideas in the community. A flexible analytic workbench supports the work. At the other extreme of granularity, which the authors call “nanogenetic analysis,” Martin et al. (2013) use keystroke-level data to track and classify student trajectories in the state-space of a game.

Two salient and well-received papers in this area explore alternative data sources for learning analytics based on embodied activity. The paper by Schneider, Abu-El-Haija, Reesman, and Pea (2013), a best paper nominee, also exemplifies sequential analysis at the level of individual actions. Schneider et al. used eye tracking on collaborating dyads and represented gaze location and transitions as graphs. They combined the graphs for each dyad to represent shared gaze. Graphs with denser shared gaze locations and transitions indicate mutual orientation towards the task. Further research is proposed to identify network metrics that can predict task performance. Worsley and Blikstein (2013) present research on how to identify expertise of students automatically from object manipulation data and gestures. More specifically, multimodal learning analytics techniques for understanding and identifying expertise as students engage in a hands-on building activity are presented. The authors were able to identify key elements in how to segment and compress object manipulation codes. In addition, they show how dynamic time warping combined with clustering can be used to classify student expertise accurately. The research outcomes also motivate further promising, but challenging, research in the area of behaviour analysis with a wide variety of sensing devices.

Alternative metrics can also be considered within online environments. Wise, Zhao, and Hausknecht (2013) draw our attention to the importance of “invisible behaviour,” such as student message reading, as well as message writing in online forums. They also discuss a framework for applying learning analytics that includes their integration with the goals of learning, student agency in interpreting diverse measures, providing time and space for learner reflection on analytics, and supporting discussion between students and instructors as partners.

3.2.3 Social Learning Analytics

Social learning analytics (Buckingham Shum & Ferguson, 2012) is one of the more active areas of research and discussion; the most cited paper at LAK ’13 was an overview of the topic (Ferguson & Buckingham Shum, 2012). Interesting work on social learning analytics, focusing on understanding the literacies embedded in social media participation, was presented by Ahn (2013). The research employed an exploratory factor analysis to examine how raw activity data may group into particular participatory practices, as well as regression models that explore whether these participatory practices predict new media literacy skills. More specifically, data from Facebook was used to examine relationships between the behaviours of learners resulting in a promising approach to identify new media literacy skills related to 21st century skills. Many papers in behaviour analysis also addressed social learning concerns.

3.2.4 Applications

The application of learning analytics in MOOCs is of major interest, and important challenges related to both popularity and low completion rates of learners are a central concern in MOOCs. The highly downloaded short paper by Doug Clow (2013) quantifies the drop-off of activity in three online learning sites. The full paper by Kizilcec, Piech, and Schneider (2013) goes further, presenting a classification method that identifies different engagement trajectories in MOOCs based on learner interaction with video lectures and assessments: “completing,” “auditing,” “disengaging” and “sampling.” With this classifier, the paper posits a promising approach to deconstructing gained disengagement based on interaction data with materials commonly used by MOOCs. The paper also draws our attention to the fact that learners may be accessing MOOCs with different motivations, and that, for many of them, meeting their objectives does not necessarily include completion. At the time of writing, the paper had already been cited 68 times by researchers in the field, illustrating the high potential of this work for the learning analytics field (source: Google Scholar).

3.2.5 Challenges

A challenge articulated by several researchers is the collection of datasets that can be shared for research purposes (Drachsler et al., 2010; Verbert, Manouselis, Drachsler, Duval, 2012). The overall objective is to collect data captured in real-life settings, from different learning environments, and to make such data available for researchers to enable comparison of research results. Niemann, Wolpers, Stoitsis, Chinis, & Manouselis (2013) present research regarding interchanging data collected from different learning environments. The authors first studied the data types and formats that learning environments use to represent and store learner data. In addition, they developed crosswalks between different schemas, so that datasets can be combined. The paper presents interesting issues that must be addressed before aggregated sets of learner data can actually be used to support learning analytics research.

3.2.6 Reflections

Two of the LAK ’13 best paper nominees offered reflections on the field itself. A thought provoking paper by Knight, Buckingham Shum, and Littleton (2013) examined the relationships between pedagogy, assessment, and epistemology, challenging us to identify where we stand in our own learning analytic practice. The authors outline the analytic implications of several established pedagogic approaches; for example, that “transactional approaches may emphasise use of facts; constructivist the broad (and contextual) application of skills; subjectivist the self-efficacy and motivators of students; apprenticeship the dynamic practical based learning which may occur through high level membership of communities of practice; connectivism the ability of students to build up, link and curate their knowledge 'networks.'" Thus, reasoning in converse, learning analytics are not neutral but may imply or encourage certain pedagogical views. The authors go further to argue that learning analytics embody epistemological assumptions about the nature of knowledge in how they approach the assessment of knowledge. Becoming aware of these epistemological assumptions is consequential because learning analytics can be used to perpetuate assessment regimes "detrimental to the wider enterprise of education," or alternatively assessment regimes that emphasize learner agency and the contextualized nature of knowledge. The paper is recommended for those wanting to approach their analytic practice in a deliberate and reflective manner.

Moving now to an even broader perspective, Balacheff and Lund (2013), grapple with the question of how multiple disciplines come together in learning analytics, and its relation to other research areas, such as educational data mining. In multidisciplinarity, the subject of study is approached from each disciplinary perspective without theoretical, conceptual, or methodological integration, while interdisciplinary research involves such integration. A bridge between the two may be found in strategies of "productive multivocality" (Suthers, Lund, Rose, Teplovs, & Law, 2013), methods of bringing together the various "voices" of different theoretical and methodological traditions in dialogues that not only provide different points of view on the subject but also expose assumptions and make epistemological positions explicit. Balacheff and Lund (2013) call for examining and making explicit the "problématiques" of learning analytics: the coherent frameworks by which we express problems and why they are interesting to solve along with the approaches for solving them. They challenge us to confront "the different understandings of Learning Analytics, without necessarily choosing any one particular definition as the one that is destined to become canonical."

4 CONCLUSIONS

The main findings of the analyses conducted on LAK '13 can be summarized as follows:

- LAK '13 was the venue of an open and international community of researchers. This community currently consists of a core of 40 researchers that attracts an orbit of 100 additional authors.
- The statistical distributions of the author community of LAK '13 are in line with that of other scientific communities.
- The body of knowledge central to learning analytics is still small but already shows signs of consolidation.
- Five main topics have been identified in the LAK '13 program — visualization, behaviour analysis, social learning analytics, learning analytics for MOOCs, and learning analytics issues (ethical, scalability, etc.) — as well as a sixth topic of reflecting on the field itself.
- Learning analytics is at a crucial stage in its development that calls for individual and collective reflection and dialogue to arrive at an understanding of the field. This is not necessarily a single core definition, but rather a locus for productive dialogue and collaboration between multiple theoretical, methodological, and practical perspectives concerning what problems are important to solve and how the analytics we are promoting reflect our epistemological assumptions and pedagogical stances.

This kind of analysis should be performed periodically to gain insight on the evolution of the field, as well as to confirm the healthy status of its surrounding community.

ACKNOWLEDGMENTS

Katrien Verbert is a Post-doctoral fellow of the Research Foundation – Flanders (FWO)

REFERENCES

Ahn, J. (2013). What can we learn from Facebook activity? Using social learning analytics to observe new media literacy skills. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 135–144). New York: ACM.

Arnold, K. E. (2010). Signals: Applying academic analytics. *Educause Quarterly, 33*(1), n1.

Arnold, K. E., & Pistilli, M. D. (2012, April). Course Signals at Purdue: Using learning analytics to increase student success. *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK ’12), 29 April–2 May, Vancouver, BC, Canada (pp. 267–270). New York: ACM.

Balacheff, N., & Lund, K. (2013). Multidisciplinarity vs. Multivocality: The case of “Learning Analytics.” *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 5–13). New York: ACM.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. US Department of Education, Office of Educational Technology, Washington, D.C., 1–57.

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Educational Technology and Society, 15*(3), 3–26.

Camilleri, V., de Freitas, S., Montebello, M., & McDonagh-Smith, P. (2013). A case study inside virtual worlds: Use of analytics for immersive spaces. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 230–234). New York: ACM.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *Educause Review, 42*(4), 40.

Clow, D. (2012, April). The learning analytics cycle: Closing the loop effectively. *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK ’12), 29 April–2 May, Vancouver, BC, Canada (pp. 134–138). New York: ACM.

Clow, D. (2013). MOOCs and the funnel of participation. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 185–189). New York: ACM.

Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014). Current state and future trends: A citation network analysis of the learning analytics field. *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (LAK ’14). 24–28 March, Indianapolis, Indiana, USA (pp. 231–240). New York: ACM.

Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management*, *37*(6), 817–842.

Drachsler, H., Bogers, T., Vuorikari, R., Verbert, K., Duval, E., Manouselis, N., Beham, G., Stern, H., Lindstaedt, S., Friedrich, M., & Wolpers, M. (2010). dataTEL: Issues and Considerations regarding Sharable Data Sets for Recommender Systems in Technology Enhanced Learning. *Procedia Computer Science, 1*(2): 2849–2858.

Dimopoulos, I., Petropoulou, O., & Retalis, S. (2013). Assessing students’ performance using the learning analytics enriched rubrics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 195–199). New York: ACM.

Dyckhoff, A. L., Lukarov, V., Muslim, A., Chatti, M. A., & Schroeder, U. (2013). Supporting action research with learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 220–229). New York: ACM.

Elias, T. (2011). Learning analytics: Definitions, processes and potential. *Learning, 23,* 134–148.

Ferguson, R. (2012, March). The State of Learning Analytics in 2012: A Review and Future Challenges. Technical Report KMI-12-01. Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom. http://kmi.open.ac.uk/publications/techreport/kmi-12-01

Ferguson, R., & Buckingham Shum, S. (2012). Social learning analytics: Five approaches. *Proceedings of the Second International Conference on Learning Analytics & Knowledge* (LAK ’12), 29 April–2 May, Vancouver, BC, Canada (pp. 23–33). New York: ACM.

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012, May). The student activity meter for awareness and self-reflection. CHI ’12: Extended Abstracts on Human Factors in Computing Systems (pp. 869–884). New York: ACM.

Halatchliyski, I., Hecking, T., Göhnert, T., & Hoppe, H. U. (2013). Analyzing the flow of ideas and profiles of contributors in an open learning community. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 66–74). New York: ACM.

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, *52*(2), 291–314.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American documentation, 14*(1), 10–25.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 170–179). New York: ACM.

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Epistemology, pedagogy, assessment and learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 75–84). New York: ACM.

Leydesdorff, L., & Wagner, C. S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics, 2*(4), 317–325.

Martin, T., Aghababyan, A., Pfaffman, J., Olsen, J., Baker, S., Janisiewicz, P., Phillips, R., & Petrick Smith,

C. (2013). Nanogenetic learning analytics: Illuminating student learning pathways in an online fraction game. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 165–169). New York: ACM.

Monroy, C., Snodgrass Rangel, V., & Whitaker, R. (2013). STEMscopes: Contextualizing learning analytics in a K–12 science curriculum. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 210–219). New York: ACM.

Niemann, K., Wolpers, M., Stoitsis, G., Chinis, G., & Manouselis, N. (2013, April). Aggregating social and usage datasets for learning analytics: Data-oriented challenges. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 245–249). New York: ACM.

Oblinger, D., & Campbell, J. (2007). Academic analytics, Educause white paper. Retrieved 20 October 2011 from https://net.educause.edu/ir/library/pdf/pub6101.pdf

Santos, J. L., Verbert, K., Govaerts, S., & Duval, E. (2013). Addressing learner issues with StepUp! An evaluation. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 14–22). New York: ACM.

Schneider, B., Abu-El-Haija, S., Reesman, J., & Pea, R. (2013). Toward collaboration sensing: Applying network analysis techniques to collaborative eye-tracking data. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 107–111). New York: ACM.

Siemens, G. (2012, April). Learning analytics: Envisioning a research discipline and a domain of practice. *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK ’12), 29 April–2 May, Vancouver, BC, Canada (pp. 4–8). New York: ACM.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 30–32.

Southavilay, V., Yacef, K., Reimann, P., & Calvo, R. A. (2013). Analysis of collaborative writing processes using revision maps and probabilistic topic models. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April 2013, Leuven, Belgium (pp. 38–47). New York: ACM.

Stefanowski, J., & Weiss, D. (2003). Carrot^2 and language properties in web search results clustering. In E. Menasalvas, J. Segovia, P. S. Szczepaniak (Eds.), *Advances in Web Intelligence* (pp. 240–249). Lecture Notes in Computer Science Volume 2663. Berlin/Heidelberg: Springer.

Suthers, D. D., Lund, K., Rosé, C. P., Teplovs, C., & Law, N. (2013). *Productive Multivocality in the Analysis of Group Interactions*. New York: Springer.

Suthers, D. D., & Verbert, K. (2013). Learning analytics as a “middle space.” In D. D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13), 8–12 April, 2013, Leuven, Belgium (pp. 1–4). New York: ACM.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago, IL: University of Chicago Press.

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Educational Technology and Society, 15*(3), 133–148.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2013). Learning dashboards: An overview and future research opportunities. *Personal and Ubiquitous Computing (PUC) Journal*, 16 pages, Springer.

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 48–56). New York: ACM.

Worsley, M., & Blikstein, P. (2013). Towards the development of multimodal action based assessment. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 94–101). New York: ACM.

(2014). Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space. *Journal of Learning Analytics*, *1*(2), 23-47.

Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space

Simon Knight, Simon Buckingham Shum
Knowledge Media Institute

Karen Littleton
Centre for Research in Education & Educational Technology
The Open University, UK

simon.knight@open.ac.uk

Abstract: Learning Analytics is an emerging research field and design discipline that occupies the "middle space" between the learning sciences/educational research and the use of computational techniques to capture and analyze data (Suthers & Verbert, 2013). We propose that the literature examining the triadic relationships between *epistemology* (the nature of knowledge), *pedagogy* (the nature of learning and teaching), and *assessment* provide critical considerations for bounding this middle space. We provide examples to illustrate the ways in which the understandings of particular analytics are informed by this triad. As a detailed worked example of how one might design analytics to scaffold a specific form of higher order learning, we focus on the construct of epistemic beliefs: beliefs about the nature of knowledge. We argue that analytics grounded in a pragmatic, socio-cultural perspective are well placed to explore this construct using discourse-centric technologies. The examples provided throughout this paper, through emphasizing the consideration of intentional design issues in the middle space, underscore the "interpretative flexibility" (Hamilton & Feenberg, 2005) of new technologies, including analytics.

Keywords: Learning analytics, epistemology, pedagogy, educational assessment, discourse analytics, social learning analytics

1. INTRODUCTION

> *Assessment is one area where notions of truth, accuracy and fairness have a very practical purchase in everyday life. (Williams, 1998, p. 221)*

Assessment sits at the heart of learning, but is hugely controversial. This is directly relevant to learning analytics, because — we argue — learning analytics implicitly or explicitly promote particular assessment regimes. Presently, many education systems are predicated on assessment regimes seeking to accredit knowledge and skills gained by students through formal assessments — often exam based. Proponents of such exams suggest they are the fairest way to assess competence and learning under controlled, reliable conditions. Assessment, pedagogy and curriculum are fundamentally related (Harlen, 2007), but many regimes of what has come to be termed "high stakes" testing are criticized. For example, standardized assessments, including the Programme for International Student Assessment (PISA), American Standardized Assessment Tests (SATs), and English National Curriculum Assessments (Sats) face myriad problems. Not least among these is that the exams are criticized comprehensively (e.g., Davis, 1999; Gardner, 2011; Hopmann, Brinek, & Retzl, 2007) on two key validity criteria: for failing to represent adequately the types of problems people are likely to face in their everyday lives (external

(2014). Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space. *Journal of Learning Analytics*, *1*(2), 23-47.

validity), and that they fail to represent an adequate conceptualization of what it means to know — of what knowledge is (internal validity). The latter claim is that while assessments clearly measure something, a good grade does not necessarily reflect mastery (Davis, 1999). These fundamental issues are highlighted in a significant body of research (e.g., Davis, 1999; Gardner, 2011; Hopmann et al., 2007), and one of the objectives in writing this paper is to clarify the implications of these issues for the Learning Analytics community.

In this paper, Section 2 considers the relationship between assessment systems and the sorts of epistemic challenges students might encounter. Section 3 introduces the concept of epistemic beliefs, and Section 4 goes on to discuss the relationships between learning analytics, and the triad of epistemology, pedagogy, and assessment. Section 4.2.1 then introduces pragmatic, socio-cultural approaches to learning analytics, which we suggest are well placed to probe or assess facets of learning that other learning analytics may not adequately address. To exemplify this argument, we draw a parallel between the psychometric measurement of epistemic beliefs and high stakes testing regimes (Section 5). The parallel is expanded in Section 6, which offers a detailed example of how our own orientation might make design decisions in the Learning Analytics "middle space." We conclude by reflecting on some broader ways in which these considerations might play out in various analytic approaches, maintaining the significance of the triad throughout.

2. WHY WORRY ABOUT EPISTEMOLOGY?

A primary concern of this paper is the relationship between epistemology, pedagogy, and assessment. Epistemology is the philosophical study of what knowledge is, and what it means for someone to "know" something. Central to the field of epistemology are questions regarding the nature of truth, the nature of justification, and types of knowledge, e.g., knowing *how* (skills) or knowing *that* (facts). Whatever "knowledge" is, *"it is uncontroversial, pre-philosophically, that education aims at the imparting of knowledge: students are educated in part so that they may come to know things"* (Siegel, 1998, p. 20). Thus, pedagogy may in part be seen as the study of *how* to impart this knowledge to students — the science and development of approaches to teaching and learning for knowledge. However, epistemology's relationship to the more familiar concepts of pedagogy and assessment is a topic of educational debate (Davis, 1999; Dede, 2008; Kelly, Luke, & Green, 2008; Williams, 1998), and we will consider this in relation to learning analytics throughout this paper.

Harlen (2007) depicted a triadic relationship between pedagogy, assessment, and practice. Influenced by this, and Katz's (2000) description of "competency, epistemology and pedagogy: curriculum's holy trinity" we depict the triad as in Figure 1.[1]

In this picture, epistemology *could* be seen as driving assessments aimed at uncovering student knowledge, and driving pedagogy to build high quality knowledge to that end. In this view, assessment is targeted at the learning of high-level knowledge — it is assessment *for* learning. However, these relationships are not fixed; neither pedagogies nor epistemologies necessarily *entail* the other (Davis & Williams, 2002) (although they may implicate). Furthermore, as we will explore in this paper,

[1] We could also introduce the notion of "folk psychology" as a mediating factor between a teacher's views on knowledge and pedagogy; for example, if we hold that some (particular) children will *never* learn x, we are unlikely to attempt to teach it (a pedagogical "move") regardless of our epistemological stance regarding the nature of "x" (Olson & Bruner, 1996). However, in that paper, Olson and Bruner implicate epistemology in a number of their points regarding "folk pedagogy."

assessment is always concerned with devising proxies for "knowledge," around which there are philosophical (epistemological) and methodological issues. Some epistemological stances hold that it is not possible to "map" the knowledge that students hold onto their responses in assessments in reliable and valid ways. This issue is further confounded by the methodological limitations of all assessment methods, and by extension learning analytics. The adequacy of analytic techniques to illuminate learning should be open to enquiry.

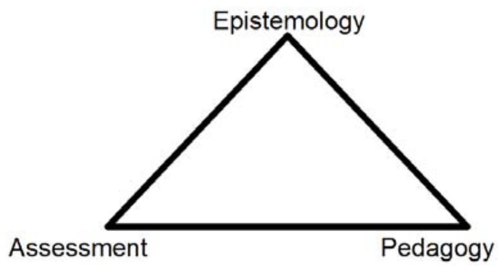

Figure 1. The Epistemology–Assessment–Pedagogy triad

The situation is, therefore, a complex one — which facet of the triad has primacy over the others is not clear in either theory or practice, and may be dynamic according to need and circumstance. However, relationships between the three can certainly be identified, and throughout this paper, we draw out some of these with respect to learning analytics — which may be conceptualized as a component of assessment. Furthermore, we suggest that, although the relationship may not be a necessary one, assessment regimes do implicate particular epistemological stances. Our position is that attention to this middle space — bounded by epistemology, assessment, and pedagogy — provides key considerations for the design of learning analytics.

Consider the following example from Denmark to illustrate the argument that, implicitly or explicitly, epistemological assumptions fundamentally shape pedagogy and assessment, and hence, the kinds of learning analytics that one deploys to achieve those ends. In Denmark, a pilot project was conducted permitting the use of the Internet (but not communication sites) to support students in five school-leaver subject exams.[2] This made it possible to set questions requiring the use of multimedia and individual Internet searches where students may be given unfamiliar resources and permitted to source information for themselves from the Internet. For example, a student might be asked to write about a poet whom they have not studied (preventing rote learning), having been provided with a poem, plus another by a contemporary, a short biography, and perhaps an image from the time. Thus, while Danish students are expected to evidence "knowledge-that" — knowledge of facts — they must also exhibit a higher level of "knowing-how," for example around information processing, synthesis, and metacognitive abilities — which remain unassessed in countries restricting access to external resources that might enhance the student's capability. While this, of course, remains a controlled assessment context, the example illustrates how even within a system reliant on exams, those exams might be conducted on a rather different epistemological grounding. Assessment regimes such as the Danish example may be taken to reflect a holistic epistemology in which *how* one comes to know is as important as *what* one comes to know, and in which it makes little sense to pick out individual tokens of knowledge in decontextualized ways (Davis & Williams, 2002; Davis, 1998, 2005; Katz, 2000).

[2] Steen Lassen (a Danish Education Minister) on the piloting of Internet access in exams: http://vimeo.com/8889340 subsequently adopted by some Danish universities (Cunnane, 2011).

We can contrast such assessments with high stakes testing regimes whose construct validity and external validity have been questioned. For instance, Davis (1999; 2006) argues that such instruments neither assess those facets of learning they set out to test, nor those facets of learning that would likely be utilized in the everyday deployment of knowledge in any particular domain. Davis has argued that high stakes testing is inadequate for understanding learning, in so far as its construal of that learning is necessarily restricted by a desire for highly reliable metrics of success. As such, it must exclude the nuanced understanding of student meaning-making, the social context in which learning occurs, and how knowledge is constituted and enacted. He argues that this, as opposed to acquisition, is the appropriate way to talk about knowledge. Davis draws on notions of situated cognition (Salomon, 1996) and socio-cultural approaches (Säljö, 1999) — particularly Säljö's "Literacy, Digital Literacy and Epistemic Practices: The Co-Evolution of Hybrid Minds and External Memory Systems" (Säljö, 2012). Säljö highlights that:

> From the learning and literacy points of view, such tools [memory aides and knowledge management systems of various sorts] imply that users' knowledge and skills, as it were, are parasitic on the collective insights that have emerged over a long time and which have been entered into the instrument in a crystallized form: algorithms, grammatical rules and concepts, etc. The user will manipulate the artificial memory system in a number of ways in order to see what comes out of the processing that goes on in the machine. (Säljö, 2012, p. 14)

However,

> Engaging with external memory systems thus requires familiarity with a varied set of epistemic practices that range from deciphering letters on a page through familiarity with meaning-making in relation to discourses and genres of texts and other media, to meta-knowledge about how such resources may be used. (Säljö, 2012, p. 12)

Säljö makes an epistemological claim, specifically, a socio-cultural, pragmatist claim: that there are important literacies and practices to be mastered in learning; that those should themselves be objects of assessment; and language and discourse are critical filters on our grasp of the world. Such an epistemology has implications for how we teach, what we assess, and which analytics techniques might be deployed. "Success" can no longer be defined as a matter of regurgitating, unaided, the correct information in a two-hour exam. Such an epistemology also, we argue, offers a perspective on why even technologically advanced societies such as Denmark that assess knowledge in less abstracted, socially embedded ways, information seeking and processing via the Internet and search engines remains a significant challenge for students (UAGU, 2010, p. 15) Although the Internet provides wider access to information, information is not knowledge. Student engagement with information should consider both the kinds of knowledge that we might call transferable competencies or skills — including those higher order skills often known as metacognitive abilities — and more propositional or fact-based knowledge. In this context, we might consider information management and seeking not only as a means to an ends, but as a way to encourage interaction with a complex network of information. As argued by Tsai, as not only:

> …a cognitive tool or a metacognitive tool; rather, it can be perceived and used as an epistemological tool. When the Internet is used as an epistemological tool for instruction, learners are encouraged to evaluate the merits of information and knowledge acquired

> from Internet-based environments, and to explore the nature of learning and knowledge construction. (Tsai, 2004, p. 525)

In this conception, learners are encouraged to think about the context, reliability, validity, certainty, and connectedness of knowledge.

To summarize, this section has argued that a consideration of epistemology is important to learning analytics in two related senses:

- The ways that we assess, the sorts of tasks we set, and the kinds of learning we believe to take place (and aim for) are bound up in our notions of epistemology. Learning analytics are not objective or neutral: data does not "speak for itself" but has been designed by a team who, implicitly or explicitly, perpetuate the pedagogical and epistemological assumptions that come with any assessment instrument.

- The Danish example shows concretely how epistemology relates to assessment regimes. When knowledge is seen as something that can only be evidenced in contextualized activity, and when it is embedded in one's physical and digital environment, the role of the Internet is redefined as a metacognitive tool that cannot be excluded in assessment.

These epistemological considerations foreground the quality of a student's *enquiry processes* as important, not just whether they get the right answer. An aspect of that enquiry process is epistemic beliefs — and it is to these that we will turn shortly.

Before we do so, we should note the importance of the whole triad in bounding the *middle space* of learning analytics, as defined by the programme chairs of the 2013 learning analytics conference:

> In summary, although individual research efforts may differ in their emphasis, we believe that all research in Learning Analytics should address the "middle space" by including both learning and analytic concerns and addressing the match between technique and application. Advances in learning theory and practice are welcome, provided that they are accompanied with an evaluation of how existing or new analytic technologies support such advances. Advances in analytic technologies and methods are welcome, provided that they are accompanied with an evaluation of how understanding of learning and educational practices may be advanced by such methods. (Suthers & Verbert, 2013, p. 2)

As noted earlier, it is our claim that epistemology is fundamental to the understanding of assessment, and that learning analytics fits in this scope. Furthermore, there is a bi-directional influence. Thus, design choices around assessments, often practical ones, may push towards particular epistemological stances, for example in the case of high stakes testing. In the other direction, sometimes explicitly epistemological stances, as in the Danish example, lead to particular forms of assessment being privileged. The role of *learning* here is also important; we have noted that the focus of pedagogy can be taken to be the bringing about of knowledge in students, as conceptualized by epistemology. However, practical considerations are again in play here, including the particular sorts of information privileged. Particularly pertinent is the role that we intend learning analytics to play in assessment: learning analytics that provides summative feedback is likely to relate to different types of learning, analytics, and outcome than learning analytics based on the provision of formative feedback.

(2014). Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space. *Journal of Learning Analytics*, *1*(2), 23-47.

The "interpretative flexibility" of new technologies, including analytics, is high: when we consider appropriation of technology within particular social settings, we should be mindful of not falling into the trap of technological determinism (Hamilton & Feenberg, 2005). As Crook and Lewthwaite (2010) note, our expectations for technology for transformative change should be mitigated by an understanding of those technologies in wider educational systems. Moreover, we should understand that technology's influence comes about through pedagogic change, not out of technology's direct effects (Crook & Lewthwaite, 2010). Indeed, the epistemological stance taken in this paper is one well aligned with approaches that situate technology in its wider application, in non-deterministic ways (Oliver, 2011).

As noted above, the adequacy of analytic techniques to meet learning should be open to enquiry. This involves understanding the design trade-offs one is negotiating (intentionally or unintentionally) in developing and deploying technological solutions. This paper argues that this analysis — this bounding — fundamentally involves understanding epistemology, pedagogy, and assessment and their interrelationships in the context of learning analytics. In Section 6 we return to these considerations in the context of the example introduced below — that of epistemic cognition.

3. EPISTEMIC BELIEFS

One facet of students' dynamic interaction with the world of information relates to how they conceptualize the information needed to answer a question — their *epistemic beliefs* regarding the nature of the question and how it may be answered. The sorts of assessment, and pedagogy, to which students are exposed will shape the types of epistemic challenge they encounter in their education — systems with a focus on "right answerism" and limited access to external epistemic resources offer fewer opportunities for challenging knowledge claims (Davis, 1999; Katz, 2000). This paper thus differentiates two related concepts:

1. **Epistemology,** which we introduced above, relates to the philosophical analysis and conceptualization of curriculum content and assessment for knowledge

2. **Epistemic Beliefs,** which we now introduce, relates to the intrapersonal, psychological conceptualizations that individuals hold regarding knowledge

Indeed, a key component of formative *Assessment for Learning* (AfL) may be the disambiguation of the epistemic requirements of questions — in terms of understanding the question, its context, and the knowledge required to answer the question (Black & Wiliam, 2009). That is, AfL may play a crucial role in guiding a student's epistemic beliefs *on the subject of* the epistemological assumptions built into assessment systems themselves.

Table 1 indicates four dimensions of epistemic beliefs for which there is general agreement across the various models of belief.[3] These dimensions are useful to consider in relation to student understanding of knowledge domains. For example, in the context of search engine tasks, "epistemological beliefs are a lens for a learner's views on what is to be learnt" (Bromme, Pieschl, & Stahl, 2009, p. 8). In such tasks, student search activity may be analyzed using the dimensions in Table 1 (e.g., Mason, Ariasi, & Boldrin, 2011), providing a lens onto students' understanding of their own learning, their task demands, and how they meet those demands.

[3] See, for example, Schraw (2013) for a review of the multiple theoretical frameworks.

Epistemic beliefs are thus one example of the type of construct that learning analytics may probe; however, they are also a particularly *good* example given their relationship to our everyday dealings with the world of information, and their relationship to pedagogy, assessment, and classroom practices (Hofer, 2001). Section 5 will discuss epistemic beliefs in relation to their measurement, but we shall first introduce some established approaches to pedagogy.

Table 1. Dimensions of epistemic belief (adapted from Mason, Boldrin, & Ariasi, 2009, p. 69)

Dimension	**Description**
Certainty of knowledge	The degree to which knowledge is conceived as stable or changing, ranging from absolute to tentative and evolving knowledge
Simplicity of knowledge	The degree to which knowledge is conceived as compartmentalized or interrelated, ranging from knowledge as made up of discrete and simple facts to knowledge as complex and comprising interrelated concepts
Source of knowledge	The relationship between knower and known, ranging from the belief that knowledge resides outside the self and is transmitted, to the belief that it is constructed by the self
Justification for knowing	What makes a sufficient knowledge claim, ranging from the belief in observation or authority as sources, to the belief in the use of rules of inquiry and evaluation of expertise

4. OUR LEARNING ANALYTICS ARE OUR PEDAGOGY

Buckingham Shum (2012) used the shorthand “our learning analytics are our pedagogy” to draw attention to the arguments set out in more detail above: that the types of analytic we chose to deploy, and the ways in which we deploy them implicate particular approaches to learning and assessment. This is particularly important given that any use of analytics will be in the context of a wider educational ecosystem, as Crook and Lewthwaite (2010) noted. It is this relationship between the types of analytic we deploy and our pedagogies that we now consider.

4.1 Pedagogy and Analytics

The relationship between learning analytics and pedagogy is important because they are both bound up in epistemology — what knowledge is. This section explicitly introduces the relationship between a number of established pedagogic approaches and learning analytics. These are not intended as comprehensive reviews, but rather as brief overviews of how the relationship between pedagogy and learning analytics might be conceptualized. The following section expands on some key ideas here, before moving on to explicate the core topic of this paper — a socio-cultural learning analytic — and one proposed instantiation of a learning analytic based on this approach.

4.1.1 Transactional or instructionalist approach

Transactional approaches hold that learning entails the transfer of knowledge from the knower

ISSN 1929-7750 (online).

(teacher) to the learner (student). They are characterized by a perspective on assessment in which success is "out there," assessable in the degree of correspondence between the claims that learners make, and the facts that they have been taught. Clearly there is a role for mastering facts in many curricula, and the technological ease with which these may be automatically assessed underlies their current dominance of learning analytics and automated assessment (formative and summative).

Analytics Implications: Learning analytics based on transactional approaches will tend to focus on simple metrics such as test scores, not requiring deeper analysis of more complex artefacts, or the processes by which they were derived.

4.1.2 Constructivist approach

Constructivist models focus on those forms of learning that occur in the learner's guided exploration of and experimentation with the world, typically in classrooms or online environments. Constructivist models are likely to measure success as quality of construction, with learners experimenting with their environment, and being capable of using tools that are appropriate for their given age.

Analytics Implications: Learning analytics with a constructivist focus will focus on *progress,* particularly through tracking and judging the modifications made to a set of materials, resources, or tools selected and arranged by the educator. An example of analytics in this tradition would be tracking the evolution of digital artefacts within the Scratch visual programming environment and community (Maloney, Resnick, Rusk, Silverman, & Eastmond, 2010).

4.1.3 Subjectivist or affect-based approach

Subjectivist perspectives can be characterized as de-emphasizing learning qua academia, with more attention to personal affect. While individual affect is a concern for educationalists, it is rarely if ever the *overarching* concern in the consideration of learning. One context in which affect is important is learning in complex socio-technical challenges: while there are certainly better and worse answers, there is too much information and no known best solution. Information seeking in such contexts can draw on subjectivist approaches that measure whether the user is "satisfied" with the information they have found. Another context in which self-reporting is an important proxy for learning is research into dispositions (Deakin Crick, Broadfoot, & Claxton, 2004) and "mindsets" (Dweck, 2006) — learner willingness to engage with opportunities that will challenge them, or stretch other transferable competencies such as questioning or collaborating.

Analytics Implications: In tandem with other approaches, learning analytics based on subjectivist approaches are likely to provide motivation assessments for understanding *why* someone is (or is not) engaging in particular actions. Such analytics may focus on self-reporting through survey tools (Buckingham Shum & Deakin Crick, 2012) or affect-based semantic mark-up such as blog tagging (Ferguson, Buckingham Shum & Deakin Crick, 2011), alongside automated approaches such as textual sentiment analysis.

4.1.4 Apprenticeship approach

Apprenticeship approaches are sometimes used in learning analytics with an interest in whether the learner has become part of a community of practice or enquiry. In this view, success is about "being part of" a given group; it is bound up in notions of communities of practice — that "to know x" is to act towards x in some way defined by (or reflected in) the behaviours of some community or other.

Analytics Implications: Analytics based on apprenticeship approaches are likely to focus on classifying expert and novice users, and the shift from novice to expert. Such analysis may explore behavioural markers that mirror those made by "experts," but may not explore the reasons or meanings implicated in such moves. Epistemic Network Analysis of user data from gaming environments is designed to quantify the degree to which learners demonstrate behaviours valued in a professional community (Shaffer et al., 2009). The creation of social capital might be considered another proxy for community membership, overlapping with the next category.

4.1.5 Connectivist approach

Connectivism claims to highlight a perspective on epistemology that translates into a learning analytics framework. Within this view, learning is about understanding how to connect ideas appropriately, and where to find such information. The suggestion is that in the case of the connectivist knower "the act of knowing is offloaded onto the network itself" (Siemens, 2006, p. 33). Within this perspective then, success is about building connections between ideas.

Analytics Implications: Connectivist approaches use network analysis to explore the "connectedness" of a learner's knowledge — in terms of both concepts, and social connections. Analytics would look at how networks' size, quality, and changes over time can serve as proxies for effective learning (Dawson, 2010; Haythornthwaite & de Laat, 2010).

4.1.6 Pragmatic, socio-cultural approach

Pragmatic approaches (building on, for example, Dewey, 1998) hold that learning occurs in the development and negotiation of a mutually shared perspective between learners. Pragmatists suggest that, as human knowers, our conception of some given thing is bound up in our understanding of its practical application — and that is all. When we attempt to understand truth beyond such a conceptualization of practical activity, we are likely to fail. Thus, success is in *use* — the measure of success is how useful the information is for the purposes it is employed; it is socio-culturally embedded and mediated, and may be in flux as activities are defined and redefined.

Analytics Implications: Pragmatic approaches have traditionally focused less on assessing the products of learning (except where they are being used *for* something), and more on the process. Analytics tools in socio-cultural approaches encourage learners to reflect on their own activity, in an attempt to understand how they can develop their skills in information processing, in their own particular contexts. Analytics within this approach might attend particularly to quality of discourse for learning, for creating a mutuality of perspectives (Edwards & Mercer, 1987) including in collaborative information seeking tasks (Foster, 2009; Hertzum, 2008; Lazonder, 2005). Our previous work is in this tradition, drawing on socio-cultural discourse analysis (Mercer & Littleton, 2007), and emerging conceptions of the pragmatic web (Buckingham Shum, 2006). This research foregrounds *how* students interact with information, make sense of it in their context and co-construct meaning in shared contexts. These are on-going processes that highlight the question of how learning analytics fits into the context of AfL and pedagogy.

Having summarized the sorts of relationships we might see between pedagogical approaches and learning analytics, let us turn to epistemology.

4.2 Epistemology and Analytics

The stance we take with regard to the relationship between epistemology, assessment, and learning analytics relates to the issue of whether we place analytics in the role of diagnosis or a kind of biofeedback. Is learning analytics (and assessment) serving as the end-point of, or a component of pedagogy? As a diagnostic, we seek to accredit learning through defining behavioural proxies taken as evidence of knowledge and competencies. As biofeedback, learning analytics is used to support learners in their own self-regulated learning activities, giving them feedback on changes they make and their impact on learning outcomes, but without, generally, making strong evaluative judgments regarding such changes. The former is thus more closely aligned with assessment *of* learning — often instantiated in high stakes summative assessment — while the latter is closer to assessment *for* learning — in which assessment is a continuous process through which formative feedback may be given to further develop student learning (see, for example, Black & Wiliam, 2001; Gardner, 2011). If evidencing process-centric competencies is defined as part of the summative assessment criteria, then the two categories converge. For example, the process competencies of evidencing sound argumentation in discourse, higher resilience when stretched with new challenges, or the creation of social capital within a community of practice, might conceivably be assessed summatively through analytics.

The relationships highlighted in sections 4.1.1 to 4.1.6 serve as general pointers to the sorts of relationships we might see between pedagogy and learning analytics. There we also highlight views on learning, alongside notions of how success may be defined within these approaches; that is, when these systems might accredit knowledge to the student. Fundamentally, accreditation implicates epistemological stances regarding when knowledge may be claimed (or not). The preceding analysis suggests roles for learning analytics in accrediting mastery in three senses:

1. **Mastering curriculum content:** This is the dominant focus of analytics approaches at present, seeking behavioural markers using e-assessment technologies of varying sophistication, in order to generate summaries at varying granularities, for both individuals and cohorts. (See transactional and some constructivist approaches.)
2. **Evidencing membership and processes:** This approach to learning analytics looks for behavioural proxies that indicate a student is part of a particular subgroup; positive feedback is given towards moving students into "successful" subgroups, but little attention is paid to the qualities of those groups except instrumentally. (See affect-based, apprenticeship, and possibly connectivist approaches.)
3. **Success is use:** This approach looks for students developing personal and collective representations of curriculum content and engagement in sensemaking about not only this material, but also their own analytics. One characterization of this family of approaches has been as Social Learning Analytics (Buckingham Shum & Ferguson, 2012). (See connectivist and pragmatist approaches.)

These three broad conceptualizations of learning analytics relate to the issue of whether or not we are deemed to *consume, discover,* or *create* knowledge — is it "out there" for us to take, do we need to investigate to find it, or is it emergent from the contexts in which that knowledge is applied and reified in activity? This is not a claim about learning or pedagogy, but a related claim about the status of knowledge and its assessment. We discuss this further in section 6.4 with reference to a particular example.

4.2.1 Pragmatism and socio-cultural definitions of "context"

The nuance of claims surrounding epistemology and assessment is important. In the introduction we referred to research arguing that conventional exams are designed to maximize the reliability of results, at the cost of straitjacketing what can be defined as learning (poor internal or construct validity) and thus what constitutes evidence of learning (poor external validity). Moreover, if we are to argue that individual tokens of knowledge cannot be identified (and "owned"), then we should accept that "the content of a specific item of knowledge depends in part on how it is related to other knowledge" (Davis, 2006). Thus, socio-cultural setting, interaction, and the purposes for which any artefact or knowledge — in the broadest sense — is being used are all of fundamental importance in understanding how people make meaning and learn. Contextual sensitivity is thus a key facet of pragmatist approaches.

Pragmatic approaches, broadly, are likely to focus on the dynamic nature of information needs, and the discourse and other artefacts that mediate our relationship with information in the world. It is not a postmodern approach, in the sense that postmodern approaches take either a relativist approach (there is no fixed truth) or a normative one (the dominant theme is correct at that time) to knowledge, but rather one that focuses on use, and meaning, over accreditation of facts to things in the world.

Given the salience of context in this approach, it deserves further explication. As with learning analytics generally, context may be taken as very mechanistic; for example, the claim that a person in place/course/role/ability band "x" should see resource "y," or other approaches that would include time, topic, or social-group resource discovery. No doubt, some of these features will prove useful, and indeed the use of semantic web technology in social learning analytics (Ferguson & Buckingham Shum, 2012) may be particularly interesting. However, in addition to computational, context is often defined in terms of temporal, linguistic, aptitude, and geo-spatial metadata. In discourse, however, we draw attention to the following:

1. We emphasize the *discourse in which, and through which, context is constituted* (Edwards & Furlong, 1978; Potter & Edwards, 2003). That is, we take the discourse to have a multifaceted role in constituting, and helping learners make sense of, the context.

2. Discourse is fundamentally associated with the *sense-making* that occurs in respect of any particular task being undertaken; the *use* being targeted is fundamental to context. Stark examples highlight this importance; for example, when we ask students to critique versus summarize a paper, we expect rather different outcomes. Assessment regimes that make this explicit may facilitate the design of analytics that capture the context of "doing x for purpose y."

3. Assessment regimes, and the broad range of tools, technological and otherwise used by learners, also act as mediating artefacts impacting how people perceive their task and its solution, which all shape the context of use.

A conception of learning analytics that recognizes the importance of context is central to epistemic beliefs:

> A sophisticated epistemology entails context-sensitive judgements. Thus, they point out that it is not very sophisticated to view the idea that the earth is round rather than flat as "tentative" whereas theories of dinosaur extinction do require a more tentative stance. (Barzilai & Zohar, 2012, p. 42)

Similarly, building spurious connections between ideas in an attempt to evidence a "complex" view of knowledge is less sophisticated than demonstrating the need for moderation, and so on. Context is thus key to understanding epistemic beliefs, the analysis of which seems highly suited to the *biofeedback* approach to formative assessment analytics, introduced earlier.

The next section further expands this claim in the context of psychological assessment of epistemic beliefs, firstly in "mainstream" psychological approaches, and then that of the discursive approach, which similarly holds context and discourse to be fundamental to understanding thinking. Section 6 then returns to learning analytics, drawing out the relationship between analytics, and the measurement of epistemic beliefs in our illustrative example for socio-cultural, pragmatic analytics.

5. MEASURING EPISTEMIC BELIEFS

The complexity of epistemic cognition suggests a particular perspective on how we are to understand these beliefs. No approach "mirrors" reality with a true, immutable, incontrovertible perspective on a learner's epistemic cognition. This concern is a dual one. Firstly, it is a methodological concern regarding our access to the world, our ability to "get at" what is out there. Secondly, it is a conceptual and psychological concern, regarding the nature of epistemic cognition and whether it itself is stable — developmentally, and across domains — or shaped in some way by resources or beliefs. These two concerns are reflected in the epistemic beliefs literature. Firstly, cognitive developmental models (King & Kitchener, 2004; Kuhn & Weinstock, 2002) suggest that individuals progress through a sequence of increasingly sophisticated epistemic beliefs, while multidimensional perspectives (Hofer, 2001; Schommer, 1990) suggest that epistemic beliefs can be separated into dimensions, within which levels of sophistication can be identified (Greene, Muis, & Pieschl, 2010, p. 248). However, both of these assume a fixed, uni-directional developmental trajectory, where beliefs are seen as global across (and within) domains. The resources view, in contrast, emphasizes the interaction of believer, with resources, highlighting that at various points in any task a cognizer may invoke differing resources (Hammer & Elby, 2003).

Secondly, methodologically the developmental models have tended towards interviews and laboratory tasks, while multidimensional models have emphasized paper and pencil self-report measures (DeBacker, Crowson, Beesley, Thoma, & Hestevold, 2008). Both of these approaches reflect the fixed perspective on beliefs from which theory they stem. Importantly, although three major survey instruments have been developed and deployed, — including in search engine tasks (Lin & Tsai, 2008; Schommer, 1990) — they are heavily criticized for their psychometric properties (DeBacker et al., 2008). Furthermore, while some studies have used interviews (Barzilai & Zohar, 2012; Mason et al., 2009), think-aloud protocols (Barzilai & Zohar, 2012; Ferguson, Bråten, & Strømsø, 2012), or systematic observations (Scherr & Hammer, 2009) such methods may be limited in their insights, particularly where self-report data is to be used and interpreted by researchers. Importantly, they are also not appropriate for the study of online, collaborative, or geographically and temporally spread activities — in particular, online information seeking, or information processing more broadly. These approaches reflect the epistemology of current assessment regimes, as indicated in Section 2, and seem to implicate the view of "fixed" psychological constructs — whether intelligence or epistemic beliefs, as further discussed throughout Section 3.

Building on sections 4.1.1 to 4.1.5, we can identify a number of analytic tools and their relationships to particular forms of data. Some forms of analytics rely on a belief that particular methods (self-report in particular) are: a) true reflections of reality, b) whole reflections of reality (i.e., they cover all the relevant ground), and c) probe "real" constructs. However, while self-report measures may be useful particularly as discussion prompts with students, they are not necessarily *the most* useful approach for many purposes. In both assessment and psychological testing, they suffer from issues of validity (Section 2). Thus, other learning analytics tools may prove more useful.

That is not to write off such tools, which may be particularly useful in exploring population differences and providing basic metrics to distinguish between people (candidates for a job, for example). Rather, it is to note that they do not align well with formative assessment (although some may be used for such purposes); they have a tendency to “teach to the test” — or in the case of psychometrics, to rely on tightly culturally bound constructs; they are often used to make claims about states of the world as “products” rather than an exploration of processes and wider contexts.

In contrast, while those adopting a resources view of epistemic beliefs may also utilize such methods — in particular those involving think aloud and interview data — they also accord well with Österholm’s discursive stance, which suggests that we should not see beliefs and communication as “two separate ‘objects’ that can affect each other, but as more integrated aspects of cognition and/or behaviour” (Österholm, 2010, p. 242). The resources view describes “the activity, the discourse, as the site where epistemological beliefs come to existence, through explicit or implicit references to prior experiences (epistemological resources)” (Österholm, 2009, p. 262). Österholm’s argument is that the resources perspective can be combined with Hammer and Elby’s (2003) resources model. In this model, epistemic beliefs are not viewed as fixed or developing cognitive models ranging over one or more domains, but rather are seen as dependent upon the resources available to the cognizer at any time. This view of epistemic beliefs as “theory-in-action” — in which context, domain, culture, and task conditions interact — accords well with the idea that context is fundamental to understanding meaning.

5.1 Trace Data for Epistemic Beliefs

While Österholm is primarily interested in spoken interactions, learning analytics may extend this interest into the exploration of users’ interactions with artefacts. A tool for such analysis may come through the use of trace data, which is more or less implicitly created by the student. For example, Stadtler and Bromme (2007) analyzed the ways participants found, extracted, and moved information, which could be used to explore information about their beliefs (e.g., visiting few websites indicates trust in those sites visited; Greene et al., 2010). Importantly in this study, users were either given evaluation prompts regarding multiple documents in the medical domain or not; those who received such prompts subsequently recalled more facts and were better able to evaluate sources. If systems of prompts promote laziness, we should be concerned. Where they improve outcomes, however, analytics should explore the best ways to implement them effectively and sustainably to support high quality pedagogy and AfL.

Furthermore, Greene et al. (2010) point out that many behaviours that would ordinarily be difficult to observe can be explicitly elicited in the context of Computer Based Learning Environments (CBLEs), for example:

> …participants who report belief in objective truth and omniscient authority may self-regulate quite differently than participants with a desire to evaluate multiple forms of justification. Likewise, participants who believe in the inherent subjectivity of all knowledge may, on average, select more representations than those who look for an objective truth. (Greene et al., 2010, p. 254)

The claim is thus that epistemic beliefs will be brought to bear on knowledge tasks in ways that can be meaningfully captured, in particular using the digital trace data left as a “by-product” of engaging in tasks (such as issuing searches, browsing websites, or messaging peers), or through the intentional creation of digital artefacts whose evolution can be traced (e.g., documents, bookmark collections, concept maps).

Trace-based learning analytics provides a means to tackle the static, decontextualized view of epistemic beliefs instantiated by questionnaire methods, offering a more authentic perspective on epistemic action than experimental contexts. However, the limit on what can be inferred about the mind of a user solely from low-level system traces is a well known problem in human–computer interaction. While trace data is unobtrusive, it necessarily gives an incomplete picture. In particular, people may have reasons for some behaviours that cannot be probed using such data, ranging from epistemic (as discussed above, for example with regard to the "flat earth" issue), to practical (ICT failures), to pragmatic (the demands of the task place a short time restriction on the activity), and so on. In summary, even though analytics regarding epistemic beliefs may be — at best — a dirty lens onto those beliefs, when analytics are considered *in action* as a tool for sense-making, they may provide an insightful tool for learners to dissect their own metacognitive and self-regulatory behaviours, as well as for educators and researchers to study them.

Trace data thus provides one means by which epistemic beliefs could be examined. However, trace could refer to many things, and as discussed in sections 4.1.1 to 4.1.5, the data collected may not represent an appropriate teaching epistemology, nor capture adequately student epistemologies (see section 2). The next section will discuss some learning analytics that may address this issue. An interesting notion then, is attempting to delve further into the sense-making significance behind particular semantic moves in a given environment. Thus, Greene et al. (2010) (see Section 5.1) described one method of trace analysis for epistemic beliefs built on information moves. Other examples of such trace capture could also be structured to gather student data in particular ways — some of which may be quite naturalistic (capturing search queries, or Facebook posts to explore "problems" encountered, or interactions made; De Liddo, Buckingham Shum, Quinto, Bachler, & Cannavacciuolo, 2011), and others of which might push students into information structuring activity in which they would not otherwise engage, such as argument mapping.

However, in encouraging such structuring by learners, and claiming to capture information about what they are doing, some may argue that we are simply reifying the constructs we have set out to explore. That is, if we are interested in epistemic beliefs, and set up a system to push students to make epistemic beliefs explicit, it does not matter whether those students *have* underlying epistemic beliefs because the system forces them into making some (it makes them reify). While for psychologists who wish to uncover underlying beliefs using the kind of psychometrics described above this is problematic, we do not see this as a concern for our project, because in our discursive, socio-cultural, pragmatic approach the interest is in beliefs as "theory-in-action." In this view, the claim is not that the measurement of beliefs is not possible, but rather that when we take measurements, the discursive context is fundamental to the practices being observed, and the ways that the beliefs are instantiated in action. Thus, learning analytics provides a means to tackle the static, decontextualized view of epistemic beliefs instantiated by questionnaire methods, offering a more authentic perspective on epistemic action than experimental contexts.

6. EXAMPLE: DISCOURSE-CENTRIC ANALYTICS FOR EPISTEMIC BELIEFS

We have reviewed the ways in which the educational triad bounded by epistemology, assessment, and pedagogy defines critical aspects of the "middle space" between learning and analytics, which can lead to very different kinds of analytics. We now present a more developed example, to illustrate how the preceding discussion can translate into tangible design.

Within a pragmatic socio-cultural perspective on the triad, *use* of knowledge, and an understanding of its rich culturally embedded connectedness is fundamental. We have emphasized the role of *context* at various points throughout this piece, indicating in particular the need to understand discourse as created in, and a creator of, context. The perspective is thus fundamentally *social* in nature, taking knowledge to be *discursive* in nature, with learners benefiting in particular from timely *formative* feedback in which language is used to construct knowledge.

Standardized assessments, just as psychometric tests, are good at population-level analytics. However, they are problematic at individual level analysis, at understanding change, and at understanding individual traits and capabilities. The perspective we have outlined in this paper suggests analytics may play an important role in assessment *for* learning, providing formative, individualized, feedback to learners and their teachers.

In the following example, we exemplify such an approach; however, while we hope the example given is illustrative, it is not intended to be proscriptive nor indeed do we claim it is without faults; rather, we seek to illustrate how the stance we have taken here has been "brought out" in the analytics — how we have brought the learning and the analytics together, using the triad to bound that middle space.

This example offers an illustration of how an analytic device might be used to offer formative feedback. In this case, our focus is not explicitly on the content level knowledge claims we can make regarding student learning, but rather on how students develop their understanding. The example offers a means to explore supporting student learning through formative feedback, and social learning of a discursive nature. Of course, further consideration should be given to the wider context of the activity and discourse we would wish to provoke.

Our example comes from work at the Open University, based around the Cohere social annotation and knowledge mapping tool (Buckingham Shum, 2008) and subsequent work on its potential for socio-cultural discourse-centric learning analytics (De Liddo et al., 2011). Cohere is a web application for mapping ideas, concepts, and arguments that can be annotated directly onto source websites. Users enter ideas — nodes with meaningful classifications — and are then invited to "make the connection" with meaningfully labelled edges, to create a conceptual graph. Both ideas and connections may also be tagged, to add a further level of semantic data. Cohere is designed as a tool to enable users to build their own structures, but also to share these, and integrate the nodes and connections of other users, thus building up communities of enquiry around particular disciplinary topics. This deliberate design decision not to impose an ontology onto users makes it explicitly pragmatic, and discursive in nature (Buckingham Shum, 2006). In the context of learning sciences and CSCL research, Cohere is a Web 3.0 (social and semantic web) era application exemplifying the knowledge-building paradigm that has inspired earlier reflective, knowledge structuring software tools (Novak, 1998; Scardamalia and Bereiter, 1993; Suthers, Weiner, Connelly, & Paolucci, 1995).

Thus, Cohere renders proxies of learner understanding in the form of a semantic network of textual nodes that may also link to multimedia resources. Three recent technical advances are also of note: a version called the Evidence Hub now feeds analytics back to end users as well as to system administrators, and enables the embedding of multimedia within nodes (De Liddo and Buckingham Shum, 2013), while the textual content of nodes can now be analyzed for significant patterns using a rhetorical parser (Simsek, Buckingham Shum, Sándor, De Liddo & Ferguson, 2013). Together, these open the possibility for analytics to offer insight into epistemic beliefs, as summarized in Table 2, which returns to the broad epistemic "dimensions" introduced earlier (Table 1), and gives examples of proxies within socio-semantic trace data, plus guidance indicative of the sorts of challenges that might be posed to students to extend their epistemic cognition and probe their learning processes.[4]

We emphasize again that is not just the representational affordances of the analytics tool, but its *use* that will ultimately determine its impact. The representational scheme of semantically typed nodes and links clearly requires learner reflection and intentional artefact construction not normally demanded by conventional social platforms, and is thus theoretically tied to a socio-culturally inspired triad. However, both the constructs and the trace should be seen in their situated context: prior research into argumentation demonstrates that many elements in the cognitive and socio-cultural environment influence the usage patterns and impact of such tools (Scheuer, Loll, & McLaren, 2010). Key variables include user familiarity with the tool, their willingness to challenge each other in public, and the task set by teachers (this is a reiteration of the claims made at the end of Section 2). Thus, these should be dynamic tools, and empirical work will be needed to explore the relationship between feedback given, representations allowed, student responses to feedback, and the impact of this on learning.

Table 2. Trace Patterns and Guidance for Epistemic Beliefs

Dimensions of Epistemic Belief	**Candidate Socio/Semantic Trace Patterns**	**Guidance/Challenge to the Learner**
Certainty *The degree to which knowledge is conceived as stable or changing, ranging from absolute to tentative and evolving knowledge*	Frequency and density of positive and negative polarity links (e.g., *supports; builds on; is consistent with*, versus *challenges; refutes; is inconsistent with*). Presence of stability markers — (e.g., current sources; bibliographic references; geographical spread). Frequency and density of meta-discourse markers in the text of nodes, which moderate assertions in scholarly ways (e.g., *in contrast to; remains an unresolved issue*)	Are there no downsides to this claim? Can you find a counterexample? Is this idea consistent across time/place? Have you looked at XY map?

[4] Following previous work (De Liddo, Buckingham Shum, Quinto, Bachler, & Cannavacciuolo, 2011) the basic analytic statistic is constructed as a percentage representation of the target type, over the total types created by the user. For example, the number of "opinion" nodes created, as a percentage of the total number of nodes created by that user.

ISSN 1929-7750 (online).

(2014). Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space. *Journal of Learning Analytics*, *1*(2), 23-47.

Simplicity *The degree to which knowledge is conceived as compartmentalized or interrelated, ranging from knowledge as made up of discrete and simple facts to knowledge as complex and comprising interrelated concepts*	Standard network analytics describing the topography of the structure Semantic analysis of node and link types describing the variety of types used by a learner, the balance in link polarity (positive/neutral/negative), and the balance of propositional (*causes; refutes; solves*) versus analogical thinking (*is a metaphor of; is analogous to*). Semantic analysis of tags on nodes and links Encapsulation of node clusters or sub-networks within a broader category expressing a higher order construct The use of different argumentation schemes, which reflect different sources of authority (*argument from expertise; argument by analogy; argument from precedent*) In collaborative maps, the social network view showing who is connected to whom, but what kinds of links	Have you considered how X and Y might be connected? Why is this node so (dis)connected? Would you not expect this network to be better connected? Does it make sense that this paper is in the hub? Does this analogy hold in all cases? (And many other critical questions used to probe arguments: Buckingham Shum and Okada, 2008) Is it significant that these two analysts are not connecting their ideas to each other?
Source *The relationship between knower and known, ranging from the belief that knowledge resides outside the self and is transmitted, to the belief that it is constructed by the self*	Presence within the text of nodes of “I think” or other restatements of fact Challenges to claims through negative link types and meta-discourse markers in node text Few additional nodes made other than those created as quotations.	What do you think of these ideas? Doesn’t your perspective raise important questions for this claim? How does the evidence relate to your view?

ISSN 1929-7750 (online).

(2014). Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space. *Journal of Learning Analytics*, *1*(2), 23-47.

Justification *What makes a sufficient knowledge claim, ranging from the belief in observation or authority as sources, to the belief in the use of rules of inquiry and evaluation of expertise*	Judgements of relevance, and supporting or explanatory notes (“this evidences/explains x”); ties to method “ideas” Use of well known Argumentation Schemes, especially evidence and arguments grounded in the attribution of authority to people (“experts”) or publications based on their status (“valuing peer review”) Rhetorical constructions in node text that show a critical use of sources	What evidence do we have for this idea? Is it “good” evidence? Why/why not? You seem to quote/cite sources without much commentary of your own — bring in your voice more.

This example thus provides space to consider the role of the epistemology, assessment, and pedagogy in bounding learning analytics. In this case, assessment — or “measurement,” analytics, data gathering, etc. — is used to support learning, and the exploration of that learning. The technological support provides space for discursive, socio-culturally motivated pedagogy. Finally, the aim of the analytic device is not to make claims about a learner’s knowledge states at any particular moment, nor to provide psychometric assessment of groups of students, but rather to provide a discussion object, and to probe student *sense-making* processes, and *understanding* — their justifications for claims, a classic definition of knowledge — and thus an epistemological component of our analysis.

6.1 Many Lenses on Epistemic Beliefs

Table 2 thus proposes one set of traces from which meaningful data could be captured. This is not, however, to dismiss other approaches discussed in Section 4.1. The epistemological approach discussed throughout this work is instead intended to indicate that what drives our learning analytics — and assessment — is not *what they are*, but rather, what we *do* with them. Our suggestion is that many of these approaches to learning analytics — these dirty lenses on the world — provide insights into different levels of learning, and tools for meaning-making. For example, with this richer than normal data model in place, it is very simple, computationally, to feed back the number of ideas and connection types used, but this may provoke meaningful dialogue regarding what these other types might be used for, or why they have not thus far been used. Similarly, constructive discourse might occur around the reasons why one student’s map is more connected (but perhaps not appropriately so) than another’s.

There is a strong relationship between analytics, assessment, pedagogy, and epistemology (Figure 1); learning analytics should be mindful of this triad, which socio-cultural analytics bridges well. Our approach should be seen as one of “many lenses” for many contexts, used in combination with the more conventional forms of learning analytics currently dominating. In this last section before concluding, we outline how the approaches discussed in Section 4.1 relate to epistemic beliefs, and some strengths and limitations of these approaches:

Learning analytics based on Transactional approaches. Approaches that emphasize fixed, “correct” knowledge, over how those facts are used to display understanding, are likely to encourage lower epistemic cognition, and implicate more “realist” epistemologies that see knowledge as a reflection of “things” in the world.

(2014). Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space. *Journal of Learning Analytics*, *1*(2), 23-47.

Learning analytics based on Constructivist approaches. Similarly, there may be an overemphasis on a limited *range* of knowledge in constructivist approaches that emphasize development qua progression, but without considering the socio-cultural context in which that progression occurs, nor the wide range of uses for which it may be deployed. This may be particularly true in constrained systems that guide students through pre-set tasks and levels of attainment to meet, pre-specified software, and so on, as compared to those exploring knowledge co-constructed in discourse Understanding the ways that students build knowledge claims — understanding connections, justifications, change over time, and nuance — is fundamental to understanding their epistemic beliefs. Knowing that a student is at stage x of y in development may be less significant.

Learning analytics based on Apprenticeship approaches. In a similar vein, apprenticeship approaches can offer useful insight into group membership and the development of a student's thinking. However, the approach described in this paper suggests the best way to think about such approaches is with respect to the functional role that such community membership plays in a student's epistemic action, and their normative standards.

Learning analytics based on Subjectivist approaches. Learning analytics based on "affect" could be useful to the analysis of epistemic beliefs, with their analysis of "satisfaction" with information, e.g., enquiry-based learning (Ferguson et al., 2011); self-efficacy in IR information seeking (Tsai & Tsai, 2003); satisfaction with search results (Huffman & Hochster, 2007). As such, affective analytics might be used to explore whether learners are prematurely satisfied with findings that a peer or educator deems to be inadequate, or if they have an appropriate sense of disquiet or frustration with a flawed argument or methodology.

7. CONCLUSION

This paper opened with the argument that the triad of epistemology, assessment, and pedagogy are fundamentally entwined. Furthermore, we suggested that a focus on high-stakes assessment — which learning analytics may well be used to perpetuate — is detrimental to the wider enterprise of education, prioritizing the reliability of tightly defined assessments over continuing, formative assessment for learning, and authentically situated learning that is harder to fit into formal examination contexts. This is problematic as it limits the ways we can challenge students in assessments, and fails to reflect their encounters with knowledge claims in the world beyond the classroom walls. Learning Analytics should be mindful of the epistemological, assessment, and pedagogic implications of their application, and should be cautious of falling into technological determinism.

We have highlighted that transactional approaches may emphasize the use of facts; constructivist the broad (and contextual) application of skills; subjectivist the self-efficacy and motivators of students; apprenticeship the dynamic practical-based learning that may occur through high-level membership of communities of practice; connectivism the ability of students to build up, link, and curate their knowledge "networks." A socio-cultural, pragmatic approach may offer an additional toolset, alongside a theoretical frame through which to use other learning analytics lenses. All are partial (in bias, and hence in their coverage of all that might be measured), but may be used in complementary ways.

Analytics from user traces provide a means to track and record previously ephemeral process data, which could benefit assessment *for* learning in significant new ways. Pragmatist approaches, which emphasize use and meaning-making over the accrediting of true statements, may have an important role here. The grasp of curriculum facts and methods remains critical but the emphasis shifts to their effective, contextualized use, in argument structures, in discussion, in problem-solving. A focus on the socio-cultural learning system draws attention to how analytics take into account the centrality of discourse for sense-making, and in constituting "context."

We have gone beyond "our learning analytics are our pedagogy" (Buckingham Shum, 2012), arguing in more detail that analytics also embody related epistemological assumptions, thus perpetuating related assessment regimes. Since learning analytics unavoidably embody these triadic assumptions, there is the risk of perpetuating the limitations of current educational practices, especially those associated with high stakes testing.

In conclusion, the stakes are high. As du Gay and Pryke (2002, pp. 12–13) observe, "accounting tools ... do not simply aid the measurement of economic activity, they shape the reality they measure." As educational institutions adopt platforms and applications that make it increasingly easy to capture particular kinds of data, conduct certain analyses, and foreground particular patterns for human attention, they reinforce their location within the triad. Educators, possibly held to account with analytics-based performance indicators, may well come to design learning in a way that will enable them to evidence impact with these new tools. Students, aware that their behaviour is being tracked, may similarly seek to "evidence" their progress, and game the more simplistic measures. Technology alone does not determine practice; moreover, as with any tool, it is not only the design of the tool, but the way in which it is wielded in context, that defines its value.

ACKNOWLEDGEMENTS

We are grateful to Cindy Kerawalla and three anonymous reviewers for helpful comments on an earlier version of this paper presented at the LAK13 conference, and to Dan Suthers and the LAK13 attendees who offered helpful feedback, which was used in developing this extended version.

REFERENCES

Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online sources. *Cognition and Instruction*, *30*(1), 39–85. doi:10.1080/07370008.2011.636495

Black, P., & Wiliam, D. (2001). Inside the black box. BERA. Retrieved from http://www.collegenet.co.uk/admin/download/inside%20the%20black%20box_23_doc.pdf

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31.

Bromme, R., Pieschl, S., & Stahl, E. (2009). Epistemological beliefs are standards for adaptive learning: A functional theory about epistemological beliefs and metacognition. *Metacognition and Learning*, *5*(1), 7–26. doi:10.1007/s11409-009-9053-5

Buckingham Shum, S. (2006). Sensemaking on the pragmatic web: A hypermedia discourse perspective. *1st International Conference on the Pragmatic Web.* 21–22 Sept 2006, Stuttgart. Open Access Eprint: http://oro.open.ac.uk/6442

Buckingham Shum, S. (2008). Cohere: Towards web 2.0 argumentation. *Proceedings of the 2nd International Conference on Computational Models of Argument.* 28–30 May 2008, Toulouse, France. IOS Press, 97–108. http://oro.open.ac.uk/10421

Buckingham Shum, S. (2012). Our learning analytics are our pedagogy. Keynote Address, presented at the *Expanding Horizons 2012 Conference*, Macquarie University. Retrieved from http://www.slideshare.net/sbs/our-learning-analytics-are-our-pedagogy

Buckingham Shum, S., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. *Proceedings of the 2nd International Conference on Learning Analytics & Knowledge.* 29 Apr–2 May, Vancouver, BC. New York: ACM, 92–101. Eprint: http://oro.open.ac.uk/32823

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society*, *15*(3), 3–26.

Buckingham Shum, S., & Okada, A. (2008). Knowledge cartography for controversies: The Iraq debate. In A. Okada, S. Buckingham Shum, & T. Sherbone (Eds.), *Knowledge cartography: Software tools and mapping techniques.* London: Springer.

Crook, C., & Lewthwaite, S. (2010). Technologies for formal and informal learning. In K. Littleton, C. Wood, & J. K. Staarman (Eds.), *International Handbook of Psychology in Education* (pp. 435–461). Bingley, UK: Emerald Group Publishing.

Cunnane, S. (2011, May 12). The Danish gambit: Online access, even during exams. *Times Higher Education*. Retrieved from http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=416090&c=1

Davis, A. (1998). Understanding and holism. *Journal of Philosophy of Education*, *32*(1), 41–55.

Davis, A. (1999). The Limits of Educational Assessment. Hoboken, NJ: Wiley-Blackwell.

Davis, A. (2005). Learning and the social nature of mental powers. *Educational Philosophy and Theory*, *37*(5), 635–647. doi:10.1111/j.1469-5812.2005.00148.x

Davis, A. (2006). High stakes testing and the structure of the mind: A reply to Randall Curren. *Journal of Philosophy of Education*, *40*(1), 1–16. doi:10.1111/j.1467-9752.2006.00492.x

Davis, A., & Williams, K. (2002). Epistemology and curriculum. In N. Blake, P. Smeyers, & R. Smith (Eds.), *The Blackwell guide to the philosophy of education*. Blackwell Reference Online.

Dawson, S. (2010). “Seeing” the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology, 41*, 736–752.

Deakin Crick, R., Broadfoot, P., & Claxton, G. (2004). Developing an effective lifelong learning inventory: The ELLI project. *Assessment in Education, 11*, 248–272.

De Liddo, A., & Buckingham Shum, S. (2013). The evidence hub: Harnessing the collective intelligence of communities to build evidence-based knowledge. *Large Scale Ideation and Deliberation Workshop, 6th International Conference on Communities & Technologies.* 29 June 2013, Munich. Open Access Eprint: http://oro.open.ac.uk/38002

De Liddo, A., Buckingham Shum, S., Quinto, I., Bachler, M., & Cannavacciuolo, L. (2011). Discourse-centric learning analytics. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 23–33). New York: ACM. doi:10.1145/2090116.2090120

Dweck, C. (2006). *Mindset: The new psychology of success*: New York: Random House.

DeBacker, T. K., Crowson, H. M., Beesley, A. D., Thoma, S. J., & Hestevold, N. L. (2008). The challenge of measuring epistemic beliefs: An analysis of three self-report instruments. *The Journal of Experimental Education*, *76*(3), 281–312. doi:10.3200/JEXE.76.3.281–314

Dede, C. (2008). A seismic shift in epistemology. *EDUCAUSE Review*, *43*(3), 80–81.

Dewey, J. (1998). *Experience and education*. Kappa Delta Pi. Retrieved from http://ruby.fgcu.edu/courses/ndemers/colloquium/ExperiencEducationDewey.pdf

du Gay, P., & Pryke, M. (2002). Cultural economy: Cultural analysis and commercial life. London: Sage.

Edwards, A. D., & Furlong, V. J. (1978). *The language of teaching: Meaning in classroom interaction*. London: Heinemann. Retrieved from http://www.getcited.org/pub/102019252

Edwards, D., & Mercer, N. (1987). Common knowledge: The development of understanding in the classroom. London: Routledge.

Ferguson, L. E., Bråten, I., & Strømsø, H. I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence: A think-aloud study. *Learning and Instruction*, *22*(2), 103–120. doi:10.1016/j.learninstruc.2011.08.002

Ferguson, R., & Buckingham Shum, S. (2012). Social learning analytics: Five approaches. *Proceedings of the 2nd International Conference on Learning Analytics & Knowledge*, Vancouver, BC. New York: ACM Press. Retrieved from http://projects.kmi.open.ac.uk/hyperdiscourse/docs/learning analyticsK2012-RF-SBS.pdf

Ferguson, R., Buckingham Shum, S., & Deakin Crick, R. (2011). EnquiryBlogger: Using widgets to support awareness and reflection in a PLE setting. *1st Workshop on Awareness and Reflection in Personal Learning Environments*, Southampton, UK. Retrieved from http://oro.open.ac.uk/19576/1/Producing_Animations_to_educate_MSM.pdf

Foster, J. (2009). Understanding interaction in information seeking and use as a discourse: A dialogic approach. *Journal of Documentation*, *65*(1), 83–105. doi:10.1108/00220410910926130

Gardner, J. (2011). *Assessment and learning* (2nd ed.). London: SAGE.

Greene, J. A., Muis, K. R., & Pieschl, S. (2010). The role of epistemic beliefs in students' self-regulated learning with computer-based learning environments: Conceptual and methodological issues. *Educational Psychologist*, *45*(4), 245–257. doi:10.1080/00461520.2010.515932

Hamilton, E., & Feenberg, A. (2005). The technical codes of online education. *Techné: Research in Philosophy and Technology*, *9*(1). Retrieved from http://scholar.lib.vt.edu/ejournals/SPT/v9n1/hamilton.html

Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *Journal of the Learning Sciences*, *12*(1), 53–90. doi:10.1207/S15327809JLS1201_3

Harlen, W. (2007). *Assessment of learning*. London: SAGE.

Hertzum, M. (2008). Collaborative information seeking: The combined activity of information seeking and collaborative grounding. *Information Processing & Management*, *44*(2), 957–962. doi:10.1016/j.ipm.2007.03.007

Haythornthwaite, C., and de Laat, M. (2010). Social networks and learning networks: Using social network perspectives to understand social learning. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, & T. Ryberg (Eds.), *7th International Conference on Networked Learning.* Aalborg, Denmark.

(2014). Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space. *Journal of Learning Analytics*, *1*(2), 23-47.

Hofer, B. K. (2001). Personal epistemology research: Implications for learning and teaching. *Educational Psychology Review*, *13*(4), 353–383.

Hopmann, S. T., Brinek, G., & Retzl, M. (Eds.). (2007). *PISA according to PISA: Does PISA keep what it promises?* Vienna, Austria: Wien Lit-Verlag. Retrieved from http://www.univie.ac.at/pisaaccordingtopisa/

Huffman, S. B., & Hochster, M. (2007). How well does result relevance predict session satisfaction? *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (p. 574). doi:10.1145/1277741.1277839

Katz, S. (2000). Competency, epistemology and pedagogy: Curriculum's holy trinity. *Curriculum Journal*, *11*(2), 133–144. doi:10.1080/09585170050045164

Kelly, G. J., Luke, A., & Green, J. (2008). What counts as knowledge in educational settings: Disciplinary knowledge, assessment, and curriculum. *Review of Research in Education*, *32*(1), vii–x. doi:10.3102/0091732X07311063

King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist*, *39*(1), 5–18. doi:10.1207/s15326985ep3901_2

Kuhn, D., & Weinstock, M. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer & P. Pintrich R. (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 121–144). Mahwah, NJ: Lawrence Erlbaum Associates.

Lazonder, A. W. (2005). Do two heads search better than one? Effects of student collaboration on web search behaviour and search outcomes. *British Journal of Educational Technology*, *36*(3), 465–475. doi:10.1111/j.1467-8535.2005.00478.x

Lin, C., & Tsai, C. (2008). Exploring the structural relationships between high school students' scientific epistemological views and their utilization of information commitments toward online science information. *International Journal of Science Education*, *30*(15), 2001–2022. doi:10.1080/09500690701613733

Maloney, J., Resnick, M., Rusk, N., Silverman, B., & Eastmond, E. (2010). The scratch programming language and environment. *ACM Transactions on Computing Education, 10*: Article No. 16.

Mason, L., Ariasi, N., & Boldrin, A. (2011). Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction*, *21*(1), 137–151. doi:10.1016/j.learninstruc.2010.01.001

Mason, L., Boldrin, A., & Ariasi, N. (2009). Epistemic metacognition in context: Evaluating and learning online information. *Metacognition and Learning*, *5*(1), 67–90. doi:10.1007/s11409-009-9048-2

Mercer, N., & Littleton, K. (2007). Dialogue and the development of children's thinking: A sociocultural approach (New edition.). London: Routledge.

Novak, J. D. (1998). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations.* Mahwah, NJ: Lawrence Erlbaum Associates.

Oliver, M. (2011). Technological determinism in educational technology research: Some alternative ways of thinking about the relationship between learning and technology. *Journal of Computer Assisted Learning*, *27*(5), 373–384. doi:10.1111/j.1365-2729.2011.00406.x

Olson, D. R., & Bruner, J. S. (1996). Folk psychology and folk pedagogy. *The handbook of education and human development*.. In D. R. Olson and N. Torrance (Eds). Blackwell Publishing, 1998. Blackwell Reference Online. 03 June 2014 <http://www.blackwellreference.com/subscriber/tocnode.html?id=g9780631211860_chunk_g97806312118604>

Österholm, M. (2009). Theories of epistemological beliefs and communication: A unifying attempt. *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 275–264). Retrieved from http://umu.diva-portal.org/smash/record.jsf?pid=diva2:228322

Österholm, M. (2010). Relationships between epistemological beliefs and properties of discourse: Some empirical explorations. In C. Bergsten, E. Jablonka, & T. Wedege (Eds.), *Madif 7, the 7th Swedish Mathematics Education Research Seminar*, Stockholm, Sweden (pp. 241–250). Linköping, Sweden: SMDF. Retrieved from http://osterholm.pcriot.com/publ-madif7a.shtml

Potter, J., & Edwards, D. (2003). Sociolinguistics, cognitivism and discursive psychology. *IJES, International Journal of English Studies*, *3*(1), 93–110.

Säljö, R. (1999). Learning as the use of tools: A sociocultural perspective on the human-technology link. In K. Littleton & P. Light (Eds.), *Learning with computers: Analysing productive interaction*. Oxford, UK: Psychology Press.

Säljö, R. (2012). Literacy, digital literacy and epistemic practices: The co-evolution of hybrid minds and external memory systems. *Nordic Journal of Digital Literacy*, *1*, 5–19.

Salomon, G. (1996). Distributed cognitions: Psychological and educational considerations. Cambridge, UK: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (1993). Technologies for knowledge-building discourse. *Communications of the ACM, 36*, 37–41.

Scherr, R. E., & Hammer, D. (2009). Student behavior and epistemological framing: Examples from collaborative active-learning activities in physics. *Cognition and Instruction*, *27*(2), 147–174. doi:10.1080/07370000902797379

Scheuer, O., Loll, F. N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Argumentation, 5*, 43–102.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, *82*(3), 498–504. doi:10.1037/0022-0663.82.3.498

Schraw, G. (2013). Conceptual integration and measurement of epistemological and ontological beliefs in educational research. *ISRN Education*. Retrieved from http://www.hindawi.com/isrn/education/aip/327680/

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Franke, K., Rupp, A. A., & Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st century assessment of learning. *The International Journal of Learning and Media, 1*, 33–53.

Siegel, H. (1998). Knowledge, truth and education. In D. Carr (Ed.), *Education, knowledge, and truth: Beyond the postmodern impasse* (pp. 19–36). London, UK: Routledge.

Siemens, G. (2006). *Knowing Knowledge*. Lulu.com. See *www.elearnspace.org/KnowingKnowledge_LowRes.pdf*

Simsek, D., Buckingham Shum, S., Sándor, Á., De Liddo, A., & Ferguson, R. (2013). XIP dashboard: Visual analytics from automated rhetorical parsing of scientific metadiscourse. *1st International Workshop on Discourse-Centric Learning Analytics, at 3rd International Conference on Learning Analytics & Knowledge.* Apr. 8–12, 2013, Leuven, Belgium. Open Access Eprint: http://oro.open.ac.uk/37391

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning*, *2*(2), 191–210. doi:10.1007/s11412-007-9015-3

Suthers, D. D., & Verbert, K. (2013). Learning analytics as a “middle space.” *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 1–4). New York: ACM. doi:10.1145/2460296.2460298

Suthers, D. D., Weiner, A., Connelly, J., & Paolucci, M. (1995). Belvedere: Engaging students in critical discussion of science and public policy issues. In Greer J (Ed.), *Proceedings of AI-ED 95: World Conference on Artificial Intelligence in Education.* Washington, D.C.: AACE, 266–273.

Tsai, C. (2004). Beyond cognitive and metacognitive tools: The use of the Internet as an “epistemological” tool for instruction. *British Journal of Educational Technology*, *35*(5), 525–536. doi:10.1111/j.0007-1013.2004.00411.x

Tsai, M.-J., & Tsai, C.-C. (2003). Information searching strategies in web-based science learning: The role of Internet self-efficacy. *Innovations in Education & Teaching International*, *40*(1), 43–50. doi:10.1080/1355800032000038822

UAGU (2010). Undervisningsministerie (Ministry of Education), & Afdelingen for Gymnasiale Uddannelser (Department of Secondary Education). (2010, September). Endelig rapport fra følgegruppen for forsøget med digitale prøver med adgang til internettet i udvalgte fag på stx og hhx (Final report of the Monitoring Group experiment with digital samples with access to Internet in selected subjects at stx and HHX). Undervisningsministerie (Ministry of Education). Retrieved from http://www.uvm.dk/~/media/Files/Udd/Gym/PDF10/101007_it_rapport_13%20september%20 2010.ashx

Williams, K. (1998). Assessment and the challenge of scepticism. In D. Carr (Ed.), *Education, knowledge, and truth: Beyond the postmodern impasse*. London, UK: Routledge.

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

Learning Analytics for Online Discussions: Embedded and Extracted Approaches

Alyssa Friend Wise, Yuting Zhao, Simone Nicole Hausknecht
Simon Fraser University, Canada
alyssa_wise@sfu.ca

Abstract: This paper describes an application of learning analytics that builds on an existing research program investigating how students contribute and attend to the messages of others in asynchronous online discussions. We first overview the E-Listening research program and then explain how this work was translated into analytics that students and instructors could use to reflect on their discussion participation. Two kinds of analytics were designed: some e*mbedded* in the learning environment to provide students with real-time information on their activity in-progress; and some *extracted* from the learning environment and presented to students in a separate digital space for reflection. In addition, we describe the design of an intervention though which use of the analytics can be introduced as an integral course activity. Findings from an initial implementation of the application indicated that the learning analytics intervention supported changes in students' discussion participation. Five issues for future work on learning analytics in online discussions are presented. One, unintentional versus purposeful change; two, differing changes prompted by the same analytic; three, importance of theoretical buy-in and calculation transparency for perceived analytic value; four, affective components of students' reactions; and five, support for students in the process of enacting analytics-driven changes.

Keywords: Online learning, computer mediated communication, learning analytics, asynchronous discussion groups, student participation

1. INTRODUCTION

The field of learning analytics is concerned with the collection and analysis of data traces related to learning in order to inform and improve learning processes and/or their outcomes (Siemens et al., 2011). Within this space, a distinction can be made between classes of analytics based on the types of data collected and the kinds of decision making targeted (Ferguson, 2012). At a macro level, administrators and policy makers have the opportunity to use learning analytics to make programmatic or legislative decisions. In such situations, data on past learning events is used to make decisions about future ones; these choices tend to be based on relatively long data time-cycles (Clow, 2012), affect large numbers of people, and involve outcome-type data; for example, summative assessments, performance indicators, and the like (Buckingham Shum & Ferguson, 2012). For these purposes, data may also be aggregated at various levels (i.e., student, class, department, institution, et cetera).

In contrast, at a micro level, learners and teachers have the opportunity to use learning analytics to make more local decisions about the current learning events in which they are involved (Clow, 2012). In this case, the relevant data relates to tracking learning processes and an important element of interpretation is having a model of learning for the particular environment — i.e., a research-based framework for understanding what productive activity in the specific learning context looks like. In some cases, the model may be specified such that analytics data is processed and interpreted according to some system of rules leading to automatic changes in the learning system (e.g., Roll, Aleven, &

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

Koedinger, 2010). In other cases, data may be processed into categories according to the model, but then presented to stakeholders (students, teachers, administrators, et cetera) to support decision-making (e.g., Jovanovic et al., 2008).

This paper reports on a learning analytics application of the latter type designed to support decision-making by students and teachers involved in asynchronous online discussions as a learning activity. The work builds on an existing research program investigating how students contribute and attend to the messages of others in online discussions. Both contributing and attending to discussion messages are important activities in realizing the theoretical potential of online discussions to support knowledge construction (Wise, Speer, Marbouti, & Hsiao, 2013). However a substantial research base shows that in actual asynchronous discussions learners often pay limited attention to others' posts (Hewitt, 2003; Thomas, 2002), resulting in conversation patterns that can be characterized as incoherent rather than dialogic (Herring, 1999; Webb, Jones, Barker, & van Schaik, 2004). Early research into these problems reported disturbingly low overall statistics, suggesting that the problems were global and, to some extent, systemic products of asynchronous discussion environments (Hewitt, 2003; Kear, 2001; Swan, 2003). These findings spurred efforts to improve online discussion tools to support productive engagement and discussion (e.g., Scardamalia, 2004; Marbouti, 2012). However, more recent work disaggregating data across individuals has revealed that students in fact engage in very different kinds of behaviours when they participate in asynchronous online discussions (Wise, Perera, Hsiao, Speer, & Marbouti, 2012; Wise, Speer et al., 2013; Wise, Hsiao, Marbouti, Speer, & Perera, 2012). This suggests that discussion participation can also be improved with more targeted efforts to provide guidance to students individually using learning analytics.

The paper describes how we attempted to take up this opportunity by using our pre-existing research program to inform the development of a learning analytics application for online discussions. We begin by laying out our theoretical framework and empirical findings about "listening" (attending to others' posts) in online discussions. We then explain how this work was translated into guidelines for practice and analytic metrics with which students and instructors could assess and reflect on their discussion participation. We describe two classes of analytics for online discussions that we developed — embedded and extracted analytics — as well as the design of the larger intervention framing their use. The latter half of the paper then describes and presents findings from an initial implementation of the learning analytics application and its implications for the future design of analytics interventions for asynchronous online discussions.

2. FROM RESEARCH TO ANALYTICS

The increasing amount of information automatically captured by online learning environments is attractive as a source of data with which to better understand and support learning processes. However, there is a wide gap between the kinds of data easily capturable in learning environments such as online discussions and the kinds of constructs established as pedagogically valuable (Buckingham Shum & Ferguson, 2012). To address this problem, it is vital to build bridges from both directions: for learning analytics researchers to draw on existing educational research and theory as they develop their applications; and for educational researchers to leverage data mining and information visualization work to explore how they might develop useful learning analytics based on their research programs. Here we describe our experience of the latter process with the hope that it can inform other educational researchers moving into the "middle space" of learning analytics.

2.1 Theoretical Framework and Research Base

In this work, online discussions are conceptualized from a social constructivist perspective as a venue in which learners can interact to build both collective and individual understanding through dialogue (Kamel Boulos & Wheeler, 2007; Stahl, 2005). Scholars differ in their theoretical accounts of the mechanisms underlying the learning process, referring variously to the importance of being exposed to multiple viewpoints, articulating one's own ideas, experiencing socio-cognitive conflict, and the negotiation and internalization of group understandings (Lipponen, 2002; Stahl, 2005; Wise, Speer et al., 2013). However, at a fundamental level all explanations depend on two basic processes in which learners must engage: "speaking" (externalizing one's ideas by contributing posts to the discussion); and "listening" (taking in the externalizations of others by accessing existing posts) (Wise, Speer et al., 2013). Speaking in online discussions is clearly visible to others; this may explain why the bulk of research on and guidance for participation in online discussions is focused on posting activity (Hew, Cheung, & Ng, 2010). However, while listening is largely invisible, it is also critical for discussions that build understanding in the ways described above (Wise, Speer et al., 2013). The E-Listening Project is a four-year research effort funded by the Social Sciences and Humanities Research Council of Canada to understand better how students attend to others' contributions in online discussion. Work on this project has documented many of the different kinds of listening behaviours in which students engage (Wise, Perera et al., 2012; Wise, Speer et al., 2013; Wise, Hsiao et al., 2012) and has shown empirical connections between students' listening and speaking behaviours (Wise, Hausknecht, & Zhao, 2013; Wise, Hausknecht, & Zhao, 2014).

While the language of speaking and listening draws on a metaphor based in face-to-face conversations, online discussions offer different affordances and constraints for these activities (Wise, Hausknecht, & Zhao, 2014; Wise, Perera et al., 2012). Specifically, in asynchronous online discussions, learners have greater control over the timeline and pace of their engagement (Jonassen & Kwon 2001). This creates opportunities for thoughtful listening and reflective speaking (Lipponen, 2002), but also additional challenges of time management, especially for prolific discussions (Peters & Hewitt, 2010). For this reason, helping learners to actively monitor and regulate how they speak and listen in online discussions is an important tool for supporting productive engagement in discussions.

Given the above-described goal of using dialogue to build individual and collective understandings and the existing research base on asynchronous online discussions, particular speaking and listening behaviours can be characterized as more or less productive. Here we summarize the different dimensions of listening and speaking identified by the E-Listening project, relating them to the online discussion literature more broadly. First, in terms of *speaking quantity*, multiple posts are needed to respond to others' ideas and questions, elaborate on the points made, and negotiate with others (Pena-Shaff & Nicholls, 2004). These posts should be distributed throughout the discussion (rather than concentrated at the start or end), relating to the *temporal distribution* of participation. In addition posts of moderate length best support rich dialogue since very short posts tend to be shallow in their presentation of ideas (Dennen, 2001) but long posts are often perceived as overwhelming and are thus not read (Peters & Hewitt, 2010). Precise specifications of moderate length may differ by context and age-level; in higher education, this is often specified at around 100 to 200 words. In terms of *speaking quality,* posts that are clear, critical, and connected to the existing conversation support the generation of insight and understanding of the topic (Rovai, 2007). Posts whose arguments are based on evidence and/or theory can also trigger others to build on or contest the points productively (Clark, Sampson, Weinberger, & Erkens, 2007), and responses that clarify points, elaborate or question existing ideas, or synthesize different ideas together help deepen the exploration of ideas and move the discussion forward (Pena-Shaff & Nicholls, 2004). In terms of listening activity, there are also multiple dimensions requiring attention. Considering *breadth of listening*, viewing a greater proportion of others' posts exposes students to a greater diversity of ideas (Wise, Hsiao et al., 2012). *Depth of listening* is important as an indication of the degree to which learners consider the ideas of others (Wise, Hsiao et al., 2012); greater listening depth is predictive of richer argumentation in posts made (Wise, Hausknecht, & Zhao, 2014). *Listening reflectivity* (revisiting one's own and others' posts from earlier in the discussion) can provide context for interpreting recent posts and examine how thinking has changed. Revisiting others' posts has been shown to be predictive of more substantive responses to others' ideas and possibly increased reflection on how one's own ideas have changed (Wise, Hausknecht, & Zhao, 2014). Finally *temporal distribution* of listening is important since engaging in multiple sessions during a discussion and integrating listening and posting in the same session can support learners in making connections between ideas and contributing posts that productively build on previous ones (Wise, Speer et al., 2013).

2.2 Issues in Translating a Research Program into Learning Analytics

While established learning research programs have much to offer the development of learning analytics, the process of translating research methodologies and measures into useful learning analytics metrics is far from trivial. Research programs work to identify traces of learning that are meaningful to researchers; however, such measures are often quite complex and/or theoretically grounded. They thus may not be intuitively meaningful (or meaningful in the same way) to students, instructors, and administrators. For this reason educational researchers need to go through a process of selection (and possibly modification) of metrics to make them understandable and useful to the target audience. There may also be opportunities to support the use of more complex analytics through the intervention design: how the analytics are introduced into the learning environment and how the activity of interpreting and making decisions based on them is framed. This raises another important consideration: while research programs generally use measures to understand what has happened in a learning event with eventual consequences for practice, in an analytics context stakeholders use metrics to make immediate decisions about their behaviour. Thus, researchers must carefully consider the anticipated effects of providing certain metrics to learners and teachers at particular points in the learning process. Finally, in choosing measures, it is also important to consider the accelerated timeline of data collection, processing, and interpretation required for learning analytics compared to research analyses.

Once measures have been decided on, additional work is required to figure out how to present the traces of learning to students and teachers in a form that will be useful to them and frame the process of interpretation and decision-making as an integrated aspect of the learning activity tied to goals and expectations. While not specific to the case of translating a learning research program into analytics, these aspects of analytics design are as important as the selection of the measures. In the following sections, we describe all three aspects of our learning analytics application: the traces of learners' activity in online discussions we chose to capture, how we presented these traces to learners, and how we framed the inclusion of analytics as part of discussion activity to guide their use in productive decision-making by learners and teachers.

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

2.3 Two Kinds of Approaches to Learning Analytics

Below we describe two different classes of analytics for learners that we developed for online discussions: *embedded* analytics and *extracted* analytics. Embedded analytics are traces of activity integrated into the learning environment itself that can be used by learners in real-time to guide their participation. In this case, interpretation of the analytics and participation in the learning activity are unified as a single endeavor. An advantage to embedded analytics is that they can be used seamlessly to support metacognitive monitoring during participation in the learning activity itself. However, a weakness of being embedded is the possibility of being ignored; i.e., there is no reason to assume that students will naturally use the analytic affordances of the tool to support their participation individually or as a group. For this reason, such use needs to be specifically encouraged by structuring it in to the learning activity parameters.

Extracted analytics are traces of activity extracted from learning environment and presented back to learners for interpretation as a separate exercise from participating in the learning activity itself. That is, while the presentation of the analytics may be integrated into the overall learning environment, the activity of interpreting them is separate from that of engaging in the learning activity. For example, analytics presented to learners via most dashboard systems fall into this second category.[1] The distinction between embedded and extracted analytics will be further clarified through the description of the specific instantiation of each that follows in the subsequent sections.

2.3.1 Embedded Analytics

In this work, we have chosen to use a specific asynchronous discussion forum because of its inherent affordances for providing embedded analytics. The *Starburst* discussion forum (Marbouti, 2012) was developed to present discussion threads as a hyperbolic (radial) tree structure, allowing students to see the structure of the discussion and the location of their comments within it (see Figure 1). Posts are represented as scaled, coloured spheres connected by lines indicating their reply relations. When a post is selected, it enlarges to the maximum size and moves to the centre of the diagram while the other posts are rearranged around it; sphere size decreases with distance from the currently selected post. For each student, new posts are shown in red and previously viewed posts are shown in blue. The initial (seed) post always remains yellow. The design rationale and general benefits over traditional linear text-based forums have been described previously (Marbouti, 2012) and include more purposeful reading and replying behaviours by students, the ability to focus on a part of the discussion in the context of the whole, and increased reading of posts in a connected fashion. Easy navigation between connected posts also provides context as needed for each comment, thus the reply tool in *Starburst* omits the commonly provided auto-quote function to encourage students who use direct quotes from previous posts to do so purposefully.

In addition to these general benefits for interaction, *Starburst* also provides an embedded visual analytic of listening and speaking behaviours in terms of how the group's discussion is proceeding and how each individual is attending and contributing to it. At the group level, the graphical representation of the tree's branches allows for easy inspection of the structure of the discussion. Learners can see how many

[1] While the metaphoric notion of a dashboard suggests that one can pay attention to both the activity (driving) and instrumentation panel (dashboard) simultaneously, or at least in very quick succession, in practice learning analytics dashboards are generally presented as dedicated screens, apart from the main learning activity interface.

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

different threads have been created thus far and how deep each is, using this information to inform their decisions about where to read and contribute (Marbouti, 2012). They can also examine which threads and posts are receiving the most attention (responses) and if any are being neglected. For example, in Figure 1 the post labelled "Transfer"…? has not yet been taken up by the group. This kind of analytic can be useful in addressing the problem of inadvertent thread abandonment or death (Hewitt, 2003).

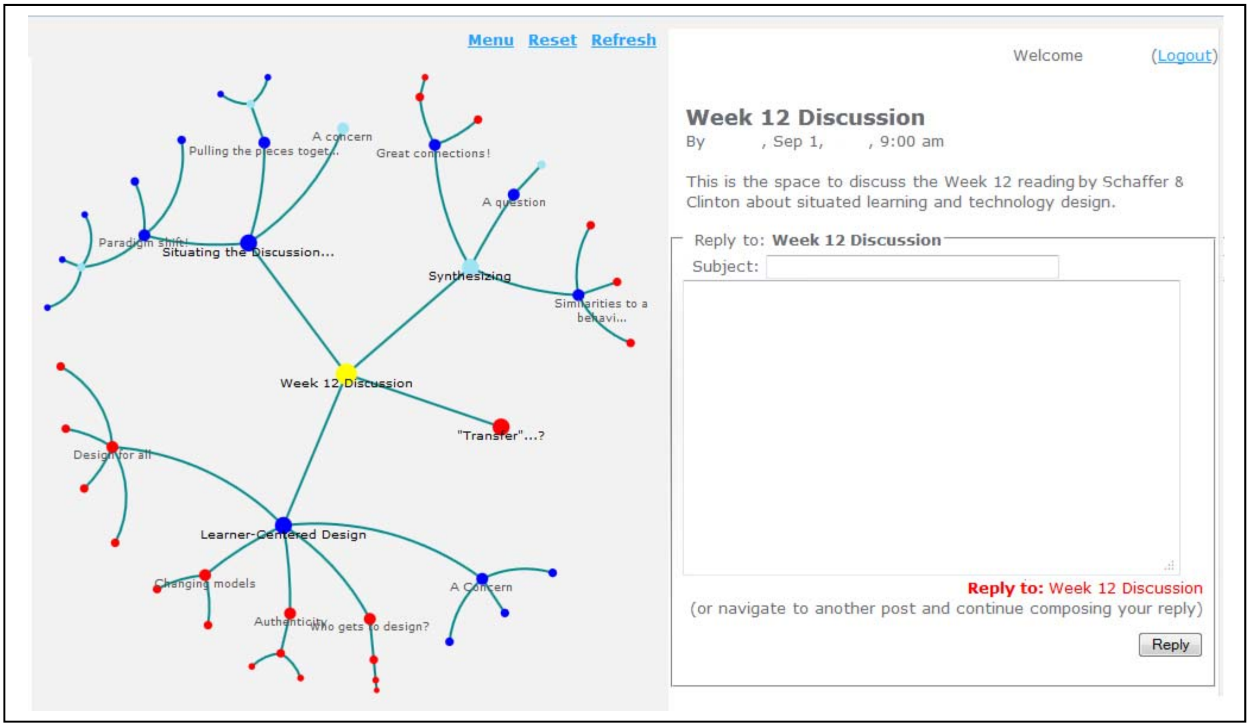

Figure 1. The Starburst discussion tool adapted for analytics

At the individual level, the red/blue colouring of the posts helps each student easily track which parts of the discussion they have attended to already and which parts they have not. For example, in Figure 1 the student has been heavily attending to one thread (top left), moderately attending to a second (top right), and very minimally attending to a third (bottom). In addition, for the current application, we made an adaptation from the previous version of the forum to colour the active learner's posts differently (light blue). In this way, we provide an analytic to each student of how he or she has contributed to the discussion thus far in terms of quantity (high or low volume compared to overall quantity of discussion), breadth (distribution throughout the discussion), and intensity (multiple contributions to a specific thread). Students can also easily see to which of their posts others have replied and to which they have not. In the example shown in Figure 1, the learner has made five posts in two of the three active threads. Several posts have stimulated further discussion, in particular the one labelled "Synthesizing." Here the learner can also see his or her round-trip interaction (Anderson, 2008; Henri, 1992) with another student responding with "A question." The learner can also see that no one has yet addressed the post entitled "A concern."

2.3.2 Extracted Analytics

In contrast to the embedded analytics described above, other useful information about student online discussion activity does not easily lend itself to presentation through the graphical interface (e.g., temporal distribution of participation). Thus, in our work with extracted analytics, we sought to make log-file trace data of speaking and listening activity visible to learners. The metrics selected were developed based on our prior research investigating how students attend to the messages of others in online discussions described earlier (Wise, Hausknecht, & Zhao, 2014; Wise, Perera et al., 2012; Wise, Speer et al., 2013; Wise, Hsiao et al., 2012) and are summarized in Table 1.

The metrics chosen for the extracted analytics were a subset of indices from the most important dimensions/variables identified in the research program that provided a useful signal with a minimum of noise (leading to an emphasis on count rather than time-based measures; see Wise, Hausknecht, & Zhao, 2014). We also specifically chose metrics thought to be readily understandable by students (with brief explanation), and that were complementary to rather than duplicative of the embedded analytics. There were, however, two intentional overlaps. First, the metric *Number of Posts Made* is viewable through the embedded analytics; however, the total number of posts made is less salient than their distribution, thus providing this sum and the average for the group (not easily determinable from the interface) was deemed useful additional information. Second, the metric *Percent of Posts Read* is similar to the embedded red/blue colour tracking of posts viewed in the interface; however, while the interface tracks all posts *viewed*, this metric is only based on posts actually *read* (not scanned) and thus provides complementary (and often quite different) information.

Table 1. Summary of discussion participation metrics

Metric	Definition	Participation Criteria
Range of Participation	Span of days a student logged in to the discussion.	*Temporal distribution*
Number of Sessions	Number of times a student logged in to the discussion.	*Temporal distribution*
Percent of Sessions with Posts	Number of sessions in which a student made a post, divided by his/her total number of sessions.	*Temporal distribution*
Average Session Length	Total length of time spent in the discussion divided by his/her number of sessions.	*Temporal distribution*
Number of Posts Made	Total number of posts a student contributed to the discussion.	*Speaking quantity*
Average Post Length	Total number of words posted by a student divided by the number of posts he/she made (including manually inserted direct quotes).	*Speaking quantity*
Percent of Posts Read	Number of unique posts that a student read divided by the total number of posts made by others to the discussion.	*Listening breadth*
Number of Reviews of Own Posts	Number of times a student reread posts that he/she had made previously.	*Listening reflectivity*
Number of Reviews of Others' Posts	Number of times a student reread others' posts that he/she had viewed previously.	*Listening reflectivity*

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

Data processing of these analytics is implemented using a toolkit consisting of a combination of mySQL queries and Excel VBA macros. Initially, log-file and posts tables are extracted from the discussion forum database and merged into a single spreadsheet file. This file lists each action taken by a student in the system in a row with the following information: action type (view-post, create-post, edit-post, delete-post), a time-date stamp, ID of user performing the action, ID of post being acted on, length of post being acted on, and ID of user who created post being acted on. A series of macros then clean the data, separate data by user, calculate action duration (subtracting sequential time stamps), divide actions into sessions-of-use (based on a 60-minute abandonment threshold, see Wise, Speer et al., 2013), and make adjusted estimates for duration of session-ending actions (based on the relevant post's length and the average speed of the indicated action by that user). View actions made on a user's own posts are re-coded as self-review actions and all view and review actions are sub-categorized as reads or scans based on a maximum reading speed of eight words per second; a conservative estimate based on Hewitt, Brett, & Peters (2007). Finally, the nine variables are calculated based on the definitions shown in Table 1. A summary table of the metrics is then created for each learner (see Figure 2). This form was chosen as a straightforward way of presenting the analytics not requiring particular visual literacy skills and thus unlikely to cause confusion. It also allows for concise presentation of multiple metrics. In future work it may also be valuable to explore other graphical ways to represent the metrics. In addition to each student's individual metrics, we also provide class averages as a local reference point. Reference points of greater aggregation (e.g., for other sections of the class, historical averages) could also be included as desired.

Metric	Your Data (Week X)	Class Average (Week X)
Range of Participation	***4 days***	5 days
Number of Sessions	***6***	13
Average Session Length	***33 min***	48 min
Percentage of Sessions with Posts	***67%***	49%
Number of Posts Made	***8***	12
Average Post Length	***149 words***	125 words
Percentage of Posts Read	***82%***	87%
Number of Reviews of Own Posts	***22***	13
Number of Reviews of Others' Posts	***8***	112

Figure 2. Sample learner analytics summary

2.4 Intervention Design

As discussed above, an effective learning analytics intervention involves more than simply providing well-designed metrics to learners; it necessitates setting up a frame for their use as an integral part of course activity tied to goals and expectations (Wise, 2014). The design of our analytics intervention was based on the principles of *Integration*, *Diversity*, *Agency*, *Reflection*, and *Dialogue* (Wise, Zhao, & Hausknecht, 2013). These are used to address possible concerns about rigidity of analytics interpretation, hegemony of optimizing to only that which can be measured, and impediment of learner's development of metacognitive and self-regulative learning skills (Buckingham Shum & Ferguson, 2012; Clow, 2012; Duval & Verbert, 2012).

Integration of analytics with the expectations of the learning activity and fabric of the course experience is critical for the analytics to be used by students in meaningful ways. Our intervention design helps students to understand the goals of the learning activity by providing clear guidelines about the online discussions, their purpose, and what is expected of students in terms of quantity, quality, and timing of posting, as well as broad, deep, integrated, and reflective attention to the posts of others (see Figure 3a). Analytics are then introduced in the context of this framework. For the embedded analytics, the appropriate sections of the participation guidelines describe how the interface provides information about particular elements of participation (see Figure 3a); for the extracted analytics, a separate guideline sheet is given with a chart describing each metric and how it relates to the participation criteria (see Figure 3b). In this way, metrics are not presented simply as a set of numbers, but ones that have clear meaning in the context of the learning activity.

a)

Discussion Participation Guidelines

Attending to Others' Posts

Broad Listening: Try to read as many posts as possible to consider everyone's ideas in the discussion. This can help you examine and support your own ideas more deeply. However, when time is limited it is better to view a portion in depth, then everything superficially.

**The visual interface shows posts that you have viewed in blue and new ones in red to help you track this.*

b)

Learning Analytics Guidelines

Attending to Others' Posts	
% of Posts Read	The proportion of posts you read (not scanned) at least once. *It is good to read as many posts as possible to consider everyone's ideas in the discussion However, when time is limited it is better to view a portion in depth, then everything superficially.*

Figure 3. Excerpt from (a) discussion participation guidelines and (b) learning analytics guidelines

Diversity in analytic metrics and *Agency* in interpreting their meaning in the context of the learning activity are important in shaping analytics as an empowering tool rather than an exacting master. Our intervention is designed to include multiple measures (as described above; see Figure 2). Importantly, the guidelines present the metrics as a starting point for consideration, not as absolute arbiters of one's engagement in the activity. To further support individual student agency in using the analytics, the intervention design has students keep an online reflective journal in which they are encouraged to set personal goals for participation and to use the analytics to help monitor these goals. This activity also supports the principle of *Reflection,* as students are given their analytic metrics in the journal space along with goal-setting and reflective prompts (see Figure 4). Students are consciously given time to write in their reflective journals and the instructor has access to and the ability to comment on students' reflections, supporting the principle of *Dialogue* around the interpretation of the metrics in the context of each student's current goals.

Week 7

Below you are provided with some detailed information about your discussion participation for the past week generated from the system clickstream data. Please refer to the analytics guideline sheet to aid your interpretation and remember that more is not always better. Note anything interesting you see in the data in the observations box and then use this information to help you answer the reflection questions.

1. What do you think went well in the group's discussion this week? What could be improved?
2. What do you think went well in your individual discussion participation in this week's discussion? What could be improved?
3. How does your participation in this week's discussion compare to previous weeks?
4. How well did you do in meeting the goal(s) you set for your participation in this week? How do you know?
5. Please set at least one goal for your participation in the coming week and write it below.

Figure 4. Sample reflective journal question prompt

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

3. IMPLEMENTATION AND METHODS

3.1 Research Questions and Design

The initial pilot of this learning analytics application was conducted as a semi-automated implementation with a small group of learners in an authentic class setting. We argue that such lightweight testing of the efficacy of theoretically grounded analytics implemented with the minimally necessary toolset is a valuable validation step prior to developing a full-blown learning analytics platform. In conducting the implementation, we asked three core research questions:

1. What were students' reactions to the embedded and extracted analytics?
2. Did the students use the analytics to try to adjust their participation? If so, how?
3. Did students' participation behaviours change through the term? If so, how?

A mixed-methods design was used to answer the research questions. Student reactions to and uses of the analytics were assessed through analysis of entries in their reflection journals and interviews conducted at the end of the term, while changes in behaviour were identified through direct examination of the log-file data. Data was examined for individual students as well as across the class to identify both unique instances and overall trends.

3.2 Ethics Procedures

Study approval was obtained from the office of research ethics at the researchers' university. The research protocol included provisions to ensure secure data collection and storage on university servers, and participant confidentiality was guaranteed via aggregated data and/or pseudonyms. After obtaining instructor consent, all students in the participating class were informed about the purpose of the study, provided with an explanation of what their involvement would entail, and invited to give their consent; those choosing to participate were able withdraw from the study at any point, although none did. To prevent any possible impact on students' educational experience, the course instructor was not present during participant recruitment and did not know the identity of students choosing to participate in the study until after the course was completed and grades were submitted.

3.3 Context, Participants, and Learning Activity

The implementation was conducted in a semester-long blended graduate seminar on educational technology consisting of nine first- or second-year doctoral students and one instructor. The seminar met once a week in a face-to-face session with a series of ten week-long online discussions interspersed between meetings. The first week of discussion was facilitated by the instructor to model good practice and to give students a chance to acclimate to the tool; each subsequent discussion week was facilitated by one of the course's nine students.

At the start of the class, students were introduced to the *Starburst* discussion forum and engaged in a conversation about the goal and purpose of online discussions in the course. They were also provided with discussion participation guidelines and an explanation of how the embedded analytics could be used to help support these. Finally, an online wiki-based reflective journal was created for each student as a space for goal setting and reflection on his or her discussion participation. At the beginning of each four-hour class session, students were given 10 to 15 minutes to write in their reflective journals based on the prompts described earlier. Between classes, the instructor was invited to read students' comments and respond as needed. For the first half of the term (five week-long discussions), the reflections were based solely on students' perceptions of the discussions and the embedded analytics. This was done to separate out the effects of the embedded analytics and provide a baseline for comparison once the extracted analytics were introduced. In the second half of the term (five more week-long discussions), students' extracted analytics were calculated for each discussion week and pasted into the reflective journal to serve as part of the prompt for reflection. Both embedded and extracted analytics were presented explicitly as objects of reflection to understand and effect change on one's discussion participation and not as part of the course evaluation.

3.4 Data Collection

3.4.1 Interviews

Shortly after the end of the term, all nine students and the instructor were each interviewed about their experiences using the learning analytics as part of the online discussion activity. The hour-long semi-structured interviews consisted of four parts: 1) questions about how the participant understood the purpose of online discussion and their participation habits; 2) their reaction to and use of the analytics embedded in *Starburst*; 3) their reaction to and use of the specific metrics provided as extracted analytics; 4) their experience using the reflective journal as the framing activity for the analytics.

3.4.2 System Data

The log-file data that had been processed into extracted analytics was taken to investigate changes in student behaviour in the discussions. The same metrics were also calculated for the discussions in the first half of the term (when extracted analytics were not provided to the students). In addition to the metrics shown to the learners (see Table 1), two additional metrics used in the research program (percentage of others' posts viewed and percentage of scans) were calculated for all ten discussion weeks to develop a more thorough understanding of learner behaviour in the discussions.

3.5 Data Analysis

3.5.1 Common Themes in Reactions to the Analytics

All interview recordings were transcribed and imported into Atlas.ti. Any text referring specifically to an aspect of the embedded or extracted analytics was marked as relevant text (Auerbach & Silverstein, 2003). Using an inductive approach (Thomas, 2006), each interview was then initially open-coded for repeating ideas by one of the researchers. Files were exchanged and a second researcher reviewed the codes, adding additional ones as needed. This ensured that the coding of each interview was informed by the full breadth of the corpus and research team. Finally, the researchers worked together with the repeating ideas to identify emergent high-level themes and characterize the relationships between them.

3.5.2 Case Studies of Analytics Use and Changes in Behaviour

A case study of analytics use, including intended and actual changes in behaviour, was conducted for each of the nine students in the class based on their interviews, reflective journals, log-file data, and post-contents. Since students were encouraged to set up personal goals and choose different paths to reach their goals, it was important to look at how their behaviour changed individually. Importantly, given that the learning analytics allowed for multiple possible profiles of productive activity and improvement, we would not necessarily expect all student behaviour to change in the same way (Wise, Zhao, Hausknecht, 2013).

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

Several analyses were performed as part of the case studies to see how individual students used the analytics and how their behaviour changed. First, we looked at the goals students set in their reflection journal each week, examining their log-file data to see whether empirically these goals were met. We also reviewed the students' own reflections on their progress towards their goals. Second, we charted each student's metrics across the first five discussions (only embedded analytics) and last five discussions (both embedded and extracted analytics) to identify any changes in behaviour induced by the extracted analytics (regardless of relation to specific goals). In doing so, it was important to take each student's facilitation week into consideration since this could possibly contribute to any changes seen in participation (Wise & Chiu, in press). Finally, to contextualize the cases, we examined students' attitudes towards particular analytics from their reflective journals and interviews to find possible explanations for their changes in behaviour. After building the case study for each student, we then looked across cases for common themes and unique instances of how students used the analytics and how their behaviours did or did not change throughout the term.

4. FINDINGS

4.1 Reactions to the Analytics

4.1.1 Usefulness of Embedded Analytics

Over half the students found the graphical branching representation useful in providing information about the structure of the discussion. Specifically they reported that it outlined the flow of the conversation, showed the connections between ideas, and provided a big picture of the discussion. Using this information, some students reported that they chose to attend to threads that were more active because they were curious about what topics people emphasized, while others were attracted to isolated threads either out of curiosity as to why they had not received much response or because of the reduced effort required. Both approaches indicated that students were using analytics embedded in the interface to make decisions about how to interact with the discussion. Two students described using the information provided about the structure of the discussion strategically, visiting isolated threads when they did not have much time, and tackling the larger threads when they did. Despite these positive results, a sizable minority of students found the graphical representation of branching confusing, particularly when there were many posts in the forum, they found the interface looked too busy and thus did not communicate the structure of the discussion effectively.

With respect to the colouring of nodes to indicate new, read, and one's own posts, just under half of the students said the coloured nodes helped them to be more aware of their participation in the discussion and they used this information to direct them to which posts to read first. Specifically several students found it useful to identify where they had posted, either to see which parts of the discussion they had contributed to already and who had replied to them, or to prioritize reading new posts that replied to their own posts. Commenting on the specific colour scheme used for this analytic (red/blue/light blue), one student found the red colour of new posts "alarming" and said that it "forced [her] to open each of them and try to read," while another suggested that the blue and light blue posts could be more strongly differentiated.

Finally, acknowledging the usefulness of having embedded analytics, several students suggested additional metrics that could be added to the tool. For example, one student recommended showing people's agreement or disagreement through the interface while another suggested that it might be helpful to add other information to show which posts are more important to the discussion.

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

4.1.2 Diverse Reactions to Extracted Analytics

Students had diverse reactions to the nine extracted analytics introduced in the second half of the course. Overall, some students found them useful in providing "hard" numbers with which to assess their participation; however, others pointed out that there is much that they do not capture. In general, student reflections on the extracted analytics seemed to fall into two classes: 1) validation of things they were already aware of and 2) metrics that were surprising. Some of the surprises were positive; for example, one person felt she was not contributing enough to the discussions but the metrics showed that she was substantially above the class average. Other surprises were negative, however; for example, for many people the extracted metric of *Percent of Posts Read* was substantially lower than what they expected based on the embedded red/blue posts-viewed analytic (the difference was due to post scanning). Reactions to analytic surprises often had an affective component with students variously feeling pleased, upset, wronged, or ashamed by their metrics.

The two metrics most generally mentioned as useful were *Average Post Length* and *Percent of Posts Read*. The metric of *Average Post Length* was almost universally popular among students. Many said they preferred reading concise posts because they made the discussion "more of a dialogue and less of an essay," and thus this metric helped them assess if they were doing a good job keeping their own posts succinct. Interestingly, one student who consistently made very short posts set the contrary goal of increasing his/her post length and felt that the *Average Post Length* metric was useful in monitoring progress on this goal. The only student who said that the metric was not useful still mentioned that it might be helpful for others to keep track of their post length.

Many students also described *Percent of Posts Read* as useful, but it was a more controversial metric. In describing its usefulness, students said that the metric encouraged them to read (rather than just quickly scan) posts and some felt they became more conscious of the way they read in the forum. The majority of students also said that they tried to adjust their reading activity based on the metric. However, several students felt that the metric did not always represent their activity accurately; specifically, they were surprised that the metric indicated that they read less than they thought they did and were defensive about this difference, suggesting that the system failed to "catch" all of their reads when they read quickly. Interestingly, the majority of students who felt that their actual reads were not measured correctly still said that they found this metric useful.

Students individually put more or less meaning into the remaining metrics with a small number of students mentioning each of the metrics as useful to them. For example, a few students found *Range of Participation* helpful to know, and noted that it was especially interesting to see how much others were able to participate. *Number of Sessions* and *Average Session Length* were appreciated by one student who used them to track time spent in the discussion, saying that a longer period was needed for people to "digest and understand threads." *Percent of Sessions with Posts* was helpful for another student to "understand how much time I was ... just reading and how many times I was reading and contributing at the same time." Some students who found *Number of Posts* helpful used it to make sure they were making *enough* posts, while others found that they were posting much more than they realized and used it to try to *reduce* the number of posts they made. Students had particularly mixed responses to the *Number of Reviews of Others' Posts*: two students thought peer review was important because you can get something new by reading others' posts again; two students were confused by the concept of review; and three students specifically found the concept of review unhelpful because they could "mostly be reading someone else kind of once and getting it." Finally, although one person said it was

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

good to see the *Number of Reviews of Own Posts* and started to do more self-reviews towards the middle of the course, most other students did not find this metric particularly helpful.

4.2 Use of the Analytics

Participants used the analytics in various ways, often leading to a change in their discussion behaviours. The following themes described the main uses of the analytics found.

4.2.1 Students Used Analytics to Set (Often Recurring) Goals for Participation

Many students found the metrics useful as an evaluation of their participation and they set goals to improve their performance. However, students had different emphases on changing various metrics. As one student pointed out, "I think based on the numbers you could… choose the direction how you want to go." For example, some students paid attention to adjust their post length: "One of [my goals] being to reduce the size, I was going quite a bit over the average in some of them so [I thought] let's reduce the number of words per post." Other students focused on improving the percentage of others' posts they read by spending more time on each post.

> I found that I wanted the challenge of trying to up the percentage of overall posts that I reviewed each week. This also meant slowing down my reading since the data would not record a quick read of the information. The overall result was that I learned more and was able to get a broader sense of opinion concerning the readings.

Although collectively students had a variety of goals, each student's particular goals were often recurring; that is, students expanded or continued goals from previous weeks rather than creating new goals each time. Many explained this by acknowledging that even when they knew what they wanted to change, it was hard to follow through; it often took a certain amount of time to adjust their behaviour and achieve their goals. This could lead to some frustration, as pointed out by one student: "I thought it was useful to reflect, but I found it frustrating to reflect and set goals and then not necessarily be able to actually meet them." However, the instructor was more positive about the need for students to work over several weeks of discussion to achieve their goals.

> I think we want to be working towards feeling comfortable with each other in how we comment in the discussions, but that's something I think I have a greater appreciation for—the fact that it has to build over time and so I saw several students who said, you know, "I don't feel comfortable participating in a certain way as of yet" and then pushing themselves to do that later.

4.2.2 Change with the Analytics were Not Always Intentional

While the majority of students set goals and tried to alter their behaviours according to the metrics, others showed changes in their discussion participation that did not seem to be intentional. These were students who described not finding the metrics particularly valuable or not trusting their accuracy, and who seemed unaware (even at the end of the course) that their discussion participation behaviours had actually changed. For example, one student focused her discussion efforts on qualitative aspects of the post contents not captured by the analytics, but still her *Percentage of Posts Read* increased and remained high after the introduction of the metrics.

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

4.2.3 Using Analytics to Monitor Behaviour

Students also sometimes used the metrics to monitor, but not necessarily change their behaviour. A good example of this was with *Average Post Length*. Post length was a greatly discussed metric in the interviews as no one wanted to read excessively long posts. Many students mentioned focusing on keeping their own posts concise from the start of the course (before the extracted analytics were introduced) and wanted others to do the same. With the introduction of the metrics, students described being able to track their post length more easily.

> I think I was pretty consistent probably all the way through, probably pretty much around 100 word post is what I would try and keep to—100–150.

4.2.4 Using Oneself, Others, and Guidelines as Reference Points for Comparison

Having a reference point with which to compare one's metrics was important for all the students in interpreting their analytics. Three specific reference points were commonly used. First, students appreciated having discussion guidelines to set up the scope of their participation and described thinking about the guidelines while participating, using them as a touchstone for how they should participate in an online discussion. Throughout the course, many students described referring to these guidelines as a point of comparison for their participation metrics.

Second, by far the most common reference point used was other students in the class. All students expressed an awareness (both in their journals and in the interviews) of the class average given in the metrics, with many individuals appreciating it as a reference point. This comparison allowed for a closer look at participation in the context of the class:

> It was good to compare because all of those people I think, except one, are coming from the [same] background and so on, and myself coming from the [other] background, it shows the difference, it shows the difference. That was good for me to analyze.

For several students who were below the class average, this also usefully drove their participation goals. However, comparisons with the class average could also take a negative form depending on the student and his/her interpretation of the data. While some students found the average interesting, motivating, and helpful, others found it surprising, intimidating, and stressful. As one student expressed it, "Since all my numbers are below the average so that makes me feel, 'Oh my gosh, I'm kind of jumping out of this class' or something like that. It is kind of a little bit — sometimes depressing."

Finally, students also used themselves as a reference point, looking at their past goals, metrics, and reflections as comparison points for their current metrics. Several students noted in their reflections when their metrics had gone up or down from the previous weeks: "Compared to previous week, [my] number of reviews of others' posts has been hugely increased (excluding my facilitation week), and I did spend more time to read and understand others' posts." Some students also used multiple previous journal entries to determine whether change was occurring over time: "Actually reading the [journal entries] after two, three weeks to get back and see where you were standing and are you still standing on the same foot or you change, yeah, that was maybe interesting."

4.2.5 Validating Invisible Activity

Finally, one of the most valuable uses of using the analytics from the instructor's perspective was that it honoured and validated discussion forum activity occurring under the surface. Specifically with respect to the metrics capturing listening data, it made her aware of the intense involvement of certain students who were very engaged with the discussion but did not always post many comments. It also highlighted a lack of listening by some of the vociferous speakers. The instructor thus used the analytics as a starting point for dialogue in the reflective journal, recognizing the listening efforts of some, while pointing out the need for others to listen more deeply.

4.3 Changes in Behaviour

4.3.1 Common Changes in Percentage of Posts Read/Viewed

The most common change observed in students' discussion behaviour related to the *Percentage of Posts Read*. With the introduction of the extracted analytics in the second half of the term, this metric rose for all but two students (who were already at the maximum of 100%). Two specific patterns were seen. First, students who had a large gap between the percentage of posts they viewed and the percentage of posts they read in the first half of the term (due to scanning behaviour) started reading more of the posts they viewed, narrowing the gap in the two metrics (see Figure 5a). Second, students whose percentage of posts viewed and read were similar, but low, showed a rise in both metrics (see Figure 5b). For two students (not shown), there is a potential confound in interpreting the change seen in percentage of posts read because their facilitation week occurred just as the extracted analytics were introduced. Thus, their rise in *Percent of Posts Read* in the second half of the term could be due to the facilitator role, the introduction of the analytics, or a synergistic effect of the two.

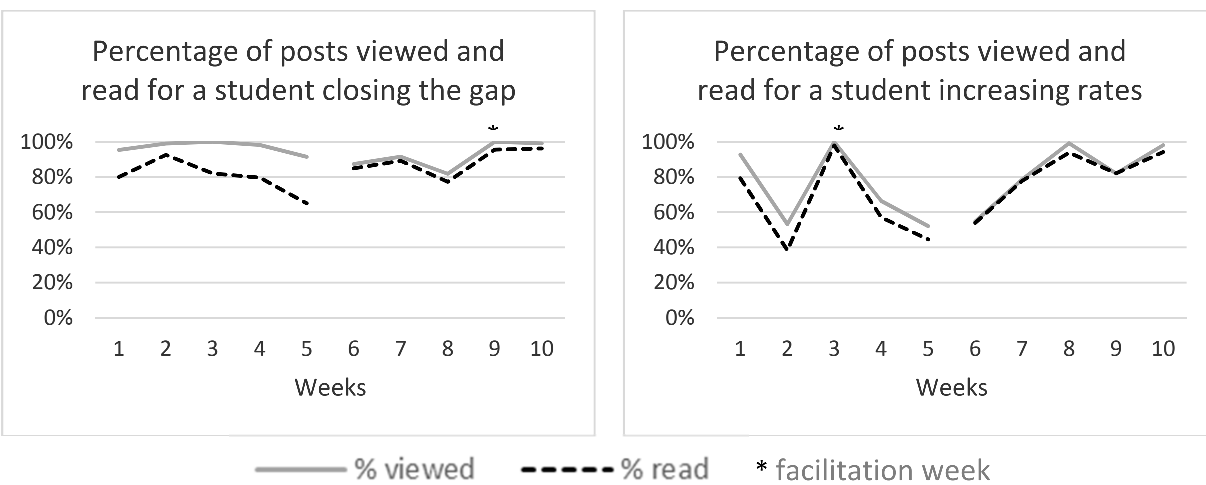

Figure 5. Percentage of posts viewed and read for students who, after the introduction of the extracted analytics, (a) narrowed the gap between posts viewed and read and (b) raised the percentage of posts both viewed and read

4.3.2 Unique Changes in Other Metrics

Besides the *Percent of Posts Read,* some students attempted to change other metrics, often accomplishing this to some degree. However, these were mostly smaller one-off changes enacted in particular weeks but not sustained over time. For example, one student wanted to extend her *Range of Participation* (days in the discussion) for Week 8, but only increased it from three to four days. Another

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

student wanted to increase his *Number of Posts Made* in Week 7; he did then contribute 15 posts (compared to 11 the week before) but dropped back down the following week. Similar to *Percentage of Posts Read*, other metrics also changed for students even if this was not explicitly set as a goal. For example, one student increased her *Number of Reviews of Others' Posts* over the second half of the term even though she did not outline this as one of her goals.

4.3.3 Change in Strategy

A few participants found that adjusting their metrics to meet their goals meant adopting new strategies for discussion participation. This is noteworthy since, as described earlier, changing strongly ingrained behaviours can be hard, thus a higher-level approach, such as a strategy shift, may be a productive tool to support change. As one student described:

> This week was interesting for me as I had some online sessions that I spent reading threads from start to finish without adding to them. This is not something I had done much of before and I found that I could follow the thread much better when I didn't stop reading to write something myself. This is reflected in the lack of reviews of my own/other's posts — I didn't need to reread postings using this method. The downside to this is that by the time I had read the whole thread through, I often didn't feel I had much to offer as far as new direction/comments/questions that hadn't already been said. So my own posting level may have been restricted.

5. DISCUSSION

The most basic yet important finding of this study was that when given access to online discussion learning analytics in the context of a framing intervention, students enacted some useful changes to their participation based on them. This serves as validation of the potential effectiveness of learning analytics for influencing student activity in online discussions. As described above, many students were purposeful in their use of the analytics, drawing on them to set, enact, and reflect on goals for their discussion forum activity. Being proactively involved and engaged in directing one's own learning is thought to support better learning processes and outcomes (Boekaerts, Pintrich & Zeidner, 2005; Zimmerman & Schunk, 2001). This is an encouraging finding regarding the possibilities of learning analytics to support students in the self-regulated learning processes of planning, monitoring, and evaluation (Winne & Hadwin, 2010).

While many students adjusted their participation according to self-set goals and what they believed was beneficial for their learning, sometimes students appeared to change their behaviour involuntarily simply due to the presence of the metrics. Such unintentional change may have positive or negative consequences depending on the changes enacted. For example, some students who felt that the metrics were not particularly helpful still increased the percentage of posts they read in the discussions; however, students who find out they are above the class average on certain metrics may involuntarily reduce their efforts in these areas, as was noted in at least one case here. This highlights that even when metrics are presented in a neutral form and with support for student agency in interpretation, the analytics still act as agents in the system with some degree of power (Latour, 2005). Thus while data may be provided for the purpose of informing decision-making, we must also always try to anticipate how it might inadvertently guide behaviour more directly.

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

This is especially important in the case of embedded analytics used in the moment rather than as part of a reflective cycle. In this study, with the help of initial stage-setting guidelines, we saw students use the embedded analytics purposefully, making strategic decisions about where to post after seeing the big picture of the discussion and where they had participated in it. This may help to avoid some typical difficulties found in asynchronous online discussions, such as new post bias (Hewitt, 2003; Wise, Marbouti, Hsiao, & Hausknecht, 2012), shallow interaction with others' posts (Thomas, 2002), and not being able to see the discussion as a whole due to disorganized threading (Dringus & Ellis, 2005). It is certainly possible, however, for embedded analytics to have inadvertent undesirable effects as well. For example, if students are simply clicking on all the red dots (unread posts) without any effort to read the posts because they find their appearance alarming, it is a distraction from meaningful participation in the discussion. This highlights a challenging design problem for listening analytics: how to balance the useful aspects of tracking which posts have been read without calling undue attention to new unread ones (Marbouti, 2012).

Turning to the extracted analytics, the most common change that students made after their introduction related to the percentage of their peers' posts that they read. It is not certain why this particular metric was most influential for students, but it may be because it both related to a clear concept they valued (attending to others' ideas) and because it revealed information that surprised them (either that they were reading fewer posts than they thought or that their peers were reading more). Thus, this analytic may have sparked a need for students to reduce the conflict between the value they placed on listening to others and the realization that they were not doing this to an appropriate extent.

For some students, the change in percentage of posts they read was enacted as a "narrowing of the gap" between posts viewed and read. In other words, these students made an effort to actually read (rather than just scan) the posts they opened. While *Percentage of Posts Read* originally was conceptualized as a measure of listening breadth, in this situation the change in the metric to match the percentage of posts viewed indicates an increased *depth* of listening since students were actually reading (not just scanning) a greater percentage of the posts they opened. However, for students already reading most of the posts they opened, a rise in the percentage of posts read indicates that they were actually attending to a greater number of posts overall — an increase in *breadth*. This distinction is important because a recent study (conducted after this analytics implementation was complete) showed that depth, but not breadth, of listening was predictive of richer argumentation in student posts (Wise, Hausknecht, & Zhao, 2014). We thus have a situation in which the same analytic resulted in two distinct changes in student behaviour with different values for learning. This emphasizes both the importance of connecting learning analytics to an empirically validated model and the need for clear usage guidelines as part of the intervention design.

Another listening quality found to be important in Wise, Hausknecht, & Zhao (2014) was reflectivity (revisiting previously read posts), which predicted more substantive responses to others' ideas. Within the current study, however, students did not particularly value the metric regarding review of others' posts. Though the participation guidelines explained the value in revisiting previously read posts to provide context for recent posts and to examine how thinking had changed, the majority of students either were confused by the concept of reviewing posts or felt that reading a post once was enough. This demonstrates that when analytics are based on concepts where the value is not obvious to students, they may need particular targeted support in understanding why (and how) to use them.

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

In general, the extent to which particular metrics were valued and used by students appeared to vary based on two different elements. First, students tended to put more weight on metrics to which they had theoretical buy-in, in other words to which they saw the connection to learning. In this application, the major support for developing such connections came from the participation guidelines describing the expected behaviours in the discussions, and the learning analytics guidelines describing the rationale for each metric. While some of the explanations of the reasons for the metrics were accepted by students (perhaps because they were simpler or more intuitive), others were not, as shown by the lack of value accorded to listening reflectivity. In future work, we will explore alternative avenues to building learner value in specific analytics through both expository and interactive means. Learners may also find certain metrics more valuable if they are measured at a smaller unit of analysis than weekly averages (Wise, Hausknecht, & Zhao, 2014).

Another element that appeared to influence student perceptions of the metrics was the transparency of measurement. For example, one of the metrics that students found most useful, *Average Post Length*, was based on a relatively straightforward calculation of the number of words in a post. However, another metric, *Percentage of Posts Read* was based on a more complicated algorithm that used a reading speed threshold to exclude posts that were only scanned. This analytic was sometimes met with skepticism, defensiveness, and denial, especially when the figure provided a negative surprise to students. While it is possible that some students read faster than eight words per second (leading to underestimation of their percent of posts read), this was chosen as a generous threshold thus it is unlikely the case for most. For this reason, emotional components of student responses to the analytics are important to consider as a more general phenomenon. This is especially an issue when introducing analytics involving more complex and less transparent calculations. For example, measures of speaking quality (post contents) were not included in this first iteration of online discussion speaking and listening analytics but are certainly critical to the value of online discussions for learning. We would expect students to buy-in to these metrics theoretically as important for productive learning discussions, but how students react to receiving computer-based (computational linguistic) assessments of their posts' qualities (e.g., Rosé et al., 2008; Ferguson & Buckingham Shum, 2011) or of the extent to which their posts show uptake of the contents of prior posts (Suthers, Dwyer, Medina & Vatrapu, 2010) remains to be seen. We suspect that students receiving unexpected negative feedback (those who could benefit most from the analytics) may become defensive and question the accuracy of how the analytics are produced. This kind of situation may thus require additional intervention and support from an instructor.

A final area for which additional support may be required is the process of change itself. Many students noted that changing how they participated in the discussion was difficult, even when they knew what they wanted to change and had set goals to this effect. In this case, an extended period for adjusting one's behaviours as well as support for this process as part of the intervention design were useful. Other learning analytics interventions may also require support for students in making desired changes. One value that an instructor could provide in such situations is in suggesting ways the student could shift his/her learning strategy (rather than just targeting a specific behaviour) to lead to the desired effects. This may be particularly helpful for students who are less skilled in regulating their own learning (Nandagopal & Ericsson, 2012). In addition, it is important for the design of the learning analytics intervention to acknowledge and support change as a process that occurs over time. As was seen in this study, many students valued being able to review their analytics over several successive discussions and set recurring goals over multiple weeks. This is important, as learning itself is an evolving developmental phenomenon that is temporal in nature (Mercer, 2008). Thus the ability for students easily to review

(2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics, 1*(2), 48-71.

both their current analytics and their historical data as a reference point, as well as to have the freedom to work on improvement over multiple analytic cycles seems a valuable attribute for learning analytics applications more generally. Similar to EnquiryBlogger (Buckingham Shum & Deakin Crick, 2012), the system we designed also allowed students to track their goals and reflections over time, supporting awareness of the changes that occurred and allowing for a potentially deeper understanding of the process of altering one's activity patterns.

Besides the metrics for each individual over time, the class average also played an important role in influencing student changes in behaviour by providing a contemporaneous comparison point. While the class average can be a powerful reference point for low-activity students, due to the power of peer influence there is also a danger that its use can lead to a reduction in activity diversity (all members of the learning community collapse to a single discussion participation profile). Thus, in future work, it would be preferable to provide a range of peer activity as a reference frame to support students in determining what changes to make to their discussion participation. Giving a range or gradient of peer activity as a reference frame may be best communicated in a graphical form, rather than the numerical presentation used here. Another area for future analytics work is using peer activity to provide information about post reception; for example, using a measure of how certain posts (or learners) are attended to by others (Brooks, Greer, & Gutwin, in press) as a proxy for their influence or importance in the discussion.

6. LIMITATIONS

The current work has several limitations, primarily related to conducting the pilot in a small context with advanced and motivated students. First, the model of reflective journals with one-on-one dialogue between the instructor and each student is not sustainable at scale. To replicate this interaction in more populous contexts, instructors can consider using peer commenters or consolidating the reflections into larger units as a formal assignment. We have implemented the latter approach to reflection (without analytics) in a large undergraduate class, suggesting that it would also be viable for this purpose. Second, the *Average Post Length* metric counted all words in a post, incorporating any direct quotations taken from prior posts or readings. In the context of a discussion tool without an auto-quote function and with advanced students this was not a problem since quotes were generally used sparingly and meaningfully. However, with younger, more novice students, overuse of extended direct quotations may artificially inflate this metric, requiring refinement of the calculation algorithm.

Finally, and most importantly, the specific findings in this context may not generalize to the larger population of students taking part in online discussions as an educational activity. This is especially true for students just starting post-secondary studies, less keen on learning, or studying different kinds of subject matter. Further testing with a diverse range of students, subject areas, and instructional levels will reveal the extent to which the specific results found here are robust for the larger population of online and blended learners (or subsets thereof), and across different learning contexts. As well, we expect other patterns to emerge when more student data is available for analysis. In contrast, some of the conceptual issues we uncovered relate to putting embedded and extracted online discussion analytics into practice more generally. Thus, phenomena such as unintentional discussion behaviour change and the presence of affective reactions to the metrics may provide conceptual generalizability useful to future online discussion learning analytics efforts.

7. CONCLUSION

This paper has described one effort to develop and test a learning analytics application based on an existing educational research program. We presented a theoretical framework and empirical findings about students' speaking and listening in online discussions, used this to develop analytics both *embedded in* and *extracted from* the learning environment. In addition, we designed an intervention to frame analytics use as an integral part of the learning activity. Together, the careful design of embedded analytics, extracted metrics, and their framing for use contribute to a system that employs learning analytics as a guide for sense-making of discussion participation with the goal of empowering students to take responsibility for regulating their own learning processes.

While the results of this study are promising for the use of learning analytics to support student participation in asynchronous online discussions, we emphasize that the sample was small and not representative of the general population of students taking part in online discussions. Thus, the extent to which specific findings (e.g., the usefulness of the embedded analytics, the dominance of changes in the percentage of posts read, and the setting of recurring goals) may generalize remains for future work. We suggest, however, that the higher-level issues related to putting analytics into practice revealed by this study have useful conceptual generalizability. Thus we present the following five concerns to be considered in future work regarding online discussion learning analytics targeting students. First, with the introduction of the analytics, students may enact not only purposeful, but also unintentional changes in their activity. Second, the same analytic can lead to multiple distinct changes in student behaviour. Third, for students to use particular analytics they need to value them; this may be affected both by their theoretical buy-in to the measure and the transparency/perception of the accuracy of its calculation. Fourth, the affective component of students' reactions to the analytics needs to be taken into account. Finally, while current analytics systems work to help us understand diagnostically *what* needs to change, students may also require additional support in learning *how* to enact that change.

ACKNOWLEDGEMENTS

We would like to thank the Social Sciences and Humanities Research Council of Canada for their generous support of this work and Farshid Marbouti for the use of *Starburst* and for making the adaptations to use it in this work.

REFERENCES

Anderson, T. (2008). Towards a theory of online learning. In T. Anderson (Ed.), *The theory and practice of online learning* (2nd ed.). (pp. 45–74). Edmonton, AB: Athabasca University Press.

Auerbach, C. F., & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. New York: New York University Press.

Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Eds.). (2005). *Handbook of self-regulation*. San Diego, CA: Academic Press.

Brooks, C., Greer, J., & Gutwin, C. (in press). The data-assisted approach to building intelligent technology enhanced learning environments. In J. Larusson & B. White (Eds.), *The handbook of learning analytics: Methods, tools and approaches.* New York: Springer.

Buckingham Shum, S., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. In S. Buckingham Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd Conference on Learning Analytics and Knowledge* (pp. 92–101). Vancouver, BC: ACM.

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society*, *15*(3), 3–26.

Clark, D. B., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review*, *19*(3), 343–374.

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. In S. Buckingham Shum, D. Gašević, & R. Ferguson (Eds.), *Proceedings of the 2nd Conference on Learning Analytics and Knowledge* (pp. 134–138). Vancouver, BC: ACM.

Dennen, V. P. (2001). *The design and facilitation of asynchronous discussion activities in web-based courses: Implications for instructional design theory* (Unpublished doctoral dissertation). Indiana University—Bloomington.

Dringus, L. P., & Ellis, T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education*, *45*(1), 141–160.

Duval, E., & Verbert, K. (2012). Learning analytics. *E-Learning and Education, 1*(8) urn:nbn:de:0009-5-33367

Ferguson, R. (2012). *The state of learning analytics in 2012: A review and future challenges*. Technical Report KMI-2012-01. Milton Keynes, UK: Knowledge Media Institute*,* The Open University. Retrieved from http://kmi.open.ac.uk/publications/techreport/kmi-12-01

Ferguson, R., & Buckingham Shum, S. (2011). Learning analytics to identify exploratory dialogue within synchronous text chat. In P. Long, G. Siemens, G. Conole, & D. Gašević (Eds.), *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 99–103). Banff, AB: ACM.

Henri, F. (1992). Computer conferencing and content analysis. In A. R. Kaye (Ed.), *Collaborative learning through computer conferencing: The Najaden papers* (pp. 115–136). New York, NY: Springer.

Herring, S. (1999). Interactional coherence in CMC. *Journal of Computer-Mediated Communication, 4*(4). doi: 10.1111/j.1083-6101.1999.tb00106.x

Hew, K. F., Cheung, W. S., & Ng, C. S. L. (2010). Student contribution in asynchronous online discussion: A review of the research and empirical exploration. *Instructional Science*, *38*(6), 571–606.

Hewitt, J. (2003). How habitual online practices affect the development of asynchronous discussion threads. *Journal of Educational Computing Research*, *28*(1), 31–45.

Hewitt, J., Brett, C., & Peters, V. (2007). Scan rate: A new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. *The American Journal of Distance Education*, *21*(4), 215–231.

Jonassen, D. H., & Kwon, H. (2001). Communication patterns in computer mediated versus face-to-face group problem solving. *Educational Technology Research and Development*, *49*(1), 35–51.

Jovanovic, J., Gašević, D., Brooks, C., Devedzic, V., Hatala, M., Eap, T., & Richards, G. (2008). LOCO-Analyst: Semantic web technologies in learning content usage analysis. *International Journal of Continuing Engineering Education and Life Long Learning*, *18*(1), 54–76.

Kamel Boulos, M. N., & Wheeler, S. (2007). The emerging Web 2.0 social software: An enabling suite of sociable technologies in health and health care education1. *Health Information & Libraries Journal*, *24*(1), 2–23.

Kear, K. (2001). Following the thread in computer conferences. *Computers & Education*, *37*(1), 81–99.

Latour, B. (2005). Reassembling the social: An introduction to actor-network theory. New York: Oxford University Press

Lipponen, L. (2002). Exploring foundations for computer supported collaborative learning. In G. Stahl (Ed.), *Proceedings of the International Conference of Computer Supported Collaborative Learning* (pp. 72–81). Boulder, CO: ISLS.

Marbouti, F. (2012). *Design, implementation and testing of a visual discussion forum to address new post bias* (Unpublished master's thesis), Simon Fraser University, Burnaby, BC.

Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *The Journal of the Learning Sciences*, *17*(1), 33–59.

Nandagopal, K., & Ericsson, K. A. (2012). An expert performance approach to the study of individual differences in self-regulated learning activities in upper-level college students. *Learning and Individual Differences, 22*(5), 597–609.

Pena-Shaff, J. B., & Nicholls, C. (2004). Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education*, *42*(3), 243–265.

Peters, V. L., & Hewitt, J. (2010). An investigation of student practices in asynchronous computer conferencing courses. *Computers & Education*, *54*(4), 951–961.

Roll, I., Aleven, V., & Koedinger, K. R. (2010). The invention lab: Using a hybrid of model tracing and constraint-based modeling to offer intelligent support in inquiry environments. In V. Aleven, J. Kay, and J. Mostow (Eds.), *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 115–124). Pittsburgh, PA: Springer.

Rosé, C., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237–271.

Rovai, A. P. (2007). Facilitating online discussions effectively. *The Internet and Higher Education*, *10*(1), 77–88.

Scardamalia, M. (2004). CSILE/Knowledge forum®. In A. Kovalchick & K. Dawson (Eds.), *Education and technology: An encyclopedia* (pp. 183–192). Santa Barbara, CA: ABC-CLIO.

Siemens, G., Gašević, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., Duval, E., Verbert, K., & Baker, R. S. (2011). *Open learning analytics: An integrated & modularized platform*. [Concept paper]. Society for Learning Analytics Research.

Stahl, G. (2005). Group cognition in computer-assisted collaborative learning. *Journal of Computer Assisted Learning*, *21*(2), 79–90.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, *5*(1), 5–42.

Swan, K. (2003). Learning effectiveness: What the research tells us. In J. Bourne & J. C. Moore (Eds.), *Elements of quality online education, practice and direction* (pp. 13–45). Needham, MA: Sloan Center for Online Education.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, *27*(2), 237–246.

Thomas, M. J. (2002). Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning*, *18*(3), 351–366.

Webb, E., Jones, A., Barker, P., & van Schaik, P. (2004). Using e-learning dialogues in higher education. *Innovations in Education and Teaching International*, *41*(1), 93–103.

Winne, P. H., & Hadwin, A. F. (2010). Self-regulated learning and socio-cognitive theory. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed.) (pp. 503–508). New York: Elsevier, Oxford.

Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In S. Teasley & A. Pardo (Eds.) *Proceedings of the 4th Conference on Learning Analytics and Knowledge* (pp. 203–211). Indianapolis, IN: ACM.

Wise, A. F., & Chiu, M. M. (in press). The impact of rotating summarizing roles in online discussions: Effects on learners' listening behaviors during and subsequent to role assignment. To appear in *Computers in Human Behavior*.

Wise, A. F., Hausknecht, S. N. & Zhao, Y. (2014). Attending to others' posts in asynchronous discussions: Learners' online "listening" and its relationship to speaking. *International Journal of Computer-Supported Collaborative Learning, 9*(2) 185–209.

Wise, A. F., Hausknecht, S., Zhao, Y. (2013). Relationships between listening and speaking in online discussions: An empirical investigation. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar (Eds.), *Proceedings of the International Conference on Computer Supported Collaborative Learning, Vol. I* (pp. 534–541). Madison, WI: ISLS.

Wise, A. F., Hsiao, Y. T., Speer, J., Marbouti, F., & Perera, N. (2012). Initial validation of "listening" behavior typologies for online discussions using microanalytic case studies. In J. van Aalst, K. Thompson, M. Jacobson, & P. Reimann (Eds.), *Proceedings of the International Conference of the Learning Sciences* (pp. 56–63). Sydney, Australia: ISLS.

Wise, A. F., Marbouti, F., Hsiao, Y. T., & Hausknecht, S. (2012). A survey of factors contributing to learners' "listening" behaviors in asynchronous online discussions. *Journal of Educational Computing Research*, *47*(4), 461–480.

Wise, A. F., Perera, N., Hsiao, Y. T., Speer, J., & Marbouti, F. (2012). Microanalytic case studies of individual participation patterns in an asynchronous online discussion in an undergraduate blended course. *The Internet and Higher Education*, *15*(2), 108–117.

Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y. T. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science*, *41*(2), 323–343.

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. In D. Suthers & K. Verbert (Eds.), *Proceedings of the 3rd Conference on Learning Analytics and Knowledge* (pp. 48–56). Leuven, Belgium: ACM.

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1*(2), 72-93.

Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community

Iassen Halatchliyski
Knowledge Media Research Center, Germany

Tobias Hecking, Tilman Göhnert, H. Ulrich Hoppe
University of Duisburg-Essen, Germany

i.halatchliyski@iwm-kmrc.de

Abstract: This paper introduces the scientometric method of main path analysis and its application in an exemplary study of the paths of knowledge development and the roles of contributors in Wikiversity. Data from two scientific domains in this online learning community has been used. We see this as a step forward in adapting and adopting network analysis techniques for analyzing collaboration processes in knowledge building communities. The analysis steps are presented in detail including the description of a tool environment ("workbench") designed for flexible use by non-computer experts. By identifying directed acyclic graphs, the meaningful interconnections between developing learning resources are analyzed by considering their temporal sequence. A schema for the visualization of the results is introduced. The potential of the method is elaborated for the evaluation of the overall learning process in different domains as well as for the individual contributions of the participants. Different outstanding roles of contributors in Wikiversity are presented and discussed.

Keywords: knowledge artifacts, idea flow, scientometrics, main path analysis

1. INTRODUCTION

Nowadays, it is commonplace to perceive learning and knowledge building as closely related activities on the Web. Knowledge building is based on *epistemic artifacts* (Knorr-Cetina, 2001) created and shared in a community. Bereiter and Scardamalia (2003) point out that knowledge building is essential for learning but has a wider scope in that it is not necessarily limited to explicit learning scenarios. Scientific research is an example of a distributed knowledge building activity that takes place in scientific communities and typically is not characterized as learning. According to Scardamalia and Bereiter (1994), the knowledge building pedagogy takes scientific research as a blueprint of the collaborative learning of students that needs to be facilitated. During a knowledge-building process, students discuss ideas and develop their shared knowledge in the manner of scientists. The philosophical foundation of this view dates back to Popper (1968), who explains the development of scientific knowledge as a constant process of emergence of new ideas and their gradual improvement or abandonment after discovering contradictory evidence. In fact, any learning community defines concepts and builds its knowledge base in a similar way (Stahl, Koschmann, & Suthers, 2006).

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

With the present work we offer an approach to analyzing learning processes organized in the form of online knowledge building. Online knowledge building is characterized by collaborative activities and the creation of shared artifacts within a community of learners. This form of collaborative learning is becoming increasingly popular on the Web and goes beyond formal educational contexts (Halatchliyski, Moskaliuk, Kimmerle, & Cress, in press). As this is a relatively new phenomenon and it shifts the focus from the individual learner to the knowledge processes within a community, appropriate methodologies are expectedly complex and in a very early developmental stage.

Due to the relation between scientific production and learning in communities, we aim to show that both processes can be studied using the same analytical approaches. S*cientometrics* as a research field is particularly concerned with the quantitative measurement of scientific work, and so offers a variety of potentially fruitful approaches new to the area of learning analytics (Suthers & Verbert, 2013). Scientometric methods are tailored for the analysis of knowledge artifacts, most prominently publications, and their authors. One well-known method is the calculation of the h-index as a measure of scientific reputation (Hirsch, 2005). In the context of learning communities, however, individual excellence is not a primary concern. Rather more interesting would be an approach to the long-term characteristics and the dynamics of interactive learning environments.

Hummon and Doreian (1989) have proposed a method to detect the main idea flows based on citation networks using a corpus of publications in DNA biology as an exemplar. Our work reported in this paper takes the *main path analysis* technique as a starting point in the analysis of a broad range of knowledge building processes that take place in formal as well as informal collaborative settings. After an initial promising application of main path analysis to networks of knowledge artifacts created for educational purposes (Halatchliyski, Oeberst, Bientzle, Bokhorst, & van Aalst, 2012), we now want to elaborate on the adaptation and adequate formalization of the method. Our guiding question in this endeavor is: What kind of insights can be gained from the main path analysis of knowledge creation in online learning communities? We will explore this question using data from Wikiversity[1] as an example. Wikiversity is understood by its active members as an "open learning community" in which users can actively produce learning resources for a broad range of topics in the form of web-based hypermedia. In our view, it represents a challenging and yet relevant field for exploring the potential of scientometric methodology to tackle the dynamics of computer-supported learning processes.

2. BACKGROUND

2.1 Community learning

New knowledge in the world might be the accomplishment of an individual, but it is inconceivable without the body of previously existing knowledge that in turn has been established by many other individuals. Consequently, learning and development of new knowledge must be examined in the context of the community in which they take place.

Online communities like Wikiversity facilitate learning through the creation of a shared knowledge base in the form of digital artifacts such as texts, pictures, or other multimedia. Users can passively learn by making use of the existing artifacts. Users can also actively learn by participating in the creation of new

[1] http://www.en.wikiversity.org/

artifacts. The *knowledge building theory* suggests incorporating such activity in formal education (Scardamalia & Bereiter, 1994). Students are expected to benefit from self-motivated exploration of knowledge areas when they share and build on each other's findings in a collaborative online environment. During this long-term process, the shared community knowledge develops as ideas are constantly improved by the participants. Individual learning is an outcome of the knowledge development of the whole community.

The collaborative production of digital knowledge artifacts has become widespread since the emergence of Web 2.0. Widely and easily available tools such as wikis afford a long-term process of mass collaboration, as artifacts are built piece-by-piece and individual contributions have variable sizes. Moreover, a single contribution to an artifact can be revised or be built upon in order to produce newer versions. Every change to the shared artifacts of a wiki community can be logged as an individual contribution activity, but the ongoing development of the knowledge base is an emergent product of the community as a whole. Intersubjectivity and shared meaning-making are epiphenomena of the interaction among individuals in a community (Stahl et al., 2006). From the systemic view of the *co-evolution model* of individual learning and collaborative knowledge building (Cress & Kimmerle, 2008), a community and the participating individuals function as two different types of systems that co-evolve through mutual fertilization. Knowledge development is reflected in the changing shape and content of the artifacts.

Knowledge artifacts often hold connections among themselves that are marked by higher-level semantic structures like topical relations, problem-solution chains, discourses, etc. Regardless of whether these connections are deliberately made by the participants in a community or whether they are automatically produced by the online environment, hypermedia links bear meaning. This meaning is an integral part of the knowledge created by a community. It is also subject to change, as connections are added or deleted in parallel with the artifact development.

In sum, learning in a community represents a complex process dependent on the activities of many participants and supported by the use and development of artifacts as learning resources. The process evolves with the constant change of the shared knowledge base at the level of single resources or their interconnections.

2.2 Temporality of a Learning Process

The learning of an individual or of a whole community is a process that essentially develops over periods of time. New knowledge is built upon existing knowledge. A knowledge base develops gradually as its information content evolves. Single ideas become more concrete, they can flow together or split into independent directions, marking a convergence or divergence in the development process (Halatchliyski, Kimmerle, & Cress, 2011). At a higher level of abstraction, the interconnections within the knowledge base also develop when new ideas are added to existing content, or when already existing connections are subsequently changed.

All these changes should be studied in order to understand the corresponding learning processes. Accordingly, the temporal dimension should be regarded as a main component of learning analytics. However, the modelling of the overall process of knowledge development is challenging, as the sequential relations between all the changes in the knowledge base need to be tracked. Any aggregation across time easily leads to a biased analysis of individual and community-level variables. A longitudinal

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

study of different points in time is also an unsatisfactory option, as it misses out on the authorship of changes made between the chosen time points. Especially difficult to grasp is the nonlinear flow of ideas that is characteristic to any learning process.

Previous work in the area of computer-supported learning has paid attention to the interactivity of collaborative processes and thereby implicitly to learning dynamics. Environment data logs have been used to describe and map interaction patterns. Their interpretation has often been supported by additional analysis of the content in the case of discussion board messages (see, for example, Hara, Bonk, & Angeli, 2000; Schrire, 2004). Suthers, Dwyer, Medina, and Vatrapu (2010) also presented a universal framework for describing interactivity in the form of uptakes between contributors independent of the environment used. Nevertheless, the field of learning analytics still needs a method to address the temporality of learning processes quantitatively. Aspects that need to be taken into account include who influenced whom, which ideas were taken up in later stages and which were not, and how differently do the participants contribute to the overall learning process. The method should also be adaptable to the multiplicity of learning environments and communities that have emerged with Web 2.0.

Different forms of sequential analysis of learner actions have also been developed in order to detect and understand the best practices of orchestration of tools and content in the learning process (Cakir, Xhafa, Zhou, & Stahl, 2005; Jeong, 2003; Perera, Kay, Koprinska, Yacef & Zaïane, 2009). Frequently occurring sequences of actions or events reveal connections between the learning history as captured in log files and learning performance. Such analysis should help warn learners against inefficient strategies and also better adapt the environment and the learning materials to their needs (Zaïane & Luo, 2001). Although it certainly accounts for the temporal dimension and thereby gives deeper insight into the learning process, sequential analysis as a data-mining technique relies on the *a priori* definition of activity and event categories. The necessary coding scheme always represents a potential weak point in the analysis as it predetermines the level of abstraction and the scope of possible patterns that can be found. The method also lacks the possibility of utilizing information on the relations between specific participants or artifacts. The latter lend themselves to analysis with a network perspective.

Social network analysis (SNA) has been used in various areas, including computer-supported collaborative learning (Aviv, Erlich, Ravid, & Geva, 2003; de Laat, Lally, Lipponen, & Simons, 2007; Harrer, Malzahn, Zeini, & Hoppe, 2007; Reffay & Chanier, 2002). The basic approach relies on representing communication events as links between the actors in the network. Applied to networks of knowledge artifacts on the Web, SNA can be an efficient approach to knowledge and its collaborative development by analyzing the meaningful structure of connections between knowledge artifacts (Halatchliyski et al., in press). The resulting network structure will very much depend on the time span during which these events are collected (Zeini, Göhnert, & Hoppe, 2012). However, the target representation no longer represents temporal characteristics. For this reason, SNA has been criticized for eliminating time. Although advances are being made to analyze the development of networks, these rarely address true network dynamics. Process temporality represents a major dimension of online learning and should not be ignored in an analysis. In this paper we present a network analysis technique that can explicitly address learning dynamics in the context of an open learning community.

3. ANALYTICAL APPROACHES TO KNOWLEDGE DEVELOPMENT

3.1 Actor-Artifact Networks

The knowledge building process develops around the creation of knowledge artifacts. A specific version of a so-called two-mode-network can be built on the basis of the relation between the actor (or author) and the artifact (or product). In the SNA methodology (Wasserman & Faust, 1994), such two-mode networks are also called affiliation networks. In the pure form, these networks are assumed to be bi-partite, that is, only links alternating between actor-artifact ("created/modified") and artifact-actor ("created/modified-by") would be allowed. Using simple matrix operations, such bi-partite two-mode networks can be "folded" into homogeneous (one-mode) networks. Here, for example, two actors would be associated if they have acted upon the same artefact (Suthers & Rosen, 2011). We would then say that the relation between the actors was mediated by the artifact. A typical example of such a transformation is offered by co-publication networks based on co-authorship. Similarly, we can derive relationships between artifacts by considering agents (engaged in the creation of two different artifacts) as mediators.

The "pure" view of actor-artifact relations as bi-partite networks has a clear mathematical-operational structure. However, there are good reasons to extend this approach: Both actors and artifacts may be interrelated in other ways than by this type of cross-wise mediation. For instance, social relations between actors may operate independently of the artifact mediation. Semantic relations may be salient between knowledge artifacts, as in the "semantic web." Mika (2007) was one of the first to elaborate on methods and potential gains of blending social and semantic network structures. Other approaches allocate actors and artifacts on different layers of a multi-layer model with homogeneous relation within each layer and heterogeneous relations in between (Reinhardt, Moi, & Varlemann, 2009; Suthers & Rosen, 2011). Such multi-relational representations may appear superior in expressiveness; however, operations in such structures are more difficult to define.

As with any other network representation, actor-artifact networks also fail to capture the notion of time explicitly. However, time may be implicitly modelled in the network relations. In the scientometric field, this is the case for citation networks: If publication X cites publication Y, we can safely assume that Y is older than X. The ensuing network structure does not contain cycles (excluding specific rare cases of cross-citation). The main path analysis method builds on such acyclic citation networks and can also be adapted to the dynamics of networks of knowledge artifacts built in the process of online collaborative learning.

3.2 Main Path Analysis

The main path analysis (Hummon & Doreian, 1989) is a network analysis technique for the scientometric study of scientific citations over a period of time. Its major application is the identification of key publications in the development of a scientific field. While many scientometric methods, such as the analyses of co-citation and co-authorship networks, stress the semantic structure of scientific work, main path analysis additionally considers the temporal structure of development. Temporality is accounted for through the very definition of a *directed acyclic graph* (DAG) where nodes are single publications and directed edges represent citations between publications. The direction of an edge corresponds to the flow of knowledge from the cited and older publication to the citing and newer

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

publication. Therefore, these links incorporate both the dimension of content relations and the temporal order of the contributions.

A DAG always contains at least one node with no ingoing edges (i.e., a source) and at least one node with no outgoing edges (i.e., a sink). In the citation network of scientific publications within one field, often one important publication is chosen as a starting point for the development of the field. This publication represents the first source. Later on, other sources that may not have cited previous publications in the field can become prominent and highly cited. Sink nodes, in contrast, represent either unimportant or very new publications not cited at the time of analysis.

The main path can be described as the most used path among all possible paths of successive edges from the source nodes to the sink nodes in a citation network. This most used path can be found by a two-step procedure: first, the traversal counts for each edge are calculated as the number of different paths between each source and sink nodes that go through this edge and, second, an algorithm is used to identify the main path based on the edge traversal counts.

This paper employs the search path count (SPC) algorithm (Batagelj, 2003), which introduces one fictitious source node and one fictitious sink node and links these to each of the actual source and sink nodes, respectively. In the example in Figure 1 the fictitious source and sink nodes are 1 and 10. Their only purpose is to simplify the original procedure (Hummon & Doreian, 1989) of weight calculation for the edges connecting the real nodes. Starting at the fictitious source node (1), the main path is identified by successively following the edge with the maximal weight to the next node until the fictitious sink node (10) is reached. At node 7 in Figure 1, there are two possible alternatives to reach the next node, because both outgoing edges have the same traversal weight, in this case the main path branches.

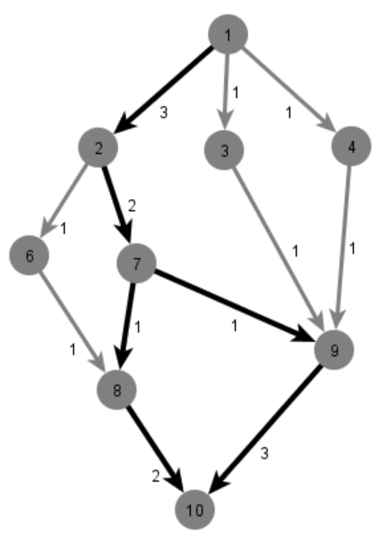

Figure 1. Example of a main path calculation

The SPC algorithm might present too strict an approach to the idea of main path, depending on the nature of the graph. For the case when the analysis requires a broader view on the main contributions in a field, Liu and Lu (2012) suggested lowering the search constraint by defining a threshold. In each step, one chooses not only the edges with the maximum weight but also edges with weight above a certain percentage of the maximum weight. In the present work, we applied a slightly modified procedure to identify the *multiple* main paths (Liu & Lu, 2012): After calculating the traversal weight of each node, we considered all the nodes with a weight above a certain threshold as part of the multiple main paths. This strategy facilitates the identification of multiple main paths of important but thematically disparate contributions that may not necessarily build one connected component.

Methods related to the main path analysis represent a structural approach appropriate for addressing the dynamics of online community learning. Depending on the nature of hyperlinks, a DAG may trace the flow of influences between ideas or the change in meanings that accompanies knowledge development. The technique allows identifying the most influential contributions and their authors in the course of the construction of a community knowledge base over time. It also facilitates the characterization of the overall discourse trajectory in collaborative learning (Halatchliyski et al., 2012).

4. EMPIRICAL STUDY

4.1 The Context of the Wikiversity Data

Wikiversity is an online learning environment operating on a wiki technology since 2006. Like its larger and older sister projects, Wikipedia and Wikibooks, Wikiversity is offered in many languages and directed at any Web user. It is not designed as the online version of an academic organization providing courses or exam certificates. It is rather an experimental open space for collaborative learning to be used by any groups of participants according to their learning goals. A major feature is the openness of the created artifacts and of the community practices to accept constructive suggestions and participation by any interested user. Thus, Wikiversity follows a learner-centred approach (Bonk & Cunningham, 1998).

As a constantly developing so-called *open learning community*, Wikiversity accumulates a rather diverse body of many types of learning resources loosely structured in scientific topics from accounting to zoology. The pages categorized under any one Wikiversity category are often set up by different users and may serve different purposes. There are separate articles but also pages connected as bigger projects or organized as courses. Nevertheless, there are often hyperlink interconnections between the different pages and contributors often join multiple projects, sometimes years after their initial start. Because of the openness, there is a great variety of participation modes within and between the different topic categories.

The development of participation is an essential part of the learning process for users. In fact, users who become more involved with the community extend their participation to many unrelated scientific topics. Even when experienced users stay within the borders of one scientific category, their contributions increasingly follow the dynamics of the shared online environment and go beyond the starting individual goals. Such possible starting goals might be, for example, the arrangement of materials for a clearly delineated course as a teacher or the participation in such a course as a student, often in connection with offline lectures in parallel. Similar scenarios of online learning and teaching in Wikiversity do occur but are not representative of the idea that the community envisions, because this form of participation is not particularly collaborative. In the long run, the learning of individuals should become interconnected, producing an interwoven socio-epistemic fabric of a community constantly open to new constructive contributions.

Because of the non-homogeneous learning practices and artifacts, the Wikiversity data represents a real challenge for a learning analytics specialist. In the following, we present our approach for discerning major patterns of learning activities and profiles of contributing participants.

4.2 Extraction and Preparation of Wiki Data

As explained in section 3.1, the main path analysis was originally developed as a method to investigate the main discourse structure of scientific fields, using networks of publications linked by citations. However, the analysis method is not restricted to this field of application. The first author and colleagues have already demonstrated how it can be applied in the educational context of computer-supported classroom discussions (Halatchliyski et al., 2012). Moreover, it can be applied to any kind of directed acyclic graph (DAG). In this paper we show how to employ the main path analysis approach to examine the development of interconnected learning resources related to a knowledge domain in the context of a wiki environment.

All analyses presented in this paper are based on data from an official dump file[2] of the English Wikiversity from 20 February 2012. We did not use the complete wiki data but employed the concept of MediaWiki[3] categories in order to identify the body of artifacts related to a specific knowledge domain. Each wiki page can be categorized under one or more headings. The categories are themselves structured into subcategories. The actual data gathering process usually starts with extracting the complete subcategory structure by following the hierarchy starting at a given top-level category. In a second step, all pages organized into at least one of the categories found in this structure are identified. It is not mandatory that each wiki page be categorized, but approximately 70 percent of all articles in the English Wikiversity belong to at least one category. Thus, we assume that our procedure yields a representative selection of the major learning resources in a knowledge domain. The chance of considering pages unrelated to a domain, which can happen when complete subcategory structures are extracted, also needs to be eliminated. One example is the category "electrical engineering" which contains "Wikiversity" as a subcategory with its large number of administrative pages that are factually unrelated to electrical engineering. Therefore, a list of subcategories for exclusion from the extraction process needs to be predefined.

As a next step, a directed acyclic graph is constructed, describing the complete flow of knowledge within a single domain in a wiki. Networks of hypermedia resources in a wiki are analogous to networks of publications interconnected by citations. Wiki pages can be regarded as publications connected by hyperlinks instead of citations. Both citations and hyperlinks indicate a flow of knowledge with a direction from a source (i.e., a cited paper or a hyperlinked page) to a target (i.e., a citing paper or a hyperlinking page).

The temporal stability of publications is crucial for the generation of a DAG from citation networks. Moreover, only works already published can be cited. In contrast to scientific publications and citations featured in their content, which are published once and then remain static from that point on, wiki pages evolve over time under the collaborative efforts of community members. Furthermore, it is quite natural that one wiki page is hyperlinked to a second page and, at the same time, the second page links

[2] http://dumps.wikimedia.org/enwikiversity
[3] http://www.mediawiki.org/

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

back to the first one, thus introducing a cycle. In order to overcome these problems, we used the Wikiversity revision logs and the page versions after each revision contained in the dump.

Regarding stability over time, revisions of a wiki page behave like classical publications. They are created (published) at a certain point in time and do not change later on. A change to a wiki page will result in a new revision and thus a modified content of that page but not in a modification of the former revision. This approach suggests using page revisions instead of wiki pages as nodes in a DAG extracted from wiki data. We distinguish between two types of directed edges in such graphs: update edges and hyperlink edges.

Update edges can be introduced between any two directly subsequent revision nodes that belong to the same page. Update edges are directed from the older revision to the newer, updated revision and, thus, represent knowledge flow over the course of the collaborative process on a single wiki page.

Hyperlink edges can be traced between two revision nodes that belong to different pages with a hyperlink pointing from one to the other. A wiki hyperlink almost exclusively points to a page and not to a specific revision and it can be interpreted as an inversely directed knowledge flow, so in the proposed DAG hyperlink edges go in a direction *opposite* to the direction of the hyperlinks in the wiki. A knowledge flow between two wiki pages is elicited at the moment of the hyperlink creation between them. Thus, a hyperlink edge in the DAG starts at the latest revision of a hyperlinked page relative to the creation time of the relevant hyperlink and points to the first revision of the target page containing that hyperlink.

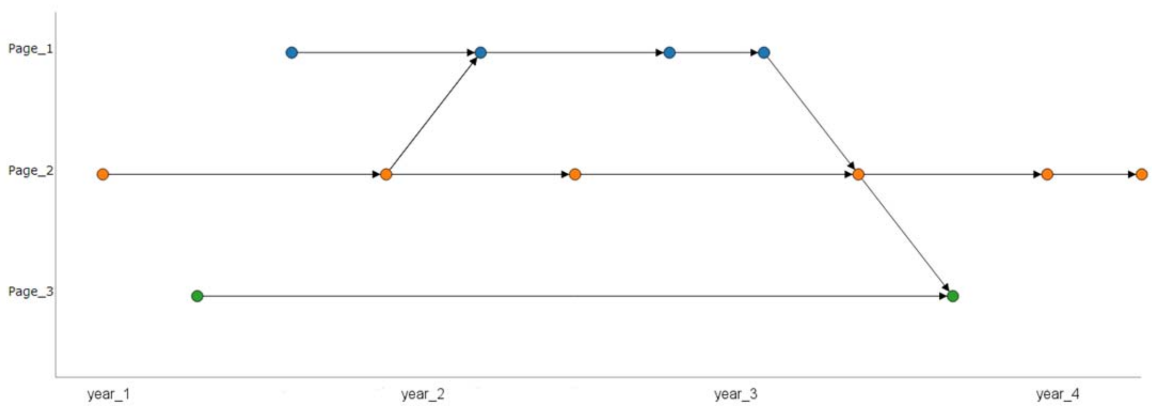

Figure 2. Swim lane diagram of a sample DAG of three articles with update and hyperlink edges

The described construction procedure results in a two-relational DAG that features update edges between revisions of a single page on the one hand and hyperlink edges between revisions of two related pages on the other hand. The procedure also guarantees that all update and all hyperlink edges are directed from a preceding revision to a succeeding revision in time. An example for such a DAG can be seen in Figure 2. In order to visualize the main paths of idea flows in a wiki, we use the visual metaphor of a "swim lane" diagram introduced in Figure 2. The page titles are shown in the left part of the diagram. All revisions of one page are represented as nodes connected by update edges and ordered in a horizontal line. The update edges of different pages are drawn parallel to one another, forming horizontal "swim lanes." Hyperlink edges between different pages are depicted as diagonal lines crossing the swim lanes. All edges point from left to right depicting the knowledge flow over time. Time is represented on the horizontal axis along the swim lanes. For any pair of nodes that belong to the same or to different pages, the node closer to the left represents the earlier of the two revisions. Node size reflects the traversal weight of a revision as calculated by the main path analysis. The more important a revision is within the paths of ideas, the larger the node is that represents it.

4.3 Results of the Main Path Analyses

Using the described method to build a DAG from wiki data, we analyzed the main paths in the two scientific domains *biology* and *electrical engineering* in Wikiversity. Both chosen categories represent well-developed domains in Wikiversity and serve as example datasets of different scales to illustrate our analysis method. Table 1 first gives a basic description of the two domains based on the revision logs in the dump.

Table 1. Descriptive characteristics of the studied domains

Domain	Pages in total	Pages on multiple main paths (90%)	Pages on main path	Edits in total	Edits on multiple main paths (90%)	Edits on main path	Authors in total	Authors on multiple main paths (90%)	Authors on main path
Biology	1268	58	8	9404	949	111	925	118	6
El. Engin.	398	34	6	4672	442	130	687	103	42

The three data blocks in Table 1 contain the number of pages, edits, and authors in the chosen Wikiversity categories. Each block shows the total count of each variable, as well as their distribution on the main path according to the SPC method (see section 3.2) and on the multiple main paths with 90 percent threshold (i.e., containing all nodes with a traversal weight above the 90th percentile).

Although the biology domain is much larger than electrical engineering in terms of page count, the latter domain is marked by a proportionally higher number of edits and authors. A clearly higher percentage of the pages in biology seem to be peripheral to the development of this domain. A similar number of authors in biology have produced roughly double the number of edits and pages on the multiple main paths in electrical engineering. This comparison reveals a higher average productivity of the authors on the multiple main paths in the biology domain. From the reverse point of view, this means that the multiple main paths in the biology domain were developed less collaboratively than those in the electrical engineering domain. Lastly, the main path in both domains is of similar length of edits and pages, but in electrical engineering, it is created by proportionally many more authors. Next, we present in detail the main path and the multiple main paths in both domains.

4.3.1 Main paths in the biology domain

The result of the main path analysis with the SPC method is depicted in Figure 3 as a swim lane diagram. The main path consists of pages from an online course on the applications of evolutionary principles held in 2009. The articles are well orchestrated, indicating a course syllabus of topics that build on one another. With only six contributors in total (see Table 1) and only two of them contributing more than two changes to the pages, the course represents a top-down approach to the design of instructional materials for a relatively passive group of learners. The revision logs reveal that the course materials did not initiate further development of the topic, as only three edits have been made since the second half of 2009, namely to the article on applications in physics (see Figure 3).

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

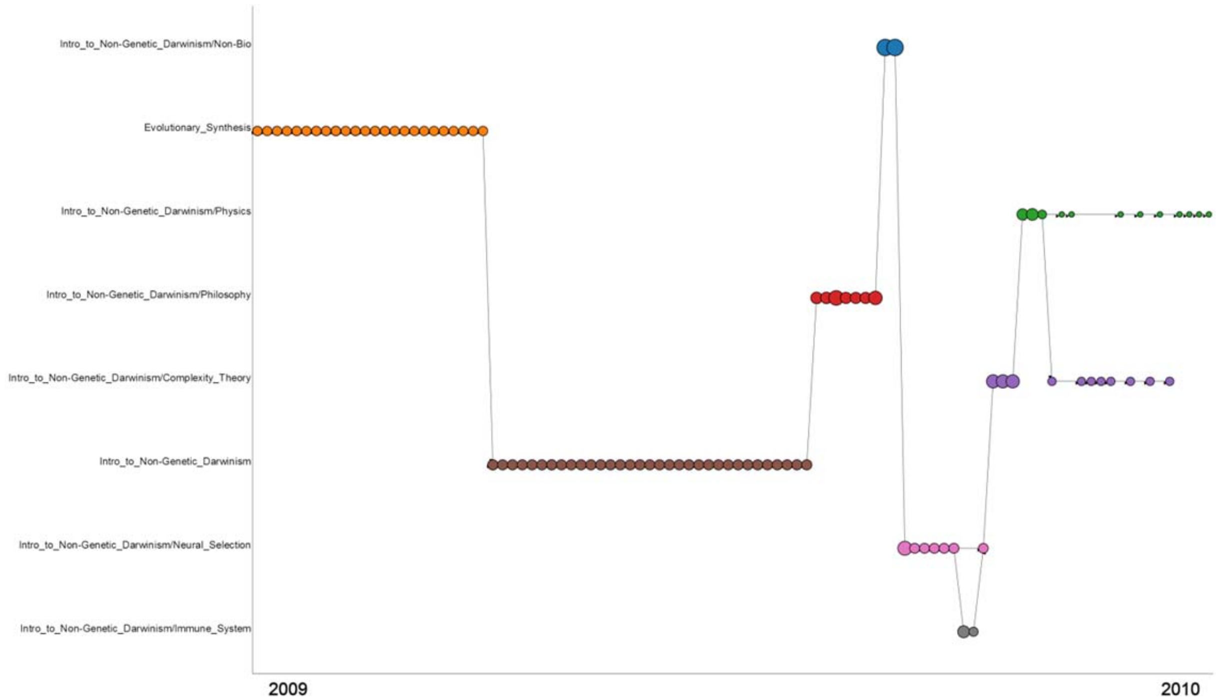

Figure 3. Simple main path in the biology domain

In order to broaden the range of important topics in the further analysis of the biology domain, we identified the multiple main paths as explained in section 3.2. Figure 4 shows the resulting swim lane diagram with additional branches of nodes and edges. Only ten percent (90th percentile threshold) of the article revisions with the highest traversal weight appear as part of the multiple main paths. Among them are all revisions presented as the main path in Figure 3.

Besides the discussed main path of the online course on evolutionary principles, several other topics appear as new separate branches: a cluster on sustainability and renewable energy from 2007 and 2008; two pages from a course on complex systems from 2011; an article about gynecological interviews gradually developed from 2007 to 2011; a small cluster on UFO research from 2006 and 2007; a larger and long-spanning cluster containing well-developed learning project pages about vitalism and consciousness, RNA interference, stem cells, life origins, human genetics, dominant group, and the connected basic biological concepts. Both branches containing the topics of vitalism and human genetics were first developed independently and later on flowed into the larger cluster. The main trajectory of that cluster starts with the topics RNA interference and cell improvement and ends with the topic dominant group.

The overall picture of the learning process in this domain suggests a heterogeneous evolution of ideas organized into separate topics. This conforms to the picture of groups of learners that followed different clearly defined interests in biology with little inter-group collaboration, except for the larger cluster of projects building on basic shared learning resources such as the general article on biology. The biology domain seems representative for the diverse and partly disconnected culture of online learning in the whole Wikiversity community.

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

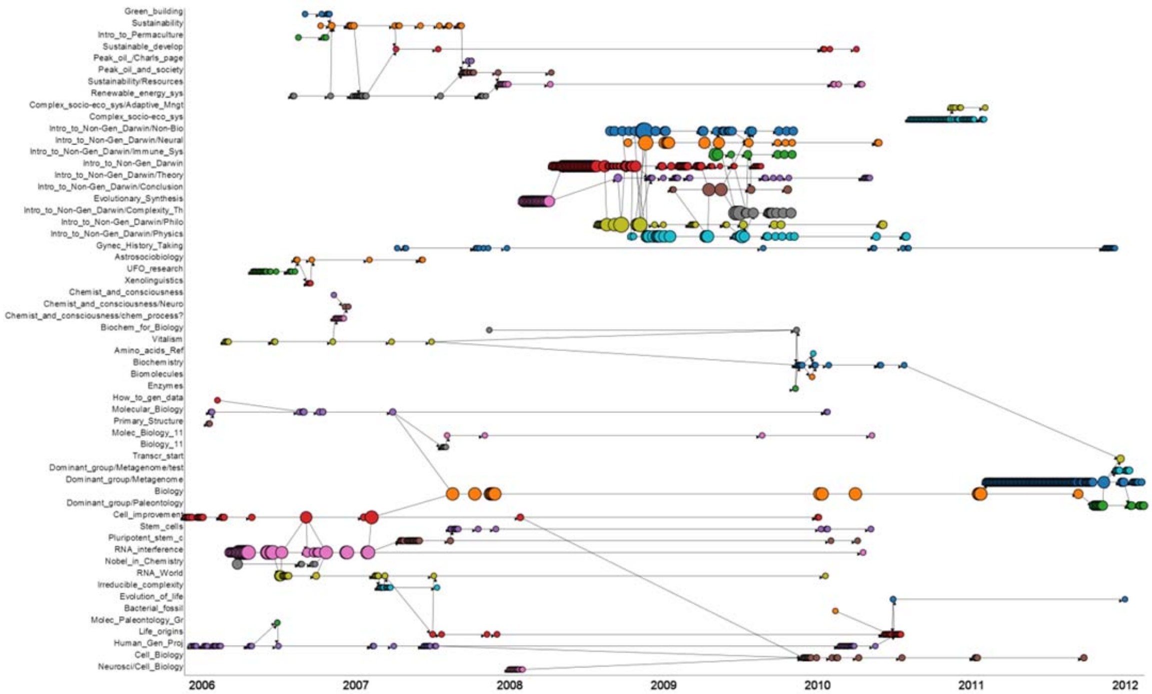

Figure 4. Multiple main paths in the biology domain

4.3.2 Main paths in the electrical engineering domain

Figure 5 shows the swim lane diagram of the SPC main path in the domain.

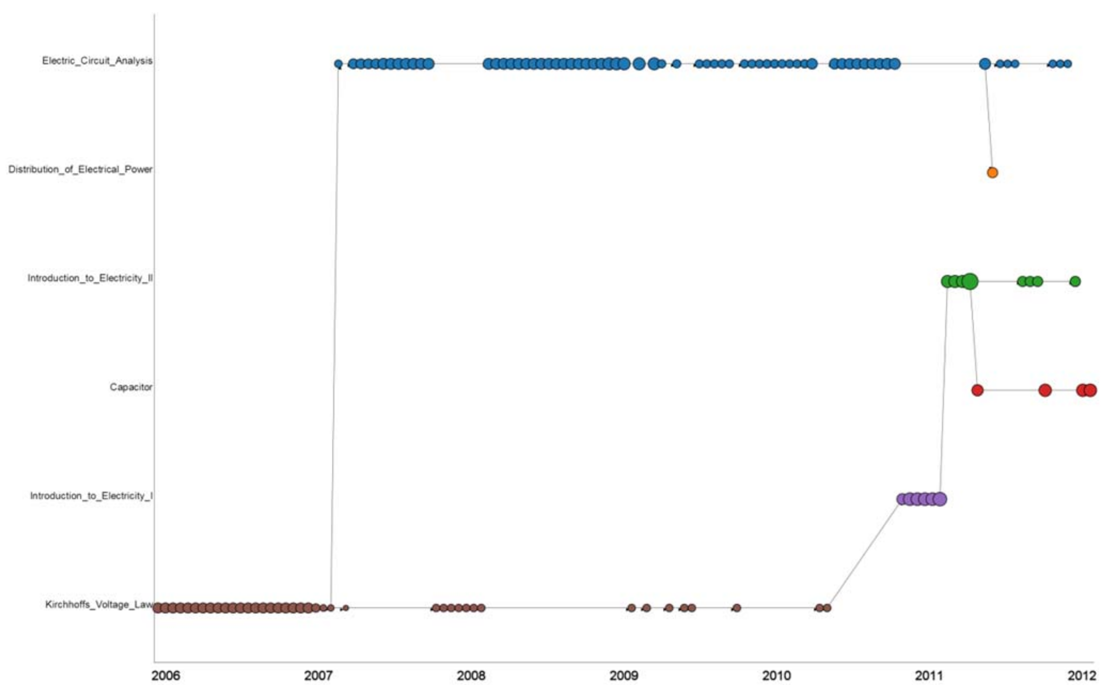

Figure 5. Simple main path in the electrical engineering domain

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

As with the main path in the biology domain, the core of the main path is the main page of an online course on electric circuits. In contrast to the course on evolutionary principles, this electric engineering course has been developed over a longer period from 2007 to 2009 and thus goes beyond the format of a course in the formal educational sense. The main path also contains an older resource from 2006 about voltage law that was later included in the course syllabus, as well as newer introductory resources on electricity from 2010 and 2011 that also referred to voltage law.

The interconnected and well-maintained articles indicate the core and narrowly interrelated topics in the domain. The creation of these core materials is an example of a truly collaborative learning process with many participating contributors (42 authors as shown in Table 1) over longer periods of time. The produced materials are structured as courses in order to facilitate any passive user encountering the topic, but the interesting learning process of the community of contributors is manifested in the collaborative creation of the study material itself.

As in the biology domain, we took a detailed look into the broader range of important topics in electrical engineering by analyzing the multiple main paths traced by ten percent of the article revisions with the highest traversal weight (90th percentile threshold). Figure 6 shows the resulting swim lane diagram that contains several new branches and additional nodes besides all the edits on the main path from Figure 5.

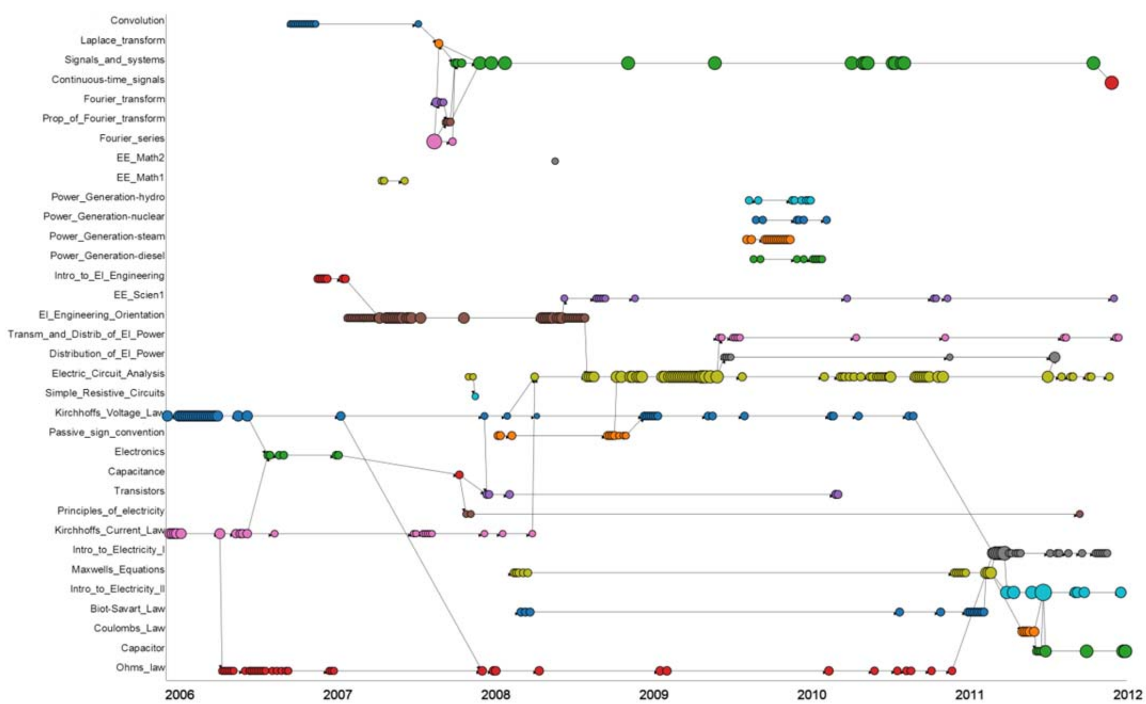

Figure 6. Multiple main paths in the electrical engineering domain

A new cluster of pages from 2006 and 2007 appearing now on the main paths covers the topic of signals and systems. The remaining separate pages of the main paths relate to mathematical tests and to a course on electrical power generation from 2008, all written by the same single author, as indicated by the revision logs.

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

The core cluster of the discussed main path now consists of many new articles covering basic electrical laws. On the main paths also appear pages from other topics structured as courses: on orientation to the domain and on transmission and distribution of electrical power. The important position in the DAG of the electric circuits topic in between the early orientation to electrical engineering and the later introduction to electricity explains why it is part of the main path in Figure 5. Although the enlarged core cluster consists of different courses and groups of topics, we found strong cross-participation of contributors across the pages in the cluster as we consulted the revision logs. In addition to the pages being thematically close, the cross-collaboration of authors presents an additional reason for the emergence of this large connected cluster.

Overall, our study showed that electrical engineering was a more compact and coherent domain than biology in the Wikiversity community. Many contributors collaborated over longer periods of time and a large number of pages, creating highly interrelated learning resources. Thus, materials organized as online courses were authored by a large number of people and serve general interests instead of that of a limited number of students for a limited period of time. The electrical engineering domain is an example of a self-organized learning community with enough time to build collaborative structures of practices and artifacts. Evaluated by main path analysis, the development resulted in more tightly interwoven topics than in the biology domain. Overall, the method revealed one large cluster of articles in both domains, as well as a few smaller ones, representing the core knowledge in those domains. This method allows for a subsequent analysis of the development of the topics over time and of the distribution of participation of their authors.

4.4 Author Profiles and Roles

After the overview of the main paths in the two domains, we turn to the analysis of the authors contributing to pages off as well as on the main paths. Here, we used the main path analysis results in combination with the revision logs in the dump. As explained in section 3.1, Wikiversity is an open virtual space and so there is no standard guideline on how authors should interact and use the environment. However, our data revealed differences in the contribution activity profiles of authors that can be interpreted in terms of a division of roles in the process of collaborative learning in a Wikiversity knowledge domain. We started by calculating for each author the number of edits and different edited pages and focused on the profiles of prominent authors who stood out among the large group of low contributors. Forty-six percent of the authors in the biology domain and 51% of the authors in the electrical engineering domain had minimal participation, just making a single edit without hyperlinks in the DAG. Respectively, 30% and 27% of the authors in the two domains who had at least one contribution on the multiple main paths did not make any other contribution. This highly skewed distribution of participation in online environments is a well-documented fact (Rafaeli & Ariel, 2008). More specifically, we see that authors who have a contribution on the main paths are generally less likely to make only a single contribution. According to this evidence, main path contributions can be interpreted to indicate high involvement in the community.

According to our interpretations of the profiles of active authors, we identified several categories of contributors: first, the role of *specialists*, who made many edits to only one or a few pages; second, the role of *maintainers* with a relatively high number of edited pages and a relatively low number of edits; third, the role of *leaders* with an outstanding number of edits and edited pages. As we show in the following, the interpretation of these roles was only accurate after taking the results of the main path analysis into account.

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

The investigated articles, and thus the contributions to them, are not of equal importance to the collaborative learning process of the community. Many articles are short stubs not interlinked with any other articles within the corresponding category. Such isolated and largely unimportant articles are not part of the main paths in a domain. Therefore, the results of the main path analyses in both domains of the study can enhance the analysis of the author roles by qualifying the number of contributions that lie on the main paths. As mentioned above, the SPC method of identifying a single main path leads to a strong focus on a small number of revisions and articles on a narrow topic. Hence, in this paper the author profiles are related to the extracted multiple main paths described in the previous subsections. Using the main path analysis in this way, a more adequate view on activity and division of roles of authors is achieved.

4.4.1 Author roles in biology

The three analyzed author roles in the biology domain are presented in the rows of Table 2 through the contribution profiles of distinctive sample authors. Each role is subdivided into type A and type B according to whether any of the contributions of an author are part of the main paths. The author activity in total and on the main paths is grouped in blocks containing the number of edits, edited pages, and edits with hyperlinks. As explained in section 3.2, hyperlinks represent knowledge flows between pages. Thus, the edits introducing a hyperlink and the edits referred through a hyperlink by another edit are important and should be regarded separately.

Table 2. Sample authors with a distinct role in the biology domain

Author profile	Author ID	Edits in total	Edits on multiple main paths	Pages in total	Pages on multiple main paths	Hyperlinked/ / hyperlinking edits	Edits with links on multiple main paths
Specialist A	278565	468	0	1	0	0 / 0	0 / 0
Specialist B	348476	10	10	1	1	0 / 0	0 / 0
Maintainer A	9357	35	0	31	0	0 / 0	0 / 0
Maintainer B	21778	43	9	41	8	0 / 1	0 / 0
Leader A	263421	1966	0	729	0	0 / 0	0 / 0
Leader B	20	552	154	112	20	31 / 35	25 / 20

The first rows, the specialist A with ID 278565 has the third highest number of edits in the domain, but these edits were all made to the same single page, moreover, none of them is part of the multiple main paths. This example shows that output quantity — the number of contributions — does not necessarily correspond to output quality — the importance for the evolution of discourse in a Wikiversity knowledge domain. The example of author 348476 adds to this finding. With ten edits in the domain in total, this is the most prolific author among the type B specialists — authors who are specialized in one single page and have at least one edit on the main paths. The low rate of activity of such specialists with important contributions would normally suggest that they should be regarded as low contributors. In the next rows, the type A and B maintainers 9357 and 21778 similarly show a low to middle rate of contribution. Maintainers mostly make small formal changes unrelated to the content of the edited Wikiversity pages. They correct spelling mistakes, organize the categorization, and sometimes also set hyperlinks, as does author 21778. Such authors typically contribute to very different domains and topics at the same time. Most of their contributions that appear on the main paths can be regarded as coincidental as they fall within a chain of important updates of the page content made by other authors. Table 2 further shows that the most prolific contributor and a type A leader in the biology domain, author 263421, didn't make a single important contribution on the main paths. A closer look into the data revealed that this author used Wikiversity to build a database on specific genes. This voluminous project was not much related to the other core topics in biology. Type B leaders, such as author 20, whose edits sometimes appear on the main paths, seem to play the most important role in the domain. Besides having the highest number of contributions on the main paths, this author also has the highest number of edits with hyperlinks. Further analyses of the data showed that authors with edits on the main paths tend to have more contributions and especially more interlinked edits than authors without edits on the main paths. Indeed, by the design of the method itself, hyperlinked and hyperlinking edits are more likely to occur on the main paths.

4.4.2 Author roles in electrical engineering

Table 3 presents the analysis of author roles in the electrical engineering domain following the structure of Table 2.

Table 3. Sample authors with a distinct role in the electrical engineering domain

Author profile	Author ID	Edits in total	Edits on multiple main paths	Pages in total	Pages on multiple main paths	Hyperlinked/ / hyperlinking edits	Edits with links on multiple main paths
Specialist A	858	44	0	1	0	0 / 0	0 / 0
Specialist B	292570	6	6	1	1	0 / 0	0 / 0
Maintainer A	3705	19	0	17	0	0 / 0	0 / 0
Maintainer B	8437	34	8	27	4	0 / 0	0 / 0
Leader A	32	245	0	75	0	1 / 0	0 / 0
Leader B	19038	867	114	133	14	20 / 35	8 / 8

As argued in section 3.3, the two domains are marked by a number of differences. Nevertheless, the studied author roles are identifiable in the same way in both domains, so the inferences about the authors in biology made in the previous subsection also apply for the authors in electrical engineering. The only difference worth mentioning is that author 19038, a type B leader in Table 3, has the highest number of contributions among all authors in the domain and at the same time has contributed the highest number of edits on the main paths. This case still corresponds to the conclusion that important

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

authors are distinguished not just by a high number of edits but also by significant contributions appearing on the main paths.

5. TECHNOLOGICAL IMPLEMENTATION

The analysis processes described in this paper have been integrated into our network analytics workbench (Göhnert, Harrer, Hecking, & Hoppe, 2013). A form of this workbench was used in the recent EU project “SISOB,”[4] which had the goal of measuring the influence of science on society based on the analysis of (social) networks of researchers and created artifacts. One area of research in this project was *knowledge sharing*. Thus the analysis techniques based on main path analysis presented in this paper were also of essential value in the project context.

We conceive workbenches as a general type of software environment designed to serve active and skilled users, without assuming the users to be computer experts. We have decided to develop a network analytics workbench as a web-based environment for several reasons, such as ease of deployment, access and update, and independence of the local computing facilities and devices. An important part of our experience with network analysis and network analysis tools is the need to combine several tools even for a single analysis process. The use of several tools sometimes also results in the need for conversion between the different data formats used by these tools. Therefore one important goal behind the development of the network analytics workbench is the integration of multiple tools and conversion mechanisms into one interface.

The workbench provides readily available processing chains for known use cases and furthermore allows for setting up new ones. The user interface (UI) is built upon a pipes-and-filters metaphor for processing chains in order to reduce the complexity of the underlying system for users who are not computer experts. An example of the UI that has been created using the WireIt[5] JavaScript library can be seen in Figure 7. In using the pipes-and-filters metaphor and being web-based, the workbench is similar to mashup projects like YAHOO pipes.[6]

In contrast to these projects, the actual processing of data in our workbench is not part of the user interface code itself but is done by a multi-agent system controlled by the workbench. The multi-agent system approach allows for combining several mostly independent tools into one workflow. These tools can be either pre-existing or newly developed. Examples of existing tools successfully integrated into the workbench are the network text analysis tool AutoMap (Diesner & Carley, 2005), the network analysis tool Pajek (Batagelj & Mrvar, 1998), and a wrapper for the R language.[7] Examples for newly developed components are a MediaWiki extraction component based on the mechanism presented in this paper and a main path analysis filter also used for the analyses presented in this paper. The communication between the web-based user interface and the agents is based on the SQLSpaces (Weinbrenner, Giemza, & Hoppe, 2007), an implementation of the tuple space architecture (Gelernter, 1985). From the user interface a description of the constructed workflow is posted into the SQLSpaces server, which contains a message for each agent (filter) type that is part of the workflow. These messages contain information about the input data and the parameter configuration of that filter.

[4] http://sisob.lcc.uma.es/
[5] http://neyric.github.com/wireit/docs/
[6] http://pipes.yahoo.com/pipes/
[7] http://www.r-project.org/

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

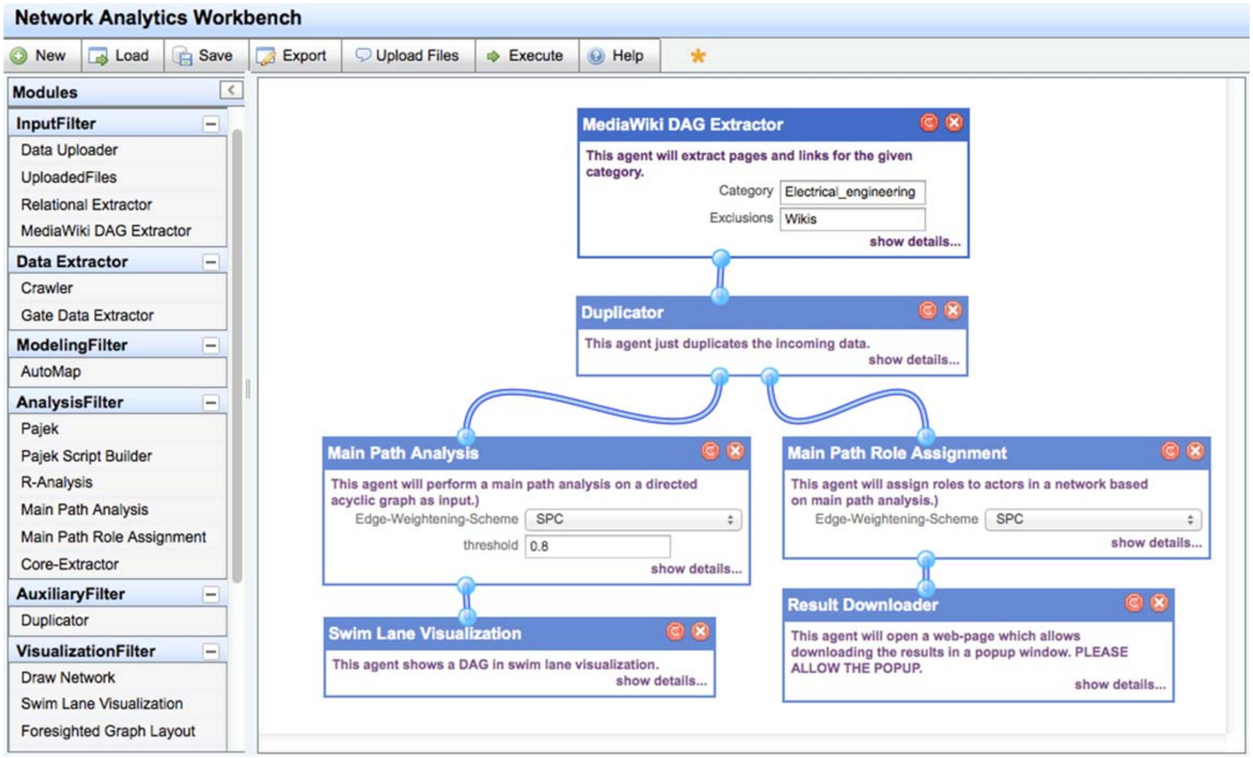

Figure 7. Screenshot of the Network Analytics Workbench

Figure 7 shows one of the workflows used for the analyses described in this paper. The first filter is used to provide input for the following filters. In this case, the filter connects to a MediaWiki database with Wikiversity data and creates a DAG for a given category from it. The extraction process follows the approach outlined in section 3.2 of this paper. The filter accepts two parameters. The name of the category for which the DAG should be extracted is a mandatory parameter. The second parameter accepts a list of categories to be excluded from the search and is optional. The next filter in the workflow presented here just duplicates all input into two parallel outputs. Thus, it allows for performing different analyses on the same possibly preprocessed input data in one workflow. In this example, the two outputs are used to perform main path analysis and analysis of author profiles in the same category of a wiki, as presented in this paper (sections 3.3 and 3.4). On the left side, the Main Path Analysis filter allows for selecting a weighting scheme to be used in the main path analysis and for defining a threshold for the multiple main path analysis. The results of this filter are then visualized using the swim lane metaphor also used throughout this paper. The other branch of the duplicator leads into the Main Path Role Assignment filter, which generates the tables used for the author profile analysis as described in section 3.4. These tables are then fed into the Result Downloader, which allows for downloading these results onto the local machine for further usage.

6. CONCLUSION

With the help of the main path analysis, we detected the core topics in the two Wikiversity domains of biology and electrical engineering. While biology had much broader scope, the collaboration of the authors was weaker. The resulting main paths had a similar size and structure to the main paths in electrical engineering, which was a small coherent domain with a relatively large group of authors and a

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

higher necessity for collaboration. Thus, the small ratio of main path versus other articles in the biology domain compared to the electrical engineering domain could be explained through differences in the level of collaboration among the authors revealed by the revision logs.

The exemplary results of the presented empirical study may be useful for the Wikiversity community as a whole. As it seems, some scientific domains like biology might benefit from strengthening of collaboration. Additional analyses may be helpful to choose appropriate directions for development, but our results point to the need for better coordination of the disparate topics in this domain. The main path analysis can also orient participants by showing them the importance of the topic they are working on. It can also reveal important reference points to other core topics in the field. A beginning contributor can be aided by a presentation of the main paths with the decision to add to an existing strand of knowledge development or to start a new peripheral one. An advanced participant in the community may benefit from the analysis as a historical reconstruction of the shared knowledge-building process, in order to compare his or her own visions and goals with the actual knowledge development of the community and to discover topical gaps necessitating further efforts. With some additional work to adapt and standardize the analysis and the necessary interventions relative to the specific goals within an educational context, the main path analysis can be used to support and even take the load off a teacher or coordinator of knowledge building.

Our approach presented in this paper is the first application of scientometric methodology for analyzing the flow of ideas in the context of an open learning wiki environment. Using the examples of the biology and electrical engineering domains in Wikiversity, we showed how main path analysis can be employed to analyze the collaborative creation of various knowledge artifacts and the learning processes of the online community. Our methods have been embedded into a web-based analytics workbench that supports the definition and re-use of analysis modules in a user-friendly visual environment.

The paper presented a procedure for creating directed acyclic graphs from wiki data and for illustrating the obtained main idea flows in swim lane diagrams. Our visualization technique allows for a unified view of knowledge flows in a network of artifacts with multiple relationships. The main path analysis results were helpful in understanding the differences in the collaborative structure of two scientific domains in Wikiversity. The results further facilitated the characterization of different roles that authors have in the community. We found that the total rate of contribution was not a sufficient criterion for identifying the most important authors in a domain. But, as the role of maintainers demonstrates, some contributions on the main paths may also not testify to the importance of an author. Instead, the total number of contributions should be evaluated in combination with the number of contributions that appear on the main paths.

For our future work, we plan to elaborate on the characterization of contributions and contributors with respect to the main paths of development in other educational knowledge-building scenarios. It appears promising to provide moderators, teachers, tutors, or the productive teams themselves with results of such analyses, in order to support reflective practices (Schön, 1983). This will raise further challenges regarding visualization and cognitive ergonomics.

ACKNOWLEDGMENTS

The work presented in this paper was funded by the ScienceCampus Tuebingen and by the European Community under the Science in Society (SIS) theme of the 7th Framework Programme for R&D (Grant agreement 266588). This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content. Last but not least, we thank Alfredo Ramos for his help with visualizing the studied main paths.

REFERENCES

Aviv, R., Erlich, Z., Ravid, G., & Geva, A. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal for Asynchronous Learning Networks, 7*, 1–23.

Batagelj, V. (2003). Efficient algorithms for citation network analysis. arXiv: Computer Science. http://arxiv.org/abs/cs/0309023

Batagelj, V., & Mrvar, A. (1998). Pajek: A program for large network analysis. *Connections, 21*, 47–58.

Bereiter, C., & Scardamalia, M. (2003). Learning to work creatively with knowledge. In E. De Corte, L. Verschaffel, N. Entwistle, & J. van Merriënboer (Eds.), *Powerful learning environments: Unravelling basic components and dimensions* (pp. 73–78). Oxford, UK: Elsevier Science.

Bonk, C., & Cunningham, D. (1998). Searching for learner-centered, constructivist, and sociocultural components of collaborative educational learning tools. In C. J. Bonk & K. S. King (Eds.), *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship, and discourse* (pp. 25–50). Mahwah, NJ: Erlbaum.

Cakir, M., Xhafa, F., Zhou, N., Stahl, G. (2005). Thread-based analysis of patterns of collaborative interaction in chat. Proceedings from: *International Conference on Artificial Intelligence in Education* (pp. 120–127). 18–22 July. Amsterdam, The Netherlands.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning*, *3*, 105–122.

de Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning, 2*, 87–103.

Diesner, J., & Carley, K. (2005). Revealing social structure from texts: Meta-matrix text analysis as a novel method for network text analysis. In V. K. Naraynan & D. J. Armstrong (Eds.), *Causal mapping for information systems and technology research: Approaches, advances, and illustrations* (pp. 81–108). Harrisburg, PA: Idea Group Publishing.

Gelernter, D. (1985). Generative communication in Linda. *ACM Transactions on Programming Languages and Systems (TOPLAS), 7*(1), 80–112.

Göhnert, T., Harrer, A., Hecking, T., & Hoppe, H. U. (2013). A workbench to construct and re-use network analysis workflows: Concept, implementation, and example case. Proceedings from ASONAM 2013: *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 1464–1466). Niagara Falls, Canada, 25–28 August. New York: ACM Press.

Halatchliyski, I., Kimmerle, J., & Cress, U. (2011). Divergent and convergent knowledge processes on Wikipedia. Proceedings from CSCL 2011: *Connecting Computer-Supported Collaborative Learning to Policy and Practice* (Vol. 2, pp. 566–570). 4–8 July. Hong Kong, China: International Society of the Learning Sciences.

Halatchliyski, I., Moskaliuk, J., Kimmerle, J., & Cress, U. (in press). Explaining authors' contribution to pivotal artifacts during mass collaboration in the Wikipedia's knowledge base. *International Journal of Computer-Supported Collaborative Learning.*

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

Halatchliyski, I., Oeberst, A., Bientzle, M., Bokhorst, F., & van Aalst, J. (2012). Unraveling idea development in discourse trajectories. Proceedings from ICLS 2012: *The Future of Learning: 10th International Conference of the Learning Sciences* (Vol. 2, pp. 162–166). 2–6 July. Sydney, Australia: International Society of the Learning Sciences.

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science, 28*, 115–152.

Harrer, A., Malzahn, N., Zeini, S., & Hoppe, H. U. (2007). Combining social network analysis with semantic relations to support the evolution of a scientific community. Proceedings from CSCL 2007: *Mice, Minds, and Society: The Computer Supported Collaborative Learning Conference* (pp. 267–276). New Brunswick, NJ, 16–21 July. Mahwah, NJ: Lawrence Erlbaum Associates.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 16569–16572.

Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks, 11*, 39–63.

Jeong, A. C. (2003). The sequential analysis of group interaction and critical thinking in online. *The American Journal of Distance Education, 17*, 25–43.

Knorr-Cetina, K. (2001). Objectual practice. In T. R. Schatzki, K. Knorr-Cetina, & E. von Savigny (Eds.), *The practice turn in contemporary theory* (pp. 175–188). London/New York: Routledge.

Liu, J. S., & Lu, L. Y. (2012). An integrated approach for main path analysis: Development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology, 63*, 528–542.

Mika, P. (2007). *Social networks and the semantic Web*. New York: Springer.

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, *21*, 759–772.

Popper, K. R. (1968). Epistemology without a knowing subject. *Studies in Logic and the Foundations of Mathematics, 52*, 333–373.

Rafaeli, S., & Ariel, Y. (2008). Online motivational factors: Incentives for participation and contribution in Wikipedia. In A. Barak (Ed.), *Psychological aspects of cyberspace: Theory, research, applications* (pp. 243–267). Cambridge, UK: Cambridge University Press.

Reffay, C., & Chanier, T. (2002). Social network analysis used for modelling collaboration in distance learning groups. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *ITS 2002 Lecture Notes in Computer Science, 2363* (pp. 31–40). Berlin, Heidelberg: Springer.

Reinhardt, W., Moi, M., & Varlemann, T. (2009). Artefact-actor networks as tie between social networks and artefact networks. Proceedings from CollaborateCom 2009: *5th International Conference on Collaborative Computing: Networking, Applications and Worksharing* (pp. 1–10). Washington DC, 11–14 November. IEEE Computer Society.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences, 3*, 265–283.

Schön, D. (1983). *The reflective practitioner: How professionals think in action.* London: Temple Smith.

(2014). Analyzing the Main Paths of Knowledge Evolution and Contributor Roles in an Open Learning Community. *Journal of Learning Analytics*, *1* (2), 71–93.

Schrire, S. (2004). Interaction and cognition in asynchronous computer conferencing. *Instructional Science, 32*, 475–502.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences*, Cambridge, UK: Cambridge University Press.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning, 5*, 5–42.

Suthers, D., & Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. In P. Long, G. Siemens, G. Conole, & D. Gasevic (Eds.), *Proceedings of the 1st international conference on learning analytics and knowledge, LAK’11* (pp. 64-74). Banff, Alberta, Canada, 27 February–1 March. New York: ACM Press.

Suthers, D., & Verbert, K. (2013). Learning analytics as a “middle space”. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the 3rd international conference on learning analytics and knowledge, LAK'13* (pp. 1-4). Leuven, Belgium, 8–12 April. New York: ACM Press.

Wasserman, S., & Faust, K. (1994). *Social networks analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Weinbrenner, S., Giemza, A., & Hoppe H. U. (2007). Engineering heterogeneous distributed learning environments using tuple spaces as an architectural platform. Proceedings from ICALT 2007: *7th IEEE International Conference on Advanced Learning Technologies* (pp. 434–436). Niigata, Japan, 18–20 July. IEEE Computer Society.

Zaïane, O., & Luo, J. (2001). Web usage mining for a better web-based learning environment. Proceedings from: *Conference on Advanced Technology for Education* (pp. 60–64). 27–28 June. Banff, Alberta, Canada.

Zeini, S., Göhnert, T., & Hoppe, H. U. (2012). The impact of measurement time on subgroup detection in online communities. Proceedings from ASONAM 2012: *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 389–394). Istanbul, Turkey, 26–29 August. IEEE Computer Society.

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum

Carlos Monroy, Virginia Snodgrass Rangel, Reid Whitaker
Center for Digital Learning and Scholarship
Rice University, USA
carlos.monroy@rice.edu

Abstract: In this paper, we discuss a scalable approach for integrating learning analytics into an online K–12 science curriculum. A description of the curriculum and the underlying pedagogical framework is followed by a discussion of the challenges to be tackled as part of this integration. We include examples of data visualization based on teacher usage data along with a methodology for examining an inquiry-based science program. With more than one million students and fifty thousand teachers using the curriculum, a massive and rich dataset is continuously updated. This repository depicts teacher and student usage, and offers exciting opportunities to leverage data to improve both teaching and learning. In this paper, we use data from a medium-sized school district, comprising 53 schools, 1,026 teachers, and nearly one-third of a million curriculum visits during the 2012–2013 school year. This growing dataset also poses technical challenges such as data storage, complex aggregation, and analyses with broader implications for pedagogy, big data, and learning.

Keywords: Learning analytics, science education, online curriculum, K–12

1. INTRODUCTION

Computers, tablets, and mobile devices offer new tools to support learning, both inside and outside the classroom. This development, in turn, has created new opportunities to understand and assess their impact on educators and learners through the collection and analysis of user interaction data. Indeed, a major advantage of online curricula over traditional textbook-based curricula is the ability to capture these user interaction or learning analytics data. Learning analytics is a relatively new, but rapidly growing area of research (First International Conference on Learning Analytics and Knowledge, 2011), whose goal, according to the Society of Learning Analytics, is "understanding and optimizing learning and the environments in which it occurs." Most learning analytics (LA) methods use data generated by the interactions between learners and the Learning Management Systems (LMSs) they use (Pahl, 2004).

Ongoing research in educational data mining and LA suggests that harnessing analytics data can advance new research methodologies in education, help educators better assess their pedagogical practices, and devise innovative educational methods, all with the goal of improving education (Elias, 2011; Campbell & Oblinger, 2007; Pardo & Delgado, 2011; Siemens, 2012). LA can leverage improvement through improved educational decision-making (Arnold, 2010; Norris, Baer, Leonard, Pugliese, & Lefrere, 2008; Vatrapu, Tplovs, Fujita, & Bull, 2011), clearer institutional and individual goal setting (Hendricks, Plantz, & Pritchard, 2008), more timely and frequent feedback for students and teachers (Ha, Bae, & Park, 2000; Hamalainen, Suhonen, Sutinen, & Toivonen, 2004; Merceron & Yacef, 2005; Suthers, Ravi, Medina, Joseph, & Dwyer, 2008), the individualization of teaching and learning (Beck & Mostow, 2008; Farzan, 2004; Heraud, France, & Mille, 2004; Lu, 2004; Talavera & Gaudioso, 2004), and the generation of a richer set of data on student behaviour and learning (Baker & Yacef, 2009; Mazza & Dimitrova, 2004; Pavlik, Cen, & Koedinger, 2009).

With the rapidly growing interest in and technical ability to leverage these data in educational settings, there also is a growing sense that many recent educational technology and big data initiatives are detached from what we know about teaching and learning, particularly in the K–12 sector (Junco, 2012; Romero & Ventura, 2010). While the use of LA has been particularly fruitful in higher education and in online courses where many to all learning activities and interactions take place online, in conventional brick and mortar schools, it remains crucial to incorporate information derived from the other activities that go on in a classroom in order to capture fully the variety of learning experiences taking place. This is particularly important in science education where most experts agree that students need to be "doing" science: conducting experiments, making observations, and building models (Crawford, 2000; Haury, 1993; Keys & Bryan, 2001). Without data on hands-on activities and classroom discussions, it can be hard to make sense of the LA data gathered by an LMS. It is, therefore, imperative to ensure that the use of LA in the K–12 sector is grounded in relevant pedagogies and serves as a complement to data on face-to-face interactive learning experiences not captured by digital artifacts such as student inquiry, classroom discussions, and hands-on activities. This comprehensive, mixed methods approach ensures that data analyses in LA are not reduced to the mere analysis of clicks and page visits.

Furthermore, LA must be leveraged in ways that both recognize and draw on existing K–12 education research. While LA has an established presence in the business sector and even in higher education, its foray into K–12 education is particularly new and is being driven in large part by the start-up community and venture capital, and less by educators or educational researchers (Culatta, 2012; DeSantis, 2012). Although there is nothing wrong with these drivers, educators are unlikely to buy into new technology and big data initiatives if they are not immediately accessible and relevant to their daily work (Means, Padilla, & Gallagher, 2010; Watters, 2012; Wayman, Cho, & Richards, 2010).

This paper expands our ongoing research (Monroy, Snodgrass Rangel, & Whitaker, 2013) and addresses this gap in the literature by outlining and discussing one strategy for incorporating LA into a broader investigation of how a K–12 online science curriculum is implemented in classrooms. The purpose of our implementation study is to increase teacher and student use of the curriculum through improved design. The strategy is meant to answer the following three key questions about the curriculum:

1. How can we incorporate the LA data into a broader effort to understand how a digital curriculum is used?
2. How can we leverage the LA data to support teachers and administrators?
3. How can we integrate the LA data into the evaluation of the curriculum's impact?

We have developed a three-pronged strategy to address these guiding questions: a *technical* prong related to the architecture supporting the collection and management of LA data, a *quantitative* prong related to how the LA data can be utilized to understand the curriculum's use and improve its design, and a final *qualitative* prong where we collect and analyze qualitative data on teacher (and in the future, student) use of the curriculum.

We begin our discussion by briefly describing the curriculum examined here and its underlying pedagogical model. The technical architecture is explained in section three. In section four, we describe

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

several specific techniques we have developed to use the LA data to measure use of the curriculum and provide examples using real data from a school district in Texas. Here, we also discuss several considerations related to the normalization of data. In section five, we describe ways to use data-visualization techniques to support teachers and administrators. Then, in section six, we situate our use of LA data within our broader data strategy for measuring the implementation of the curriculum. In order to demonstrate how this might be accomplished, we provide examples from real data collected in the fall of 2012 from a second district in Texas. Finally, we end with a discussion about the future of data analytics in K–12 education, focusing in particular on the challenges we need to overcome and the potential for educational improvement.

2. BACKGROUND: STEMSCOPES AND THE 5E+I/A PEDAGOGICAL FRAMEWORK

STEMscopes is a comprehensive online K–12 science curriculum aligned to Texas state standards and grounded in the 5E science inquiry model. The curriculum offers a range of hands-on inquiry activities, online activities, and additional supporting resources for teachers for each of the learning standards at each grade level. STEMscopes currently serves about 2,400 schools, more than 50,000 teachers, and over one million students in Texas. These figures represent approximately 40 percent of the state's school districts, making it the most used science curriculum in the state. During the 2012–2013 school year, we gathered more than one hundred million data points generated by districts, schools, and users, both teachers and students. We were able to identify the science standard covered, the grade level, specific steps in the 5E+I/A step, materials accessed, and the weekly assessment results.

The 5E pedagogical model is an inquiry-based learning approach that aims to harness learners' prior knowledge while eliciting interest in new phenomena (Bybee et al., 2006). This instructional model was created and refined by the Biological Science Curriculum Study (BSCS). The 5E acronym stands for the following five steps that comprise a full lesson cycle: engage, explore, explain, elaborate, and evaluate. In the Engage phase, students draw on prior knowledge to raise their interest in and activate their prior knowledge of the new content. During the Explore phase, students take part in activities and experiments that allow them to experience and learn new concepts and skills through active investigations. The Explain phase requires students to explain those new concepts and skills in their own words. New experiences in the Elaborate phase deepen student understanding of the new concepts. Finally, learner understanding is assessed in the Evaluate phase. Teachers are meant to implement lessons following this cycle in order to optimize learning.

The STEMscopes curriculum adds two steps to the basic 5E model: one for intervention and one for acceleration. Intervention activities are designed to support those students still struggling to master the standards, while the acceleration activities are designed for those students who have mastered the concepts and are ready to extend and apply their learning (Whitaker, 2012; Zuiker & Whitaker, 2013). With these two additional steps, the model becomes the 5E+I/A model.

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

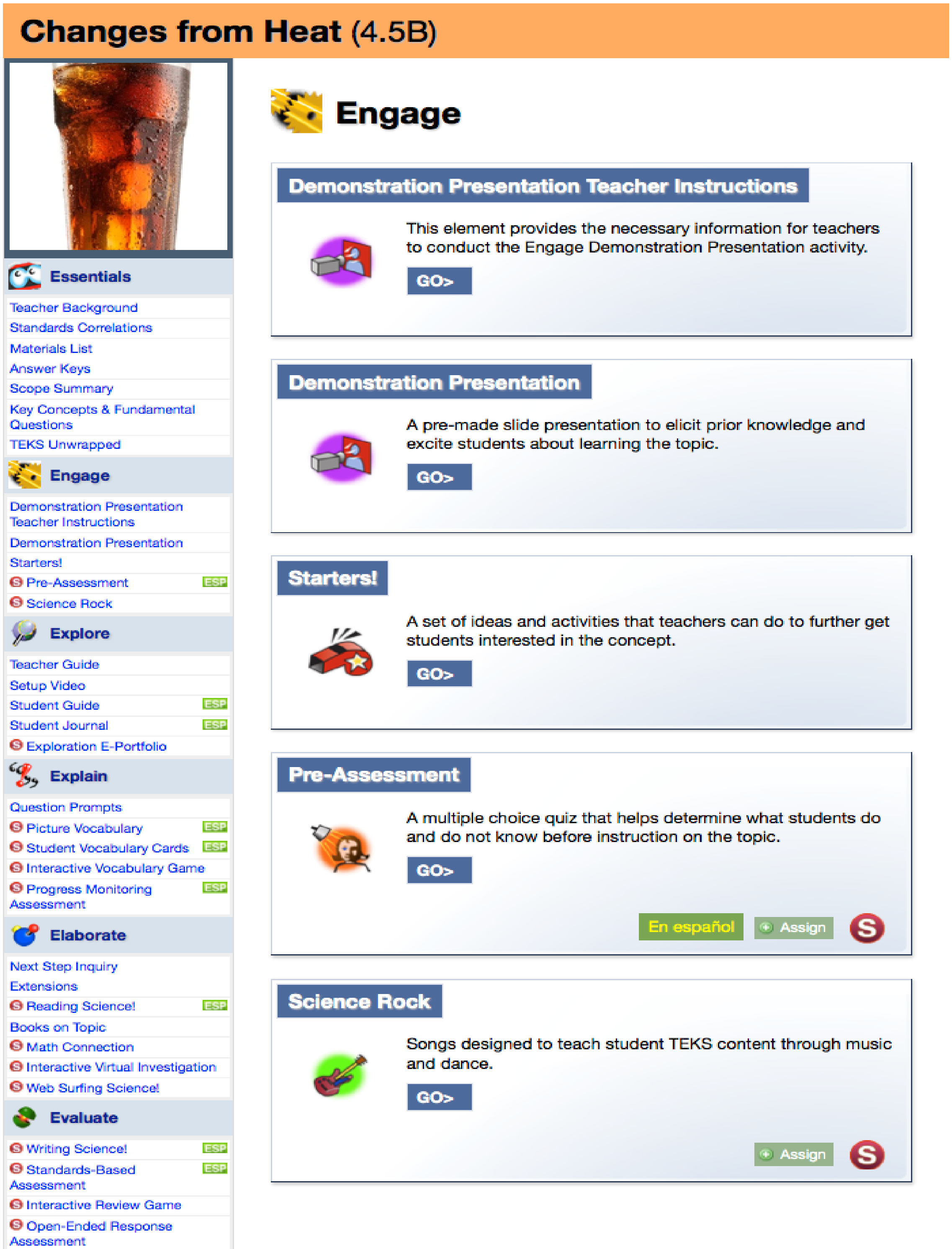

Figure 1: Screen shot of a science unit titled, "Changes from Heat." The column on the left presents elements grouped by each 5E phase (Engage, Explore, Explain, Elaborate, and Evaluate).

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

In addition to the online activities and materials, the curriculum also offers accompanying kits that contain the science materials necessary to do every hands-on activity in the curriculum. The kits allow students to conduct science experiments, build models, create simulations, and make observations, thereby offering them a deeper learning experience meant to re-create the work of scientists (Haury, 1993; Minner, Levy, & Century, 2010). In the Texas version of the curriculum[1], there are nearly four hundred standards or "scopes" in total across all grade levels. Figure 1 provides an example of what the 5E+I/A model looks like in the curriculum. Specifically, it depicts the 4th grade standard, "Changes from Heat." Components or activities for each of the seven (5E+I/A) steps are presented in the column on the left. They include materials for teachers such as background information on the learning standard, interactive games, and assessments.

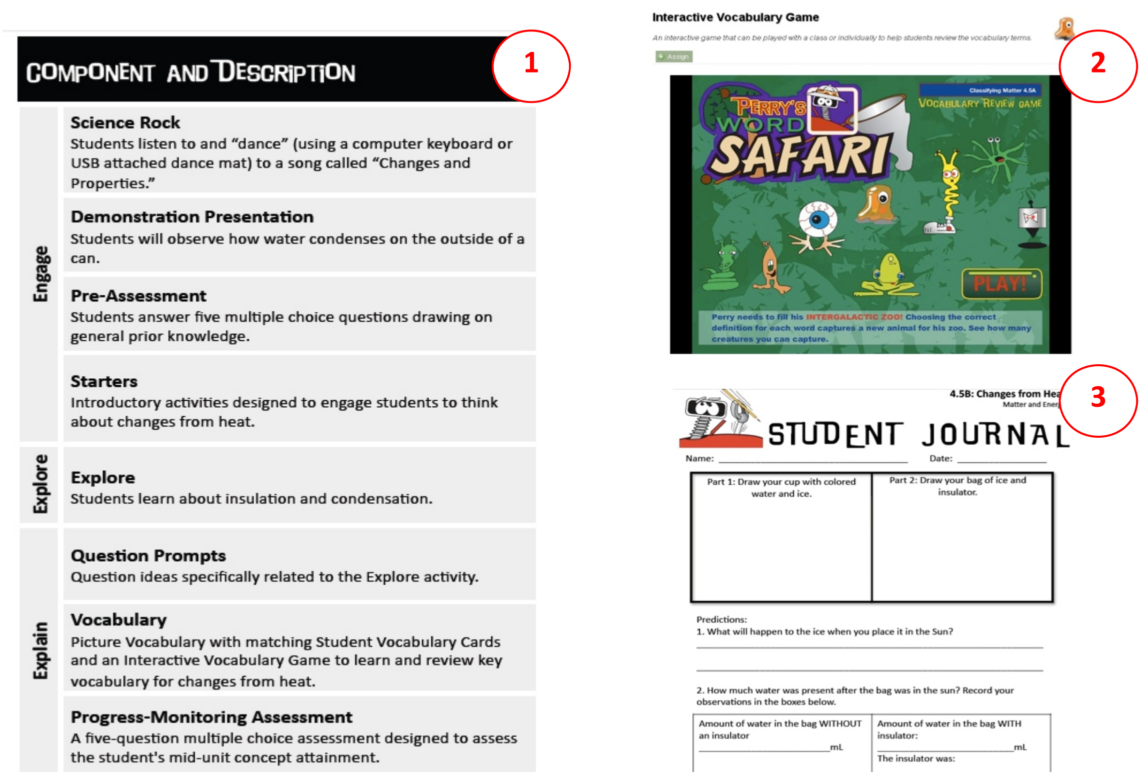

Figure 2. Various elements for one science standard: 1) a partial summary of the content of one unit, 2) an interactive game, and 3) a student journal.

Figure 2 shows three elements within a scope: 1) an overview of a unit's content in the context of the 5E+I/A model, which provides a snapshot of all the components a teacher can find for that scope and is intended for class planning purposes; 2) an interactive game in the Explain phase that offers an engaging

[1] There also is a national version created to help teachers implement the Next Generation Science Standards (NGSS), but the NGSS version does not follow the 5E+I/A model.

opportunity for students to practice new content vocabulary; and 3) a student journal sheet that can be used by students when completing hands-on activities as part of the Explore step.

3. AN ARCHITECTURE FOR SUPPORTING LEARNING ANALYTICS

As an online curriculum, STEMscopes creates an environment where data is constantly generated and stored as teachers and students use the curriculum on a daily basis. For this reason, the architecture enables the curation, storage, and processing of that large amount of data. As an example, during the 2012–2013 school year, nearly a quarter of a million interactions with the curriculum were captured each day. This volume of data demands techniques from the areas of databases, data warehouses, and big data. We have designed a distributed data warehouse environment that allows us to scale the storage and processing of user-generated data.

To serve our storage and computing needs, we employ four separate high-end computer servers from a cloud-computing provider. Figure 3 depicts the architecture and interaction among the main servers. Each one of these servers handles specific types of data and requests. The Production Server (PS) is an Nginx web-server, which processes web-client requests via a Ruby on Rails application. Account information such as user names, passwords, and account expiration dates, as well as curriculum taxonomy and state science standards are stored in the Production Database Server (PDBS). Anonymized events data generated by users interacting with the curriculum are stored in the Analytics Storage Server. These events are objects created by the Rails application on the Production Server (PS) and transmitted as encrypted packages via an HTTPS (Hyper Text Transfer Protocol Secure) session. At the end of the day, an automatic process harvests all events generated the previous day (stored in the Analytics Storage Server) and loads them into the Analytics Processing Server (APS). Once the data are loaded, a series of scripts are triggered, which calculate our analytics variables and aggregate results by user, school, and district levels. These results include information about visits by science standard, 5E+I/A step, and curriculum element used. Similarly, the analytics variables are calculated at the end of each week, month, and year. Upon completion, the calculated data are transferred to the Dashboard Data Server (DDS).

Although the current infrastructure is robust enough for our immediate needs, as we introduce more sophisticated analyses, such as machine learning, data mining algorithms, simulations, and computational linguistics, we envision the use of new methodologies in the realm of big data (Jagadish, 2012; Pavlo, Paulson, & Rasin, 2009). We recently began experimenting with Hadoop (Venner, 2009) and the Map Reduce paradigm, an open source framework that enables distributed high-intensive computing for processing extremely large amounts of data. For instance, a process employing optimized queries and accessing pre-processed aggregated data for calculating twenty-five analytics variables for all teachers during the 2012–2013 school year takes approximately eight days to complete. Thus, we need to plan a scalable infrastructure as we begin calculating content- and user-based models derived from data spanning several years, across multiple schools, districts, and states along with recent user-generated data. Among the practical benefits of this big data approach to our K–12 curriculum are to allow the use of large clustering and machine learning algorithms for supporting personalized learning in the future.

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

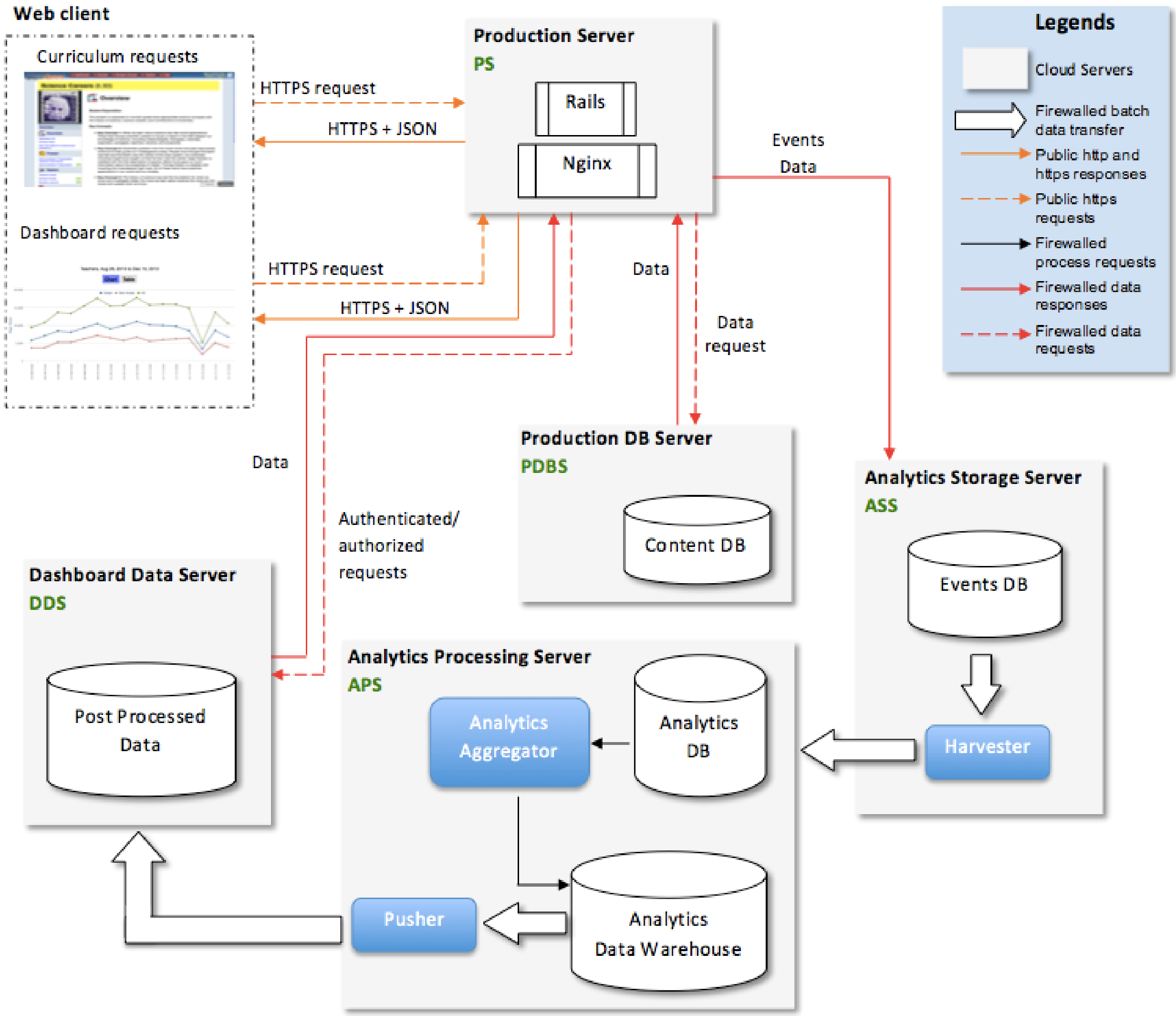

Figure 3. Curriculum architecture and processes for storing, processing, and transferring data

4. LEARNING ANALYTICS TECHNIQUES FOR MEASURING CURRICULUM USE

The first research question that we posed is how researchers can utilize LA data to understand how and how much teachers and students use the curriculum. This information is useful both to the curriculum developers as well as to researchers investigating the curriculum's effectiveness. In this section, we focus on the first. The 5E+I/A inquiry model described above guides the analysis of all data collected because without this pedagogical grounding, the data patterns at best have no meaning, and at worst can be attributed the wrong meaning (Atkinson & Wiley, 2011; Long & Siemens, 2011). The model's underlying theory of learning, constructivism, is also used to answer important questions about curriculum implementation and how students learn best (Bybee et al., 2006). For example, constructivism and the 5E model assume that students learn best by exploring concepts before they are

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

presented with formal explanations of the content. An important task, then, is to examine whether teachers and students appear to use the steps and corresponding activities in this intended order.

STEMscopes, however, also was created with the understanding that teachers want flexibility in how they teach. In practice, this means that a teacher can access and use any part of the lesson cycle when they deem most appropriate for their students, and most teachers do adapt the lesson cycle and the activities contained within them to their classroom needs (e.g., using components in a different order, changing them, or skipping some of them; Cho & Wayman, 2014; Hew & Brush, 2007; Keengwe, Onchwari, & Wachira, 2008; Koehler & Mishra, 2005; Lawless & Pellegrino, 2007; Leonardi, 2009; Orlikowski, 2000). This recognition creates a fundamental challenge for the curriculum designers who also believe that the structure of the 5E+I/A lesson cycle is the best learning path.

This paradox creates a need to understand precisely what teachers and students are doing in their classrooms, and what these patterns of use mean for curriculum development and improvement. We have developed a strategy to complement a more traditional approach to measuring use or implementation with LA data (primarily analyzing teacher use). In the rest of the section, we describe two LA measures that examine the frequency and the ways in which teachers access the curriculum, and three ways we normalize the data to compare across teachers, schools, etc., and to conduct other analyses.

All data were generated at the teacher level and come from a medium-sized school district in Texas during the 2012–2013 school year.[2] We selected this district because of an ongoing evaluation agreement and because it is one of the top districts in terms of curriculum usage. No identifiable information about district, school, or teacher was used. The original sample included 1,026 users and 53 schools. We discarded 18 users since seven were categorized as administrators and 11 did not have any visits to the actual content. We also removed 6 schools (5 high schools and one alternative education) because they were only piloting the curriculum. The final sample included 36 elementary and 11 middle schools with a total of 898 and 88 teachers, respectively.

4.1 A Pedagogy-Focused Approach to Measuring Use

It is not enough to know how often the curriculum was visited. Because the curriculum was designed using a specific pedagogical model, the 5E+I/A lesson model, our analyses must try to uncover to what extent teacher use of the curriculum follows the model. Though there are multiple ways to measure how teachers use the curriculum with LA data, here we discuss two: the proportion of visits to each of the 5E+I/A steps and the proportion of visits that are "inquiry-focused."

First, for each teacher the total number of visits to individual steps in the 5E+I/A framework was calculated. Teachers were then grouped according to their most visited step. The chart on the left in Figure 4 shows the percentage of teachers for whom each step was the most used. For instance, Explore was the most visited step for 360 (40%) elementary teachers and 13 (14%) middle school teachers; Engage was the most visited step for 242 (26%) elementary teachers, etc. Conversely, Acceleration was

[2] We have found that student use of the curriculum is relatively low, which we assume is due to the limited number of activities that students can complete online, as well as limited school access to computers and other digital devices.

the most visited for only 3 percent of elementary teachers and about 1 percent of middle school teachers. The chart on the right depicts the percentage of visits for each step. Explore was the most visited step for elementary schools teachers with 25 percent of the visits, followed by Engage with 22 percent. For middle schools, the most visited step was Explain (21%) followed by Explore and Elaborate (18% each). What these charts mean is that, overall, most visits by elementary teachers are to the hands-on activities of the curriculum. In middle school, the same is true — the most visited step is Explore. Interestingly, middle school teachers access the intervention activities substantially more than do elementary teachers.

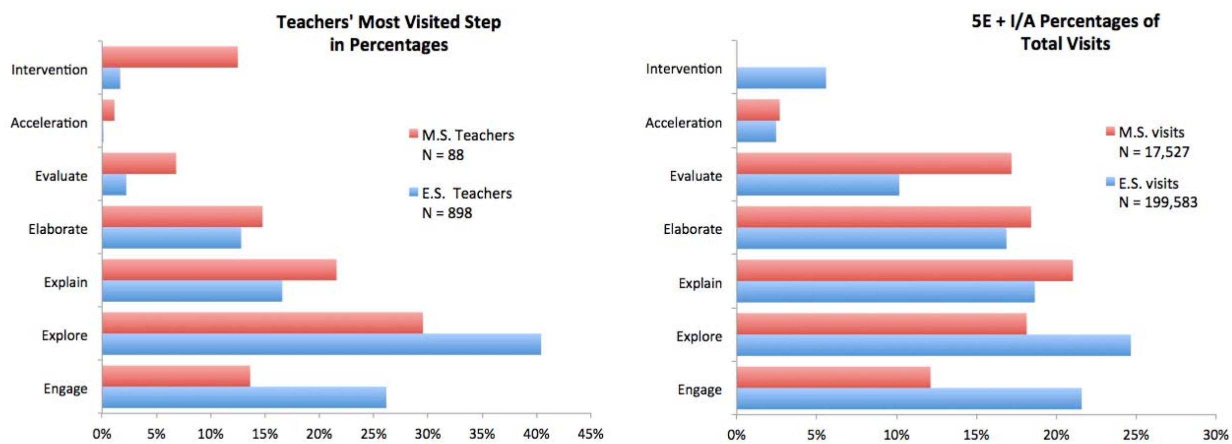

Figure 4: Percentage of teachers for each one of the most used steps (left). Percentage of visits to each 5E+I/A step (right)

Examining teacher use of the curriculum is quite complex and has continually required us to refine what we want to know, usually leading to more nuanced questions. Even the most basic question about frequency of use has many possible answers that require different approaches for calculating variables. For instance, we can capture overall use for each teacher who has an account, but this tells us nothing about how frequently they use specific components; it might simply be that they sign in and out. Therefore, we also calculated how frequently teachers use individual units (there is one unit for each state standard), different steps of the model, and other components within the curriculum (e.g., teacher background material). We also have drilled down further by weighting these statistics by the number of grades a teacher teaches or the number of science topics covered. In this way, we have begun to deal with questions of data normalization, which we will discuss more below.

Questions of time introduce additional complexity. The analytics can tell us how long a teacher is logged in, when the first and last logins were, how long a teacher spends on each page, and on how many different days a teacher actually accesses the website. These can help normalize data as well as give us additional information about how teachers use the curriculum. For example, teacher activity for a sample of teachers in the district in our study throughout the 2012–2013 school year showed a great deal of variation in actual use. These data are summarized in Table 1 below. For elementary schools (36 schools and 898 teachers), on average, teachers visited the curriculum 329 times, with a minimum of 2 and maximum of 4,784 visits; the number of days on which teachers actually log in to access curriculum content averages 30 (minimum=1 and maximum=175 days).

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

We also found variation in the time span during which teachers actively use their curriculum accounts (from first login to last login) for the period of interest. The range was from one day to 277 days, with an average of 181 days. We calculated the percent of the total time (in days) that teachers were actively logged into the website; this statistic ranged from 0% to 100% of the time with an average of 22%. We also adjusted this figure to take into account the number of grade levels the teachers teach; some teach only one grade, while others teach science to multiple grades. Finally, on average teachers visited a science standard 18 times, while the average of visits by grade level was 199 times. Results for middle schools are shown in the bottom part of Table 1.

Table 1. Teacher data on visits to STEMscopes

	Total visits	Number of different days teachers logged in	Time span for visits (in days)	Total time span weighted by number of unique days teacher logged in (%)	Number of science topics taught	Number of grade levels taught	Avg. visits by science topic	Avg. visits by grade level
Elementary School Teachers N = 898								
Mean	329	30	181	0.22	15	2	18	199
Median	191	21	212	0.14	13	1	15	128
Minimum	2	1	1	0.01	1	1	2	2
Maximum	4,784	175	277	1.00	161	9	102	1,672
Middle School Teachers N = 88								
Mean	302	22	161	0.26	16	2	15	166
Median	122	11	194	0.13	13	2	11	53
Minimum	3	1	1	0.01	1	1	2	2
Maximum	3,897	159	277	1.00	90	9	169	3,897

The user analytics variables ranged substantially across schools. Among the main reasons for variation in the number of visits, we identified four: number of standards taught, number of grades taught, longevity of the accounts, and number of different days visited. Longevity represents the number of days elapsed since the user account was created. Therefore, older accounts would tend to have more visits, while newer accounts fewer visits. In the following sections we discuss the effect of these variables on the number of visits to the curriculum.

$$IIC = \frac{\sum(En + Ex + Ep)}{\sum V}$$

Formula 1

There are many ways to measure use from a pedagogical perspective, and we have developed various indices that allow us to conduct richer analyses of teacher use of the curriculum. A comprehensive description and use of these indices, however, is beyond the scope of this paper. The Inquiry Instruction Contribution Index (IIC Index, depicted in Formula 1 below) is one way to measure use. To illustrate our approach, here we describe the application of the IIC to the total number of visits to the curriculum:

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

Specifically, the IIC measures the proportional contribution of the three most "inquiry-based" 5E steps of the lesson cycle (Engage, Explore, and Explain) to the overall curriculum usage (V) for each science standard. This index is important for pedagogical reasons: constructivism and the 5E model favour learning through hands-on, student-centred exploration, and we therefore have an interest in knowing what the proportion of teacher visits are to the more inquiry-based steps and activities. The index provides a summary statistic describing the level of inquiry in the classroom: a figure closer to 0 would suggest relatively little inquiry, whereas a figure closer to 1 would suggest more inquiry focus in a classroom.

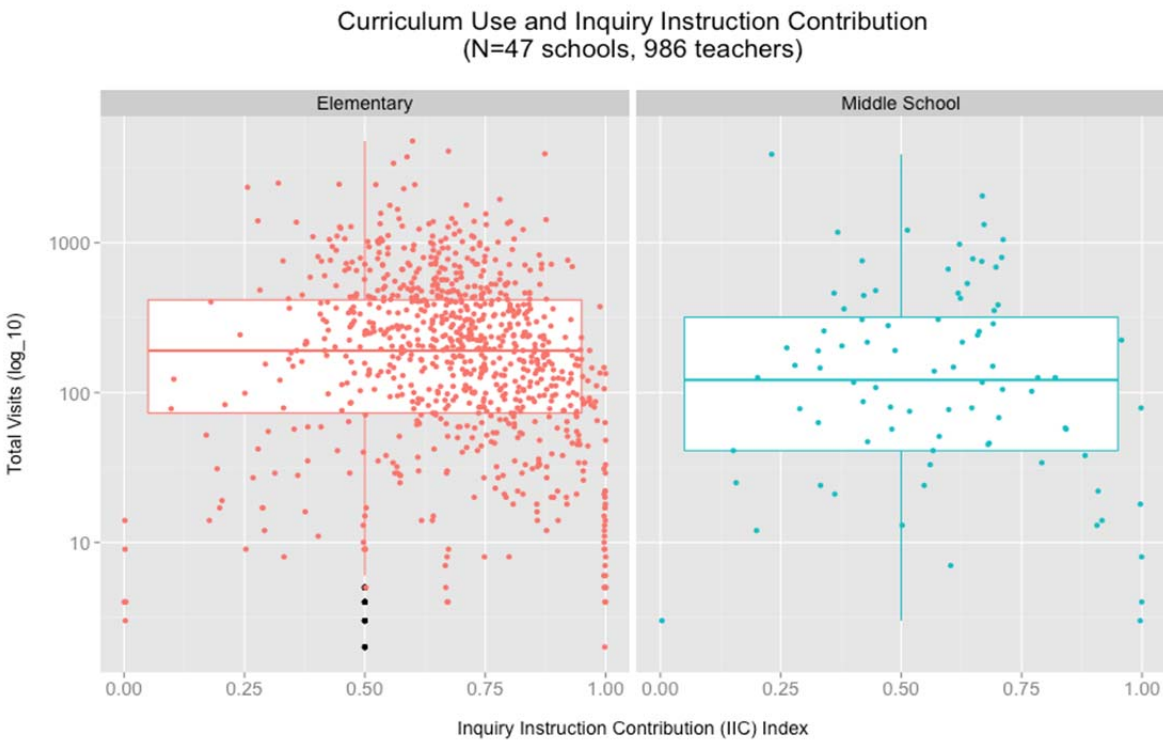

Figure 5: Boxplots depicting teacher visits to the curriculum (Y axis in log 10 scale) and Inquiry Instruction Contribution Index (X axis).

Although it is difficult to determine a "perfect" use dosage, our curriculum design team suggests that between two-thirds and three-quarters of the total curriculum visits should be dedicated to the Engage, Explore, and Explain steps. In Figure 5, we compare curriculum visits (Y-axis) with the IIC index (X-axis) for each teacher across elementary (N=36 schools, 898 teachers) and middle schools (N=11 schools, 88 teachers). The comparison indicates that the level of the inquiry focus varies according to school level.

Table 2 shows the distribution of total visits and IIC for teachers in both elementary and middle schools. Elementary schools in this district have the highest level of total use with an average of 329 and maximum of 4,784, and most of that use clusters in the higher range of the IIC index, which suggests both that the elementary teachers visit the Engage, Explore, and Explain steps a fair amount, and that there might be a relationship between higher use of the curriculum and a more inquiry-focused classroom. In contrast, the middle school teachers have lower use with an average of 302 and maximum

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

of 3,897, and their IICs are more evenly distributed, that is, the frequency of teacher visits does not appear to be related to the inquiry focus of a teacher's classroom.

Table 2. Distribution of total visits and IIC

	Elementary School Teachers		Middle School Teachers	
	Total visits	IIC	Total Visits	IIC
Minimum	2	0.00	3	0.00
1st quartile	73	0.58	41	0.42
Median	191	0.71	122	0.60
Mean	329	0.69	302	0.58
3rd quartile	416	0.83	319	0.70
Maximum	4784	1.00	3897	1.00

Broken down this way, the IIC data start to paint a picture of how teachers use the curriculum. These data might then be used to revise the middle and high school portions of the curriculum to see where improvements might be made to the more inquiry-based steps and activities. On the other hand, these data might be used to identify needs for professional development and coaching that can be embedded into the curriculum (e.g., middle and high school teachers may need support in figuring out how to integrate more hands-on activities into their lessons).

4.2 Number of science standards taught

In addition to creating LA-based measures of curriculum use, we must also ensure that the data are comparable across different units and levels of analysis. One important consideration is the number of science standards taught. Within their accounts, teachers can access all state standards throughout the school year. Because different districts and even schools may have their own curriculum scope and sequence, it is necessary to examine overall teacher use while taking into consideration the number of standards that have been taught. Without this form of normalization, a teacher in a district with one scope and sequence may appear to use the curriculum more than a teacher working under a different scope and sequence.

Figure 6 illustrates this relationship. The total number of teacher visits is on the Y-axis (in log 10 scale) and the total number of science standards taught is on the X-axis. Each point in the graph corresponds to a teacher and the dotted line is the mean number of visits. The graph points to a weak but positive relationship: as the number of science standards accessed increases, so does the number of the overall visits. Without taking this relationship into consideration, we might have a biased understanding of curriculum use.

4.3 Grade levels accessed

A second variable that influences the total number of teacher visits to the curriculum is the number of grade levels that a teacher accesses. We have found in our research that teachers access standards from more than one grade level for different reasons: some teach science to more than one grade level, while others pull activities from previous grade levels as a review, and still other teachers share accounts. Unfortunately, it is only possible to know why teachers access different grade levels if you ask them directly through a survey, interview, or focus group. It is possible, however, to control for the number of grade levels accessed as another way to normalize overall use: without this control, a teacher account that accesses numerous grade levels might have more visits to the curriculum, even though the teacher may not teach more science. Figure 7 depicts the weak but positive relationship between total visits (Y-axis, in log 10 scale) and number of grades taught (X-axis). Each point represents one teacher.

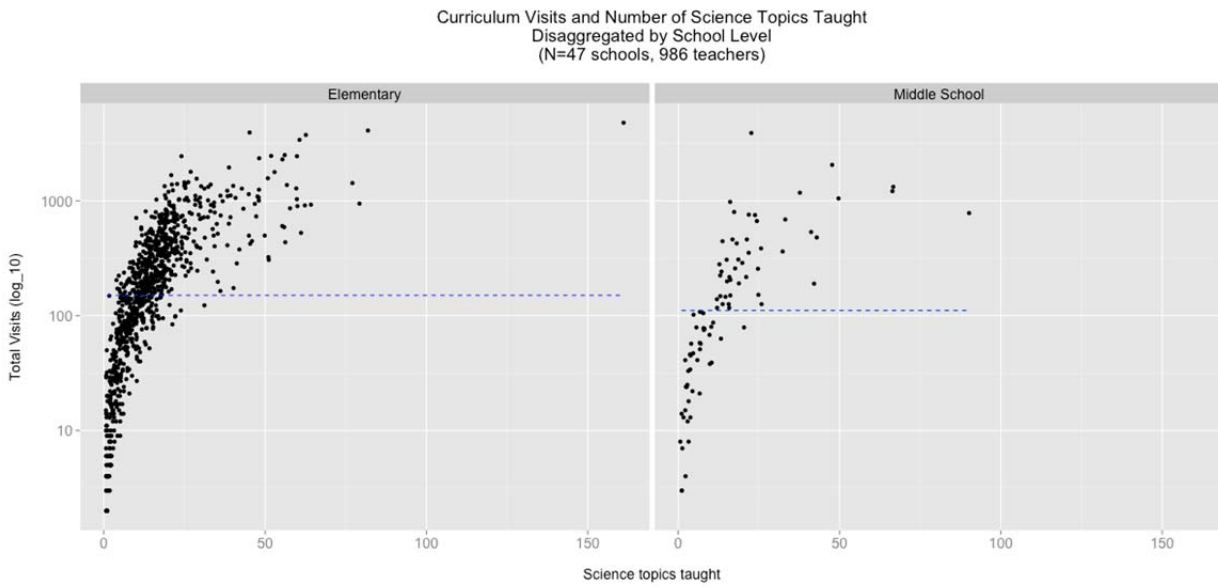

Figure 6. Total number of visits to the curriculum and number of science standards taught.

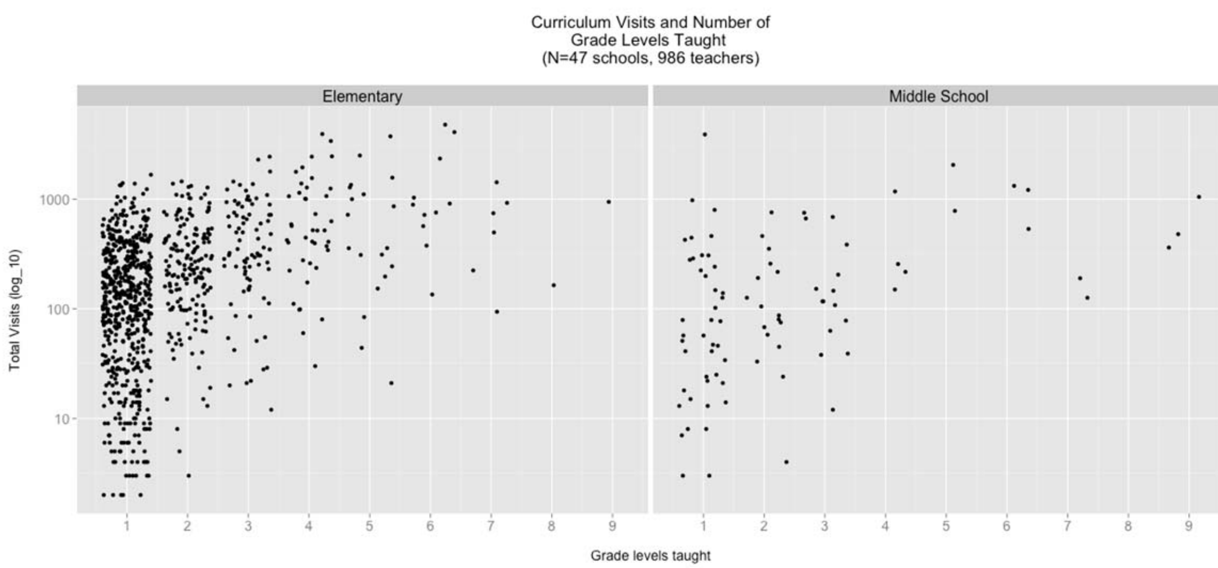

Figure 7. Total number of visits to the curriculum and number of grade levels taught.

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

4.4 Longevity of accounts

A third consideration for data normalization is the longevity of teacher accounts. Longevity refers to the number of days elapsed between the creation of an account and a second given date (e.g., the end of the school year, the last day of use, etc.). Teachers do not always activate their online accounts at the same time, which we hypothesized would affect the total number of visits to the curriculum: teachers with "older" accounts should have a higher number of overall visits, but this might not mean that they teach more science. Figure 8 shows number of visits on the Y-axis in log 10 scale and longevity of the account in days on the X-axis as of the first day of the school year, disaggregated by school level. Values closer to the left are newer accounts (negative values indicate days elapsed since the beginning of the school year and the date when the account was created), whereas values to the right correspond to older accounts (positive values indicate the number of days that accounts have been active). Most of the schools show two clusters of account longevity, one around 0 days and the other between 550 and 600 days. These clusters represent the creation of new accounts at the beginning of each school year. The plots do not show defined trend lines and therefore do not suggest a correlation between use and longevity. However, elementary schools created more new accounts once the school year had started than middle schools did.

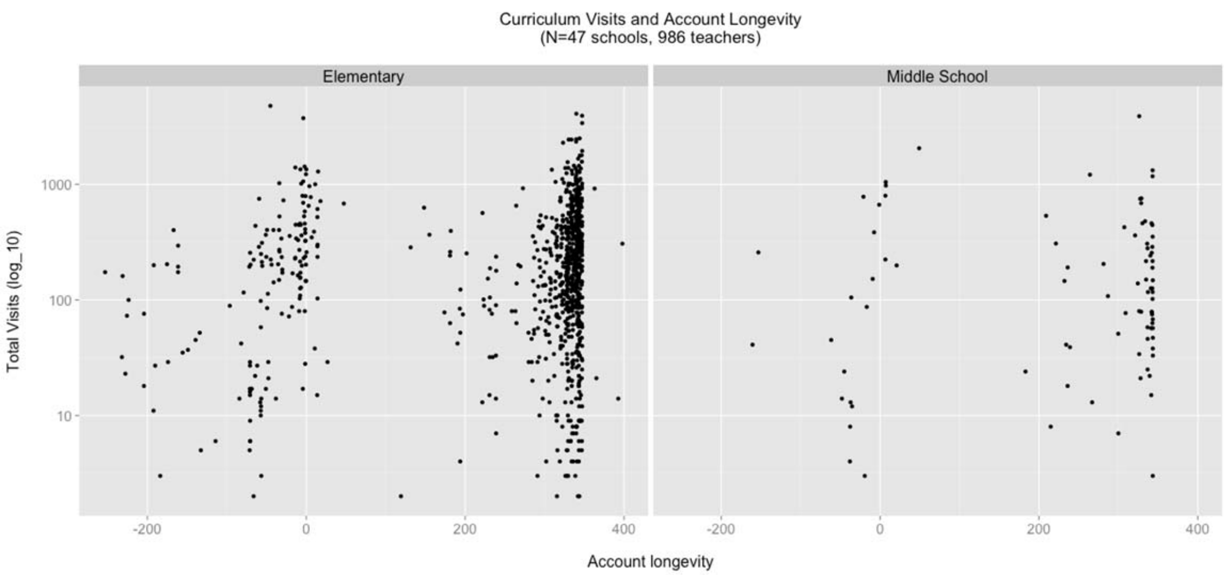

Figure 8. Total visits to the curriculum and account longevity.

4.5 Proportion of days visited

Finally, another way to think about teacher use of the curriculum is to measure the time during which they have accessed it actively. To do this, we created a variable that represents the number of unique days a teacher has accessed the curriculum divided by the teacher's total log-in time span in days (calculated as the difference, in days, between the first and last day a teacher logged in). Values closer to 1.0 indicate a teacher who has logged in virtually every day; a value of 0.5 indicates a teacher who has logged in nearly half of the time they had an active account; and values closer to 0.0 indicate a teacher who seldom logged in. For example, a teacher who uses the curriculum on a daily basis will have more unique days than one who accesses it once a week. It is important to note that this measure does not take into account the amount of content accessed within a single visit. In other words, it is possible that one teacher with more unique visits than another has accessed less content during each visit such that the two teachers have accessed the same amount of content, or such that the teacher with fewer unique visits actually has accessed more content.

Figure 9 depicts the relationship between the total visits and the proportion of days visited. Total visits on the Y-axis are in log 10 scale, and the proportion of days visited on the X-axis ranges from 0.0 to 1.0. The elementary school graph shows two clusters, the first of which depicts a growing trend in total visits as the proportion of days visited increases (between 0 and 0.5). The second cluster shows a small number of teachers with 100% proportion but very low total visits (approximately less than 75 visits for the entire school year). Middle schools show a similar pattern, with the exception that the proportion of visits ranged approximately between 0.0 and 0.35.

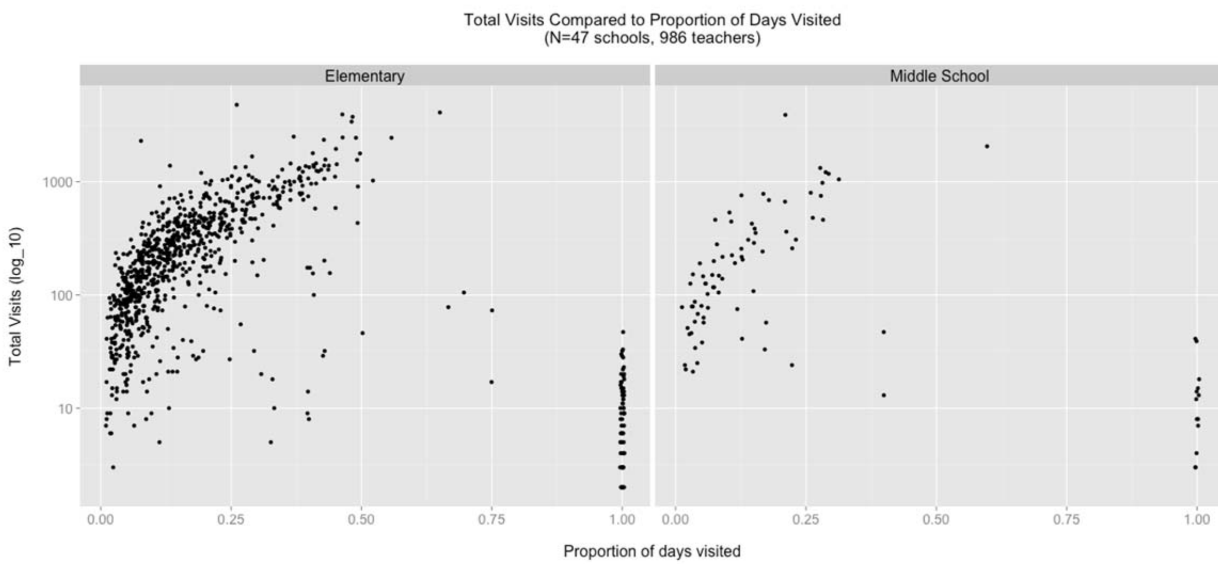

Figure 9: Total visits to the curriculum and proportion of days visited

This kind of data helps curriculum designers understand if and how often the curriculum is being used; what content is most and least used; and the dosage of 5E+I/A steps they accessed. In future, we will also be able to relate these data to student learning outcomes, both those measured by assessments embedded in the curriculum (which students can take online) and by district and state standardized assessments. For school and district administrators, this information can shed light on understanding teaching practices and design professional development to identify best teaching practices and address deficiencies in instruction.

5. LEARNING ANALYTICS AND DATA VISUALIZATION TO SUPPORT TEACHERS AND ADMINISTRATORS

In addition to using the LA data to help curriculum developers, we also process the LA data in ways to support the end users of the curriculum — teachers and administrators, primarily. In our research on curriculum use, we have learned that use varies depending on the role of the end user. For example, teachers were more interested in their own use and their students' progress, while campus-level leaders were interested in trends across classes, and district-level administrators wanted to know about use and progress across campuses (Snodgrass Rangel, Monroy, Bell, & Whitaker, 2013). Research also tells us, however, that educators often struggle to turn data into actionable information (Heer & Shneiderman, 2012; Ingram, Seashore Louis, & Schroeder, 2004; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Lachat & Smith, 2005; Wayman, 2005; Wayman et al., 2010).

Data visualization is an important tool that, in concert with strong theories, can help make sense of data (Duval, 2011). It is key, however, that the design of these visual interfaces be driven by teacher and administrator feedback. For example, several teachers with whom we have worked indicated that knowing what parts of the curriculum they had not yet used would help them know what they still needed to teach. In response, we created a heat map where colour intensity depicts the most and least used sections, with the goal of driving teacher attention to those sections. We have developed other data visualization technique, but covering all of them is beyond the scope of this paper. Here we describe three of them: timelines, heat maps, and a mastery tracker.

5.1 Timelines

Time-based visualizations can show when materials are accessed and for how long, as well as shifts in what science standards teachers access during a particular period of time. In this vein, timelines can be used to identify what teachers are teaching, when, and for how long. For example, one could see whether the life science standards are accessed before the physical science standards. The timelines can also shed light on the period of time, or duration, during which a teacher accesses a particular standard regularly. Pacing suggests the time elapsed between the completion of one standard and the beginning of a new one, which can help to identify standards that teachers may be overlapping or teaching at the same time (i.e., reviewing a set of standards while also introducing new standards). In the following paragraphs, we discuss two examples of the use of timelines.

The duration of instruction for the science topic entitled "3.5AB- Classifying Matter" for a sample of teachers in one school is depicted in Figure 10. Each horizontal bar corresponds to one teacher and bars are sorted by the first day they were used. The position on the X-axis represents the first time a teacher accessed that science topic, the length of the bar indicates the duration in days. The graph shows that "Classifying Matter" was introduced early in the school year and most of the teachers accessed it for about three to four weeks. There are also a few long horizontal bars (spanning almost to the end of the school year), this could be an indication of teachers reviewing certain materials after those science topics had been originally covered.

Another way to examine timelines is to explore how a teacher covered science topics throughout a given period of time. Figure 11 shows all science topics taught by one teacher. The Y-axis corresponds to science topics and the X-axis to time in days. This teacher covered a total of 23 science topics, of which 21 belong to 3rd grade and two to 2nd grade, respectively. This proportion most likely suggests a 3rd grade teacher accessing content from a lower grade for reinforcement purposes. Bar lengths can be clustered in three groups. The longest bars (five science topics), show duration of nearly the whole school year. A second cluster (three bars), shows duration between three and four months. A third cluster (five bars), shows duration between three and six weeks. The final cluster (nine bars) shows durations of approximately one week or less.

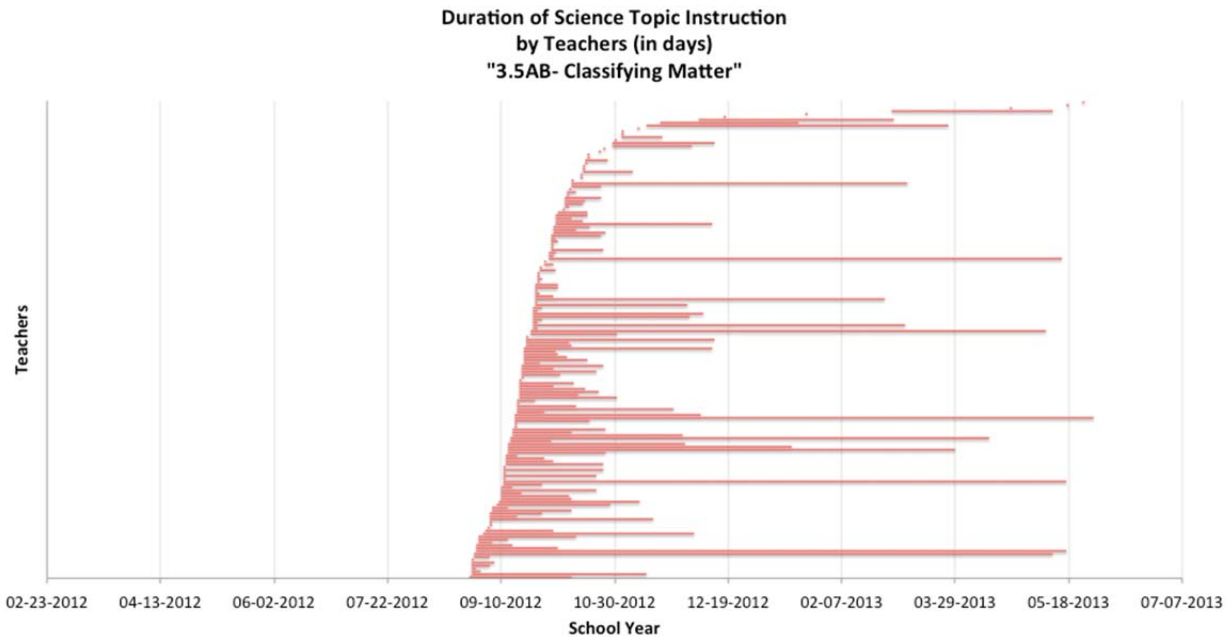

Figure 10. Timeline depicting duration of instruction for all teachers teaching the science topic entitled "3.5AB- Classifying Matter"

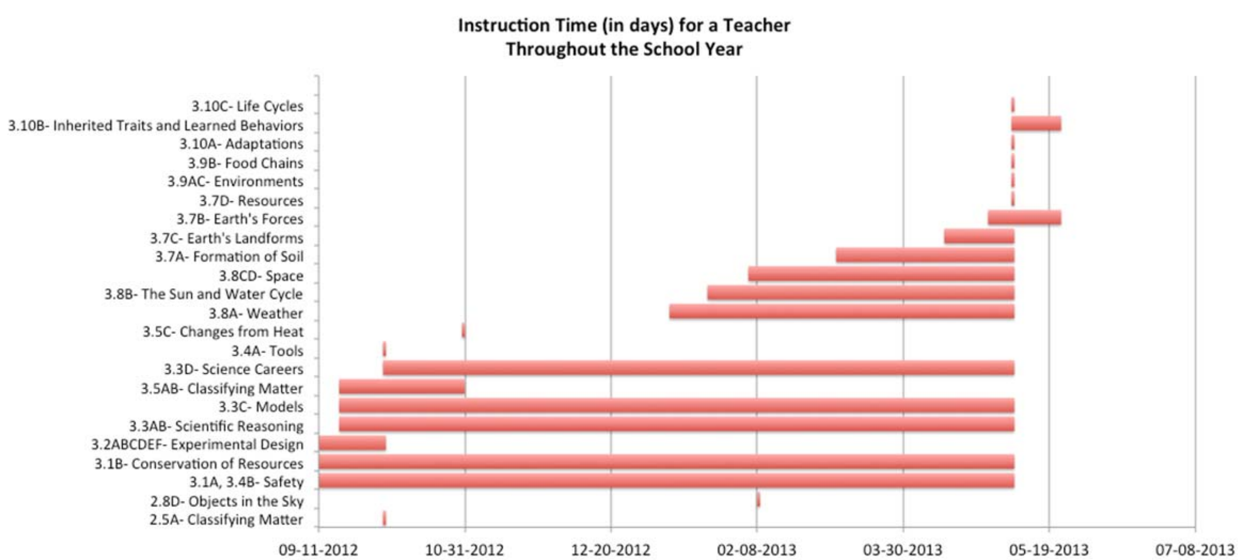

Figure 11: Timeline depicting duration of instruction of all science standards for one teacher.

5.2 Heat maps

Heat maps allow administrators and teachers to visualize how frequently different parts of the curriculum, including standards, steps, and activities, are accessed. Heat maps use colour or gray scale intensity to depict degrees of relationship between variables without focusing on exact values. In this

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

paper we discuss two cases of heat map use for our analysis. The first example (currently available to teachers on the STEMscopes dashboard) helps to understand curriculum usage at the science topic level across the 5E+I/A steps. In the second example, heat maps are intended for school or district administrators and offer overviews of instruction dosage for the 5E+I/A steps for individual teachers at the school level. Figure 12 depicts a partial view of curriculum use for 3rd and 4th grades in one school district (grade level is indicated by the numeric prefix in the TEKS column). Each row corresponds to one science topic. Blank cells indicate resources not available in STEMscopes. This graph shows high use for most topics in 3rd grade and from topics 4.5A to 4.6A, while low use can be seen from topics 3.8D to 4.4A. The most used steps are Engage and Explore, while Acceleration is the least used. The most frequently used standards are "3.8CD- Space" and "4.6A- Forms of Energy." Conversely, "4.3C- Models" is the least accessed. This type of visualization offers a quick overview of what has been taught and to what degree.

← Less More →

TEKS	Scope	Total	Overview	Essentials	Engage	Explore	Explain	Elaborate	Evaluate	Intervention	Acceleration
3.6B, 5.6D	ReSTAARt Forces and Motion	918	196	154	174	97	105	81	111		
3.6C	Magnetism and Gravity	4833	1000	523	805	671	540	468	372	324	130
3.7A	Formation of Soil	7306	1302	930	1217	898	1005	618	520	613	203
3.7B	Earth's Forces	6091	53	14	21	23	15	8	17	465	147
3.7C	Earth's Landforms	4184	891	407	764	553	537	355	261	271	145
3.7D	Resources	1358	302	175	239	111	173	116	91	98	53
3.8A	Weather	6688	1204	722	1366	1032	736	578	416	467	167
3.8B	The Sun and Water Cycle	8444	1535	953	1712	1210	871	817	562	538	246
3.8CD	Space	9542	1753	900	1544	1571	1151	1037	650	672	264
3.8D	ReSTAARt Planets	261	68	30	34	36	30	29	34		
3.9A	ReSTAARt Environments	120	27	13	25	15	15	17	8		
3.9AC	Environments	541	182	54	66	48	41	52	30	50	18
3.9B	Food Chains	351	86	19	78	22	57	35	28	24	2
4.10A	Adaptations	345	67	33	71	51	26	27	36	30	4
4.10B	Inherited Traits and Learned Behaviors	140	28	17	29	15	15	13	11	6	6
4.10C	Life Cycles	125	24	16	19	13	18	11	10	9	5
4.1A, 4.4B	Safety	686	277	30	58	114	96		111		
4.1B	Conservation of Resources	188	79	7	16	17	42		27		
4.2ABCDEF	Experimental Design	369	133	20	41	77	50		48		
4.3AB	Scientific Reasoning	253	96	7	28	40	46		36		
4.3C	Models	61	33		3	7	9		9		
4.3D	Science Careers	286	91	15	37	65	44		34		
4.4A	Tools	392	156	16	26	49	92		53		
4.5A	Classifying Matter	4836	1238	487	729	382	564	455	428	462	91
4.5B	Changes from Heat	3136	954	289	490	278	379	271	238	186	51
4.5C	Mixtures and Solutions	3150	838	268	496	282	392	269	314	220	71
4.6A	Forms of Energy	6834	1796	1105	1654	933	1090	704	758	616	178

Figure 12: Heat map depicting a partial list of STEMscopes content use for one district. Science topics included are for 3rd and 4th grades (indicated by the numeric prefix on the TEKS column).

We then investigated whether teacher visits were evenly distributed across steps or concentrated on a specific step. Figure 13 depicts six gray-scaled heat maps for elementary schools. Each row represents one teacher and columns from left to right show total visits, percentage of visits for the most used step, and seven grayed-scaled columns indicating the proportion of visits for each of the 5E+I/A steps for that teacher. The heat map scale ranges from 0% (lighter gray) to 100% (black). Rows are sorted by total visits by teacher; hence, higher users are at the top, while lower users are at the bottom.

Approximately three-quarters of teachers in "School E1" have the Explore step as the most visited one, between 27% and 49%. As the total of number of visits decreases, the most visited step varies between Elaborates and Engages. With the exceptions of teachers U_365 and U_933, access to the Acceleration and Intervention steps is extremely low. Usage in "School E2" also depicts high activity in the Explore category; however, high activity is also seen in the Elaborate step, ranging from 31% to 70% for most of

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

the higher users. Usage in "School E3" indicates a more balanced proportion for the Engages and Explores sections for approximately the top nine high users; it also suggests some degree of use for the Evaluates section. In contrast to the other two schools, "School E3" shows some degree of activity for the Intervention step.

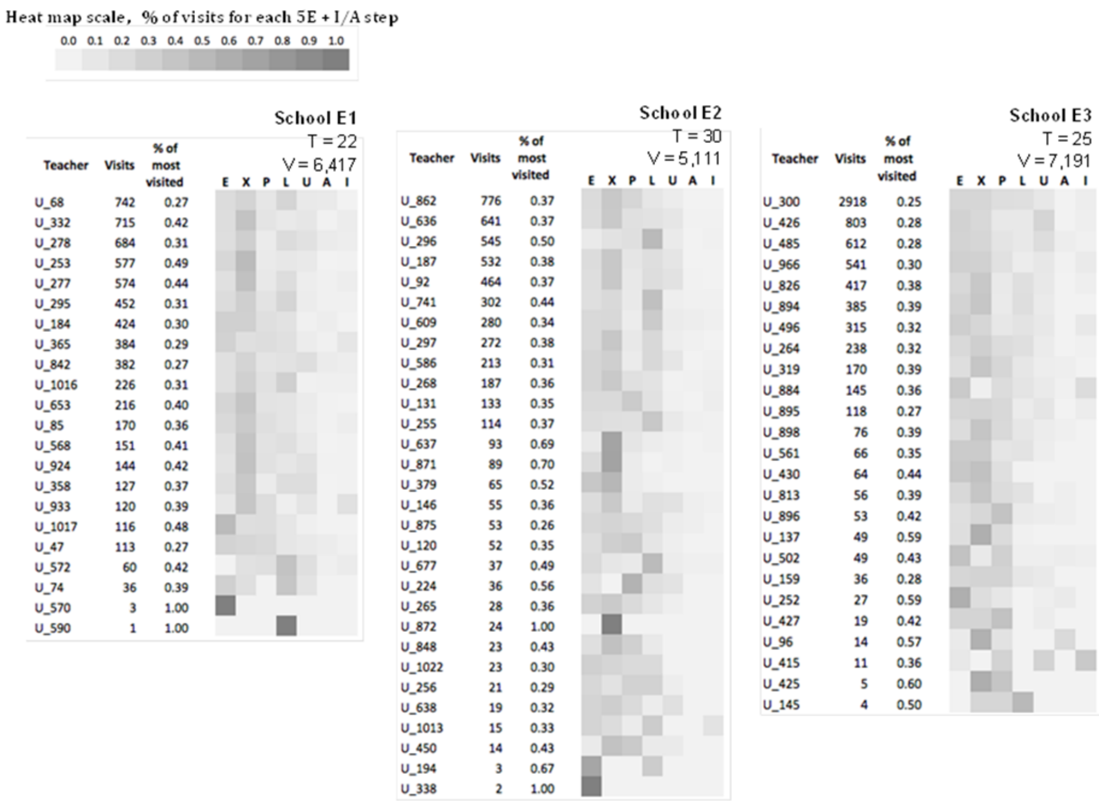

Figure 13. Heat maps depicting percentage of visits to each 5E+I/A step for teachers in three elementary schools.

Heat maps with a high concentration of cells in one or two columns would be an indication of highly skewed distributions. Therefore, the majority of heat maps show that most teachers tend to access different steps of the curriculum in a balanced way. Especially encouraging is to observe that overall the Engage, Explore, and Explain steps are evenly accessed for the highest users in the three schools. Conversely, teachers with a high concentration on one step are those located at the bottom of the graphs. Although, this is not the intended use of the curriculum, these teachers have an overall low number of visits. The low use for the Acceleration and Intervention steps can be attributed to teachers' lack of training with differentiating instruction among their students or no timely actionable information in the current dashboard, which makes it more difficult for them to identify diverse intervention levels for students.

The dosage of curriculum use for middle-school teachers shows sparser distribution and fewer teachers (see Figure 14). For instance, the highest user in "School M1," uses Evaluates the most, with 30% of usage. Darker colours on the last column indicate teachers using Intervention elements (almost half of teachers). Similar distribution can be seen in "School M2." Use patterns for "School M3" depict a concentration on Engage and Explore content for the top four highest users and a shift to Explain, Elaborate, and Evaluate content for the following two teachers. One exception is teacher U_789, who accessed Acceleration elements all the time.

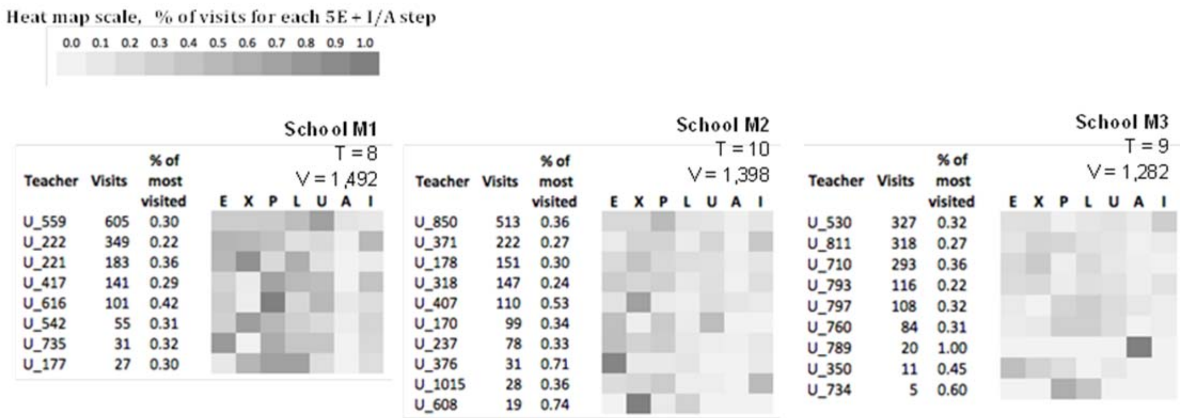

Figure 14. Heat maps depicting percentage of visits to each 5E+I/A step for teachers in three middle schools.

5.3 Mastery trackers

Finally, we have created a way for stakeholders to visualize student progress towards mastery of learning standards. This information is crucial for several reasons, not least of which is that in the U.S., public school students (and teachers and schools) are held accountable for their learning at the end of every school year in various subjects on mandatory state tests. The mastery tracker shown in Figure 16 below depicts whether individual students are making progress toward mastery on the required standards.[3]

The focus of the visualization is on progress because the mastery approach to student learning assumes that students should have more than one opportunity (or in many cases, as many opportunities as they need) to master a concept. Figure 16 shows a prototype of a dashboard with information for a class of 4th grade students. The columns show standards grouped by area (e.g., Organism and Environments), and the upward green arrows represent an improvement in performance from the previous assessment, while the downward red arrows indicate a decline in performance from one assessment to the next. Students are clustered by degree of intervention or acceleration they require, represented by the coloured cells. Students in the intervention group (yellow-coloured cells) are the ones who, although they are making progress, need a moderate amount of support from the teacher. Those in the intensive intervention group (the pink-coloured cells) require more help because they are not making sufficient progress towards mastery. Conversely, students in the acceleration group (green-coloured cells) can be challenged with more advanced activities from the acceleration step. This dashboard can help teachers better understand individual student needs, making it easier for teachers to differentiate and even personalize their instruction.

[3] The definition of mastery can be set by teachers, but usually the benchmark is 80% or higher on the final standards-based assessment.

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

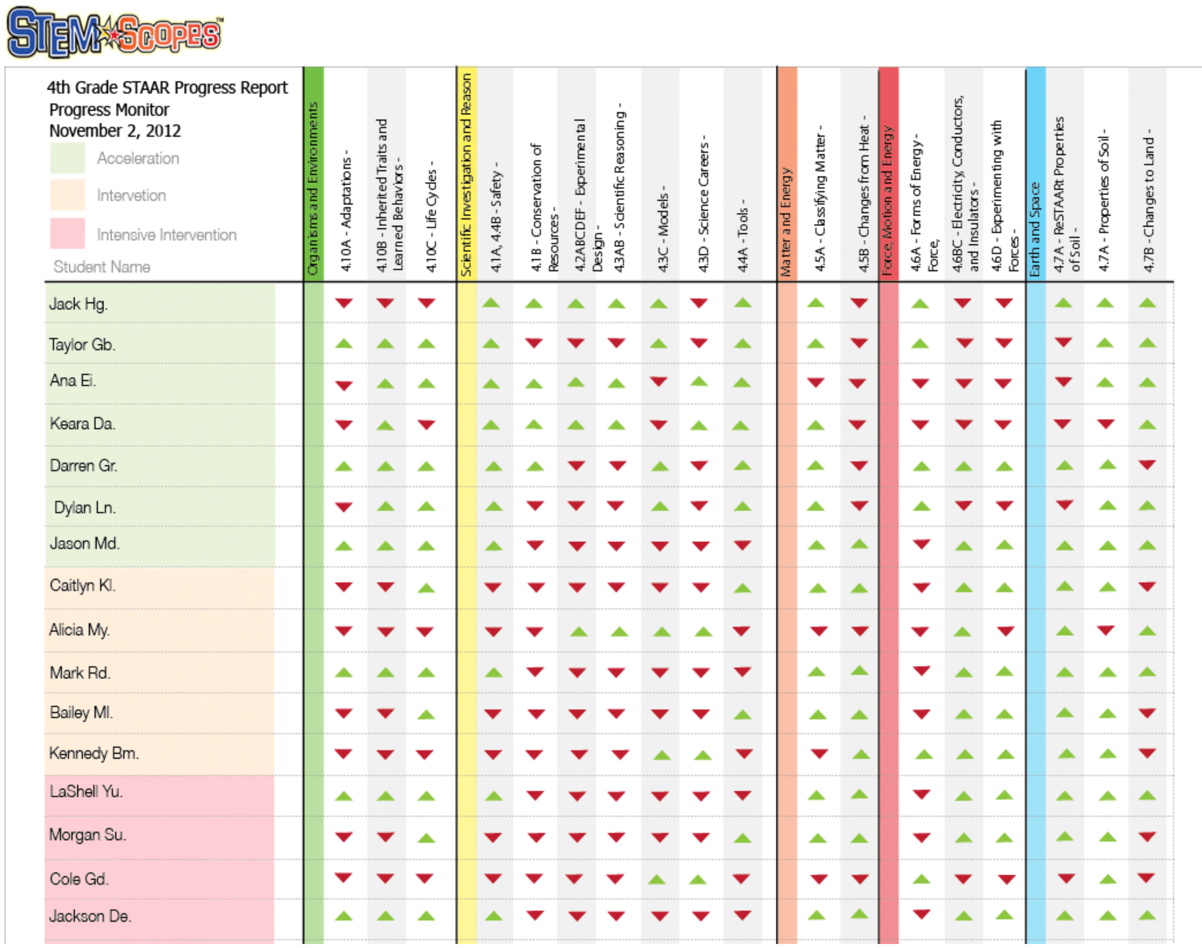

Figure 16: Dashboard prototype depicting a partial list of 4th-grade student progress for all scopes grouped by topic. Red arrows indicate grade drop from previous assessment, while green arrows show grade increase from previous assessment.

6. HARNESSING LEARNING ANALYTICS TO MEASURE FIDELITY OF IMPLEMENTATION

Research question 3 asks how LA data can be integrated into the evaluation of a curriculum. In this section, we discuss how we are integrating LA data into our curriculum evaluation work by using LA data as one measure of fidelity of implementation. In this section, we also detail the third prong of our LA data strategy; that is, how we pair the LA data with more traditional and qualitative data collection and analysis to create a holistic understanding of curriculum use and impact.

In order to understand the impact of the curriculum on student learning, we relate the level and kind of curriculum use to different student outcomes (Duriak & Dupre, 2008; Song & Herman, 2010). It is essential to understand how stakeholders are using the curriculum in order to contextualize the impact.

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

For example, if there is little use of the curriculum, then it is not possible to attribute any change in student learning to use of the curriculum; the reverse also is true. The LA data are key here because they provide a detailed and multifaceted account of curriculum use, as we described above. We utilize the LA data as a complement to more traditional forms of measuring use, which we also describe below. And while it can be difficult to scale a high-quality qualitative data collection strategy, it is difficult, if not impossible, to really understand how an online curriculum, such as STEMscopes, is being used without spending time observing its use in classrooms and talking to teachers and students about their use.

Typically, program evaluators collect data relating to the fidelity of implementation (or use) of an intervention in order to gauge the relative impact of the intervention on student learning. Dusenbury, Brannigan, Falco, and Hansen (2003) suggest that fidelity of implementation (FOI) should be measured in five ways: 1) *adherence* to the program; 2) *dose* or exposure to the program; 3) *quality* of program delivery; 4) *participant responsiveness* to the program; and 5) *program differentiation* (differentiation of components within the program). Other researchers have proposed different ways to conceptualize and measure FOI, but we focus on the previous five. In the rest of this section, we briefly describe how we incorporate LA specifically into evaluations, and then we discuss three other ways we collect data to measure impact.

We use LA data to measure fidelity of implementation in four of the five ways described in the previous section. First, LA data provide a measure of *adherence*; that is, whether a teacher or student is utilizing the curriculum in the intended way. LA data allow us to track precisely how a teacher has interacted with the program, including the proportion of inquiry instruction, as discussed in section four. Second, LA data can measure *dosage*, or how frequently and for how long a teacher engages with the curriculum. Third, LA data can be utilized to measure how *engaged* a teacher is with the curriculum by looking at, for example, time spent on a component or how quickly a teacher moves through components. Fourth, LA data can be used to measure *program differentiation*, or the impact of different components of the curriculum on student learning. This can be done by examining how a teacher interacts with various components, and weighing these data with other dosage data to estimate their impact on student learning outcomes, such as interim assessment scores as well as end-of-year test scores.

6.1 Using teacher surveys to measure fidelity of implementation

A common way to gauge teacher use of a curriculum (or other intervention) is to ask them to self-report in a survey. As part of our ongoing effort to evaluate the use and impact of the STEMscopes curriculum, in the fall of 2012 we created and administered a survey to 755 elementary school teachers in a large urban school district where some schools use the curriculum. The survey, containing 80 items, was created based on existing surveys about technology and data use. While some of the items are specific to the STEMscopes curriculum and its particular offerings, others are generalizable to other technology or curricula. The survey asks teachers about the following concepts: teacher notions of data and data use, how teachers use STEMscopes, attitudes toward STEMscopes, challenges to using STEMscopes, as well as district and school support for using STEMscopes. The survey was piloted and validated in the fall of 2012 (See Snodgrass Rangel, et al. 2013 for more information on the survey and for more detailed findings), 210 teachers responded to the survey, a response rate of 28%.

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

As part of this study, and as an example of how to combine these two kinds of data, we related teacher survey responses to their LA data to see how closely they were related. Specifically, we used correlation analysis on the associations between these three variables and several user analytics variables. We also created and used four variables from the LA data we had access to at that time:

1. The number of visits to pages containing instructional content for any of the seven steps, weighted by the number of *grade levels* accessed by the teacher's account (continuous variable)
2. The number of visits to pages containing instructional content for any of the seven steps, weighted by the number of *objectives* the teacher had taught to that point (continuous variable)
3. A categorical variable (created using quartiles) measuring the un-weighted number of visits to pages containing instructional content for any of the seven steps (1=low use, 2=medium low use, 3=medium high use, 4=high use)
4. A categorical variable (created using quartiles) measuring the number of visits to pages containing instructional content for any of the seven steps, weighted by the number of grade levels accessed by the teacher's account (1=low use, 2=medium low use, 3=medium high use, 4=high use).

First, we correlated teacher responses to the survey question asking how often they utilized the 5E+I/A steps in the suggested order and an analytics variable that measured how frequently teachers actually used all seven steps in the suggested order. Second, we correlated teacher responses to how often they utilized the 5E+I/A steps in the suggested order and the frequency with which teachers used the 5E (without the I/A) steps in the suggested order. We found that while teachers reported using the curriculum for some to most of the standards they taught, the LA data told a different story (see Table 3).

Table 3. Descriptive Statistics for Reported and Actual Use of STEMscopes

		Mean	SD	Range
Reported Use				
	Average use of teacher components	2.70	0.76	1.00 - 4.00
	Average use of 5E+I/A steps	2.48	0.64	1.00 - 4.00
	Average use of 5E+I/A steps and components in order	2.62	0.60	1.00 - 4.00
Actual Use				
	Number of 5E visits weighted by number of grades taught	103.42	125.84	1.00 - 1084.00
	Number of 5E visits weighted by number of objectives taught	20.09	13.36	2.00 - 75.00
	Average number of visits to 5E+I/A steps (categorical)	2.49	1.13	1.00 - 4.00
	Average visits to 5E+I/A steps, weighted by number of grades taught (categorical)	2.24	0.83	1.00 - 4.00

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

In the correlation analyses, we found that while almost all of the correlations were positive, very few were statistically significant relationships. The two that were significantly (but weakly) related were between the categorical variable measuring the average number of visits to a page containing instructional content from any of the seven steps (1=low use, 4=high use) and the reported average use of teacher components ($r(114)=0.195$, $p<0.05$), and the reported combined average use of the teacher components, the 5E+I/A steps, and the frequency with which a teacher uses the steps in the appropriate sequence ($r(114)=0.219$, $p<0.05$). The results of the correlation analysis can be found in Table 4 below.

Table 4. Results of correlation analysis for reported and actual use

User analytics		Reported average use of teacher components	Reported average use of 5E+I/A steps	Average use of 5E+I/A steps and components in order
Number of 5E visits weighted by number of grades taught (N=114)	Pearson Correlation	0.160	0.030	0.140
	Sig. (2-tailed)	0.090	0.780	0.130
Number of 5E visits weighted by number of objectives taught (N=114)	Pearson Correlation	0.090	-0.050	0.060
	Sig. (2-tailed)	0.330	0.570	0.560
Average number of visits to 5E+I/A steps (categorical) (N=114)	Pearson Correlation	0.195 *	0.140	0.219 *
	Sig. (2-tailed)	0.040	0.140	0.020
Average visits to 5E+I/A steps, weighted by number of grades taught (categorical) (N=113)	Pearson Correlation	0.060	-0.040	0.030
	Sig. (2-tailed)	0.550	0.710	0.730

* Correlation is significant at the 0.05 level (2-tailed)

We also correlated teacher survey responses about how frequently they use the 5E+I/A steps in the suggested order with the frequency with which they actually used all of the 5E+I/A steps in the suggested order (based on the data analytics). The results showed no significant association between a teacher's reported use and actual use of the 5E+I/A steps in the suggested order ($r(111) = -0.02$, $p = 0.84$). Second, we conducted a correlation between teacher survey responses about how frequently they use the 5E+I/A steps in the suggested order with the frequency that they actually used just the steps in the suggested order. The results showed no significant association between a teacher's reported use and actual use of the 5E steps in the suggested order ($r(111) = 0.02$, $p = 0.83$). In other words, there appeared to be no relationship between perceived and actual use regarding the 5E+I/A steps in the order intended by the curriculum designers.

The question that these two analyses raise, of course, is which measure of teacher use is correct? The answer we have settled on is that neither is "wrong" or "right," and that therefore they must be used jointly to piece together a balanced picture of teacher use. In other words, data from different sources must be triangulated. In the next two sections, we describe two more ways to collect data for the purpose of creating a more nuanced picture of curriculum use.

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

5.4 Using teacher focus groups and interviews to measure fidelity of implementation

The research team also conducted teacher focus groups to understand how teachers use the curriculum. The benefit of conducting focus groups is that we are able to engage directly with teachers and ask them about their use of the curriculum. We can follow up on their answers and probe more deeply for explanations of certain behaviours. In this way, we can look at data based on access to the curriculum next to data on what the teachers do once they have accessed the curriculum and are in their classrooms teaching students.

The focus groups we conducted were separate from the external evaluation discussed in previous sections. Instead, they have been part of internal research efforts. For example, in the fall of 2012, the research team asked the principals at two schools that use STEMscopes to put together a group of 3–5 teachers to meet with us. A total of 11 teachers participated (see Table 5). Of these teachers, all but two were self-contained teachers who taught science in addition to other subject areas. Of the 11 teachers, three were novices (first year), one had medium experience (2–5 years), four were advanced (6–20 years), and four were veterans (21+ years).

Table 5. Description of focus group participants

Total teachers	Science only	Grades K-2	Grades 3-5	Novice teacher	Medium experienced	Advanced teacher	Veteran teacher
11	2	9	2	3	1	4	4

The team created an interview protocol to guide the focus group. The interviews were semi-structured, which allowed the conversation to develop organically. The protocol asked teachers how they felt about teaching science, using technology, using STEMscopes, and what challenges they faced when using STEMscopes to teach science. Here, we only share a few findings as they relate to use of the curriculum with the intent of demonstrating the type of findings that can be extracted from direct conversations with the curriculum's stakeholders.

5.4.1 Uses of STEMscopes by teachers in grades K–2

The majority of participating teachers were K–2, self-contained teachers. Though all of the teachers were very positive about science and reported enjoying teaching science, it was clear that they had very little time to teach science, and so their use of STEMscopes and other science resources was quite limited. According to the teachers, they were only able to teach science two to three times a week, and this relatively short time was split between hands-on experience, such as an experiment, and science vocabulary. In fact, we specifically asked the teachers how often they were able to *do* science; they reported that they were working on a hands-on experience with their students almost every week.

Vocabulary, however, was the most frequently mentioned use of the curriculum; the kindergartners draw pictures of science concepts and the second graders make vocabulary cards with pictures, labels, and definitions.[4] STEMscopes' vocabulary materials were particularly helpful for the bilingual teachers,

[4] Most frequently mentioned does not necessarily mean the most frequently used. The analytics data suggest that K–2 teachers access the Explore step the most, indicating they do hands-on activities with their students.

most of whom teach science as an English as a Second Language (ESL) class where the focus is as much on learning English as it is on learning content. Also included in the discussion was the use of science to teach literacy. The teachers described spending most of their time on math and literacy to prepare students for the 3rd grade state tests, and so science often became another way to work on literacy. A teacher from one school told us how her students read science stories, saying, “this week… there is a story about plants, so then you know they are identifying roots, petals, and stems, all that stuff.” Other teachers described how the students kept science journals where they pasted in passages and diagrams to practice new vocabulary.

5.4.2 Uses of STEMscopes by teachers in grades 3–5

Teachers in the upper-elementary grades (particularly 5th grade) had more time for science instruction, often more than an hour a day, which allowed them to utilize more parts of the curriculum. However, the teachers still expressed the need for more time to cover all of the material and review material from previous grades, where students often receive little science instruction. One of the teachers described how she strives to use all of the options contained within the curriculum, saying, “I wouldn’t say [I use it] to its maximum but I would say as close as I can possibly get within the given timeframe that I am allotted.”

These two teachers indicated that they try to follow the curriculum closely by doing one “E” step on each day and using as many of the assessments as they have time for. One teacher confided that, “I don’t use every experiment that you guys [have]…I use most of them though, but I try to make sure that I, we do an experiment [from STEMscopes] every Tuesday and then I try to do one on Wednesdays also, I try two days of complete experiments.” In this way, the teacher generally is able to meet the student needs for hands-on learning and intervention at the end of the week for those students who did not demonstrate mastery on the post-assessment. Both of the teachers also told us that, for the most part, they complete each of the 5E steps in order, though one of the teachers told us that she begins each new unit with direct instruction and note-taking. Our observations of both classrooms confirmed that both teachers begin units with direct instruction and note-taking.

Finally, the teachers rely primarily on STEMscopes for their actual lessons while following the district’s scope and sequence, which directs them to teach the standards in a certain order. One of the teachers told us how she reconciles using a curriculum not officially adopted by the district: “I just read the teacher background and I just make sure that I am teaching in the sequence that [the district] is giving me to teach… I use the teacher background and all the activities from STEMscopes.”

5.5 Using classroom observations to measure fidelity of implementation

Through classroom observations, we have seen how teachers and students interact — and do not interact — with the online curriculum. Observations, though time-consuming, are important because they allow for comparing and classifying interactions. Teachers often unwittingly over-report how often they use the curriculum, and observations can cut through the problems of self-reporting and perception.

In that same internal evaluation, we observed two teachers at different schools as they taught science. Each of the teachers we observed taught science only and was a very experienced teacher. One of them taught science to grades 3–5, while the other only taught 5th grade science. The teacher who taught 3rd–

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

5th grade science had 18 years of experience as a teacher and instructional coach (though not in science), and this was his/her first year teaching science only. Furthermore, his/her previous teaching experience was in 1st grade. This teacher also helps to translate curriculum for STEMscopes into Spanish, making him/her very familiar with the curriculum. The other teacher had taught for 22 years, nine of them as a science teacher exclusively. This teacher also had previous experience teaching middle-school math. During the classroom observations, we took extensive field notes, which we later coded using an observation protocol created for science classrooms. Based on these notes, we created long and detailed "inquiry profiles" that described how teachers used the curriculum and level of inquiry in each teacher's classroom. Here, we report only general findings.

We learned a great deal about the context and quality of implementation from the observations. For instance, we noted that in these classrooms, there were no more than two computers available to students, and the teachers either printed materials from the website, or projected the website onto a whiteboard. This finding can help explain variation in teacher use — some teachers may have more technology available to them than these teachers did. Similarly, we observed that while some teachers followed the curriculum very carefully, others used fewer of the activities and instead incorporated other available resources, and others still utilized the steps and activities out of the intended order. In short, the observations reveal "the face" behind the "clicks" that LA data show us.

6. CONCLUSIONS AND FUTURE WORK

LA has the potential to bring important changes to K–12 education research and development if implemented with a pedagogically sound, mixed methods strategy, such as the one described in this paper. Incorporating the LA data into our ongoing research on curriculum has enabled us to dig deeply into patterns of use and to ask nuanced questions that relate to the pedagogy underpinning the curriculum and its effectiveness. We have also learned that on their own, LA data paints an incomplete picture of what teachers do in their classrooms with the curriculum. By bringing in more traditional measures of use, such as surveys, interviews, and classroom observation, we were able to contextualize some of the findings gleaned from the LA data.

As we have improved our technical capacity to work with the LA data, we have also created new ways to make the data accessible to the curriculum's key stakeholders, namely teachers and administrators. Data visualization has been the most powerful tool we have put into practice so far to translate data into information, and future work will examine the utility of these tools as well as develop new tools based on feedback.

There also are several hurdles that must be overcome to ensure that these new data collection and analysis tools do, in fact, fulfill their promise. First, schools and districts must address the gap in access to technology. While some schools have achieved one-to-one computing, most schools are not even close to this goal, and this has profound implications for our ability to collect reliable analytics data. Our conversations with and observations of teachers revealed that teachers and students often share accounts, and that students are limited in the activities they can complete online. This, in turn, means that the analytics data we collect may not be reliable: for some teachers the usage analytics may accurately portray their use, while for others, they may not.

(2014). A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K–12 Science Curriculum. *Journal of Learning Analytics*, *1*(2), 94-125.

A second and equally important challenge is time. Teachers often do not have time to incorporate technology and online activities into their regular instruction. It is time-consuming to plan for this incorporation, especially because it often requires the teacher to learn new software or strategies, and because it requires the teacher to create new classroom and management routines. The result of this hurdle is that the computer systems that should be collecting analytics data are not fully implemented — or are not implemented at all, which degrades the quality of data we can collect.

In our future work, we will include curriculum elements in each of the 5E+I/A steps and investigate the effects of the characteristics and properties of those elements in curriculum use. Similarly, we will disaggregate use by science standard, skills, and concepts covered on each curriculum step. These new scenarios will offer a deeper and richer picture of science teaching and learning in the K–12 realm. We will conduct studies to investigate sequences of visits, pacing, and patterns of use at different points in time. This in turn will help us to answer questions related to curriculum use changes throughout the school year and consistency of use across science standards, schools, and districts, with implications for targeted professional development for teachers and curriculum design. We will also examine specific use patterns and relate them to student learning outcomes. These, no doubt, will present the team with fresh challenges as some components of the curriculum may turn out to be more effective than others, and these results may confound the relationships we hypothesize to exist.

Though not a panacea for the challenges facing K–12, a strong LA data strategy can help researchers and educators alike make sense of the large amounts of data generated by the increased use of mobile devices, computers, and other technology in classrooms. With an improved understanding of data, students can learn from their own mistakes. Teachers can improve their lessons, work with students who need extra help, and provide extensions for those ready to move on. Analysts can improve the curriculum and the resources available to teachers and students. The strategy described in this paper was developed to analyze detailed usage patterns of the online science curriculum, STEMscopes. This approach enables analysis of large amounts of data and opens the possibilities to complex undertakings such as the creation of personalized learning environments.

ACKNOWLEDGMENTS

We would like to thank the STEMscopes production team for their work on the science curriculum. Specials thanks go to the teachers and science specialists who participated in our focus groups. Thanks also go to Amber Muscarello for her expertise on curriculum design and science teaching. Her contribution helped in the interpretation of the analytics data. Thanks to Dr. Elizabeth Bell who helped with the survey data. Finally, thanks to Dan Hoyt for reviewing this manuscript.

REFERENCES

Arnold, K. (2010). Signals: Applying academic analytics. *EDUCAUSE Quarterly*, *33*(1).

Atkinson, M., & Wiley, D. (2011). Learning analytics as interpretive practice: Applying Westerman to educational intervention. *Proceedings of the First International Conference on Learning Analytics and Knowledge*, Banff, Alberta, Canada, February 17–March 1, 2011, 117–121.

Baker, S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future versions. *Journal of Educational Data Mining*, *1*(1), 3–17.

Beck, J., & Mostow, J. (2008). How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students. *Intelligent Tutoring Systems, Lecture Notes in Computer Science*. Berlin/Heidelberg: Springer, 435–362.

Bybee, R., Taylor, J., Gardner, A., Van Scotter, P., Powell, J., Westbrook, A., & Landes, N. (2006). The BSCS 5E instructional model: Origins, effectiveness, and applications. Retrieved from http://bscs.org/sites/default/files/_legacy/BSCS_5E_Instructional_Model-Executive_Summary_0.pdf

Campbell, J., & Oblinger, D. (2007). Academic analytics. *EDUCAUSE Review, 42*(4), 40–57.

Cho, V., & Wayman, J. (2014). Districts' efforts for data use and computer data systems: The role of sensemaking in system use and implementation. *Teachers College Record Volume*, *116*(2), ID Number: 17349.

Crawford, B. (2000). Embracing the essences of inquiry: New roles for science teachers. *Journal of Research in Science Teaching, 37*(9), 916–937.

Culatta, R. (2012). From innovation clusters to datapalooza: Accelerating innovation in educational technology. Retrieved from http://www.educause.edu/ero/article/innovation-clusters-datapalooza-accelerating-innovation-educational-technology

DeSantis, N. (2012). A boom time for education start-ups: Despite recession investors see technology companies' "internet moment." *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/A-Boom-Time-for-Education/131229/.

Duriak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–349.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research: Theory & Practice, 18*(2), 237–256.

Duval, E. (2011). Attention please! Learning analytics for visualization and recommendation. *Proceedings of the First International Conference on Learning Analytics and Knowledge*, Banff, Alberta, Canada, February 17–March 1, 2011, 9–17.

Elias, T. (2011). Learning analytics: Definitions, processes, and potential. Retrieved from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

Farzan, R. (2004). Adaptive socio-recommender system for open-corpus e-learning. *Doctoral Consortium of the Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*. Retrieved from http://people.cs.pitt.edu/~rosta/docs/PDF/AH2004DC.pdf

First International Conference on Learning Analytics and Knowledge (2011). https://tekri.athabascau.ca/analytics/

Ha, S., Bae, S., & Park, S. (2000). Web mining for distance education. *Proceedings of the IEEE International Conference on Management of Innovation and Technology*, Singapore, November 12–15, 715–719.

Hamalainen, W., Suhonen, J., Sutinen, E., & Toivonen, H. (2004). Data mining in personalizing distance education courses. *Proceedings of the World Conference on Open Learning and Distance Education*. Retrieved from http://www.cs.joensuu.fi/~whamalai/articles/icde_v2.pdf

Haury, D. L. (1993). *Teaching Science through Inquiry with Archived Data. ERIC Digest*. (EDO-SE-01-1). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental, Education. Retrieved from http://www.gpo.gov/fdsys/pkg/ERIC-ED465545/pdf/ERIC-ED465545.pdf

Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Communications of the ACM*, *55*(4), 45–54.

Hendricks, M., Plantz, M., & Pritchard, K. (2008). Measuring outcomes of United Way-funded programs: Expectations and reality. In J. G. Carman & K. A. Fredericks (Eds.), *New Directions for Evaluation*, Hoboken, NJ: Wiley, pp. 13–35.

Heraud, J., France, L., & Mille, A. (2004). Pixed: An ITS that guides students with the help of learners' interaction log. *7th International Conference on Intelligent Tutoring Systems*, 57–64. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.4987&rep=rep1&type=pdf

Hew, F., & Brush, T. (2007). Integrating technology into K–12 teaching and learning: Current knowledge gaps and recommendations for future research. *Educational Technology Research and Development*, *55*, 223–252.

Ingram, D., Seashore Louis, K., & Schroeder, R. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106*(6), 1258–1287.

Jagadish, H. (2012, May 5). Big data: It's not just the analytics. Message posted to http://wp.sigmod.org/?p=430.

Junco, R. (2012, October 28). Most ed-tech startups suck! Here's where they go wrong. Message posted to http://venturebeat.com/2012/10/28/most-ed-tech-startups-suck-heres-where-theyre-going-wrong/.

Keengwe, J., Onchwari, G., & Wachira, P. (2008). Computer technology integration and student learning: Barriers and promise. *Journal of Science Education and Technology, 17(6),* 560–565.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112*(4), 496–520.

Keys, C., & Bryan, L. A. (2001). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of Research in Science Teaching, 38*, 631–645.

Koehler, M. J., & Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research, 32*(2)*,* 131–152.

Lachat, M., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk. 10*(3), 333–349.

Lawless, K. A., & Pellegrino, J. W. (2007). Professional development in integrating technology into teaching and learning: Knowns, unknowns, and ways to pursue better questions and answers. *Review of Educational Research, 77,* 575–614.

Leonardi, P. M. (2009). Crossing the implementation line: The mutual constitution of technology and organizing across development and use activities. *Communication Theory*, *19*, 278–310.

Long, P., & Siemens G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*. Retrieved from http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education

Lu, J. (2004). Personalized e-learning material recommender system. *Proceedings of the International Conference on Information Technology for Application*, Sydney, Australia, January 9–11, 374–379.

Mazza, R., & Dimitrova, V. (2004). Visualising student tracking data to support instructors in web-based distance education. *Proceedings of the 13th International World Wide Web Conference, Alternate Track Papers & Posters*, New York, May 17–22, 154–161.

Means, B., Padilla, C., & Gallagher, L. (2010). Use of education data at the local level: From accountability to instructional improvement. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development. Retrieved from http://www2.ed.gov/rschstat/eval/tech/use-of-education-data/use-of-education-data.pdf

Merceron, A., & Yacef, K. (2005). TADA-Ed for educational data mining. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, *7*(1), 267–287. Retrieved from http://imej.wfu.edu/articles/2005/1/03/index.asp

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction: What is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, *47*(4), 474–496.

Monroy, C., Snodgrass Rangel, V., & Whitaker, R. (2013). STEMscopes: Contextualizing learning analytics in a K–12 science curriculum. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, Leuven, Belgium, April 8–12, 210–219.

Norris, D., Baer, L., Leonard, J., Pugliese, L., & Lefrere, P. (2008). Action analytics: Measuring and improving performance that matters in higher education. *EDUCAUSE Review, 43*(1), 42–67.

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science, 11(4),* 404–428.

Pahl, C. (2004). Data mining technology for the evaluation of learning content interaction. *International Journal on E-Learning, 3*(4), 47–63.

Pardo, A., & Delgado, C. (2011). Stepping out of the box: Towards analytics outside the learning management system. *Proceedings of the First International Conference on Learning Analytics*, Banff, Alberta, Canada, February 27–March 1, 163–167.

Pavlik, P., Cen, H., & Koedinger, K. (2009). Learning factors transfer analysis: Using learning curve analysis to automatically generate domain models. *Proceedings of the Second International Conference on Educational Data Mining*, Cordoba, Spain, July 1–3, 121–130.

Pavlo, A., Paulson, E., & Rasin, A. (2009). A comparison of approaches to large-scale data analysis. *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, Providence, Rhode Island, USA, June 29–July 2, 165–178.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, 40*(6), 601–618.

Siemens, G. (2012). Learning analytics: Envisioning a research discipline and a domain of practice. *Proceedings of the Second International Conference on Learning Analytics and Knowledge*, Vancouver, Canada, April 29–May 2, 4–8.

Snodgrass Rangel, V., Monroy, C., & Whitaker, R. (2013). Teacher-centered design: Improving a K–12 science curriculum and its dashboard. *Proceedings of the 2013 National Association for Research in Science Teaching, NARST Conference*. Puerto Rico, April 2013.

Song, M., & Herman, R. (2010). Critical issues and common pitfalls in designing and conducting impact studies in education: Lessons learned from the What Works Clearinghouse (Phase I). *Educational Evaluation and Policy Analysis, 32*(3), 351–371.

Suthers, D., Ravi, V., Medina, R., Joseph, S., & Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers & Education*, *50*(4), 1103–1127.

Talavera, L., & Gaudioso, E. (2004). Mining student data to characterize similar behavior groups in unstructured collaboration spaces. *16th European Conference on Artificial Intelligence*, Valencia, Spain, August 22–27, 17–23.

Vatrapu, R., Tplovs, C., Fujita, N., & Bull, S. (2011). Towards a visual analytics for teachers' dynamic diagnostic pedagogical decision-making. *Proceedings of the First International Conference on Learning Analytics and Knowledge*, Banff, Alberta, Canada, February 27–March 1, 93–98.

Venner, J. (2009). *Pro Hadoop: Build scalable, distributed applications in the cloud*. New York: Apress.

Watters, A. (2012, May 4). Learning analytics: Lots of education data... Now what? Message posted to http://hackeducation.com/2012/05/04/learning-analytics-lak12/

Wayman, J. C. (2005). Involving teachers in data-driven decision making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed at Risk, 10*(3), 295–398.

Wayman, J. C., Cho, V., & Richards, M. P. (2010). Student data systems and their use for educational improvement. In P. Peterson, E. Baker, & B. McGraw (Eds.), *International encyclopedia of education* (3rd ed.) (pp. 14–20). Oxford, UK: Elsevier.

Whitaker, J. R. (2012). Responding to the need for intervention: Six easy steps prime students for mastery of science concepts. *Science and Children*, *50*(4), 75–79.

Zuiker, S., & Whitaker, J. R. (2013). Refining inquiry with multi-form assessment: Formative and summative assessment functions for flexible inquiry. *International Journal of Science Education, 36*(6), 1037–1059. doi:10.1080/09500693.2013.834489

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

Peer Promotions as a Method to Identify Quality Content

Bjorn Levi Gunnarsson and Richard Alterman
Brandeis University, USA
bjornlevi@cs.brandeis.edu

Abstract: Promoting blog content is a social activity — a means of communicating one student's appreciation of another student's work. This article explores the feasibility of using student promotions of content, in a blogosphere, to identify quality content, and implications for students and instructors. We show that students actively and voluntarily promote content, identify quality material with considerable accuracy, and use promotion data to select what to read. We explore the benefits of knowing which students are good and poor predictors of quality content, and what instructors can do with this information in terms of feedback and guidance.

Keywords: Blogging, social blogging, liking, promoting, assessment, quality content, learning analytics, knowledge community, social recommendation, recommender system

1 INTRODUCTION

The traditional classroom is changing, with devices (laptop, smartphone, et cetera) that access the Internet, connecting the classroom to the larger world. For some classes, all communication is online, making the classroom completely digital. Other classes are blended to some degree or another — part classroom, part online. Using digital communication is relatively new in learning. Neither learners nor instructors can be assumed to pick up this new learning tool automatically and fully understand its possibilities or use it effectively. To complicate matters further, the technology is rapidly changing; new applications are constantly being created and devices upgraded. What does not change is that, with each day, the integration of learning with digital technologies is further advanced. This marriage produces significant amounts of data. This is where the field of learning analytics comes into play, applying analytical methods to the content created by the learners online (Siemens, 2011).

In class, the instructor has face-to-face access to her students. Online, students and instructors become entities of the Internet, where their presentation of self is constrained by the selected mode of communication. Common communication methods are email, forum posts, blog posts, and chat rooms (possibly with audio and video). These methods of communication have different features than face-to-face communication; each of them providing alternate constraints and opportunities for contributing to a learning discourse (e.g., Baker, Hansen, Joiner, & Traum, 1999). Each contribution can be bookmarked, shared, *liked*, read multiple times, tagged, referenced, and quoted. Within the bounds of a semester-length class of students, the creation of this kind of information is a sign of social activity among students. Making visible this kind of social data to the entire class provides important peer feedback and recognition; it also potentially creates valuable data for learning analytics.

This article will show that social data is a significant element in loosely coupled online collaborative environments, beneficial to both students and instructors. We will focus on one sort of social data: *promotions* — one student anonymously and subjectively assessing the quality of another student's work by clicking a button. The promotion activity can take alternate forms; for example, *liking* or giving a *merit badge*. We show that the data can be used to solve significant problems for both students and instructors.

The field of learning analytics applies methodological scrutiny to learner-created content, such as the data created through promotions. In online learning settings, collecting data is easy; all activities can be saved for analysis. It would not be so easy to get the same data from a classroom discussion. This easy access to data has created the field of learning analytics and the data has inspired much new research. Creating sense-making visualizations for instructors and administrators (Dawson, 2009), exploring knowledge building in online social settings (Ferguson & Buckingham Shum, 2012), and data mining institution-sized data sets (Arnold, 2010; Smolin & Budakov, 2012) are examples of work done in this field.

In a technology-mediated environment, the intersection of the learning framework and the analytic methods form a middle space (Suthers & Verbert, 2013). An underlying theory of learning is a significant element of the foundation of any online learning environment: it provides both a rationale for the design of the learning environment, including the scaffolding, and a basis for analysis (p. 2). In other words, learning theory motivates the invention of technology, the use of the technology produces data, the analysis of the data can suggest methods to fix problems and enhance the learning experience for any (or all) of the stakeholders, most notably either the students or the instructor. These enhancements and solutions can then again impact the theory and produce another cycle of new problems and solutions.

Contrast, for example, the underlying assumptions of two alternate projects to support student online learning. In one case, the VirtualMathTeam project (Stahl, 2009), students work together online, in small groups, at the same time from different locations in a partial joint-focus space; learning for relatively short periods of time, by negotiating meaning in the social world (Stahl, Koschmann, & Suthers, 2006). In a contrasting case, students collaborate in a blogging environment throughout the semester — working together in a loosely coordinated fashion at different times and different locations — creating and distributing knowledge, and collaboratively building skills (Alterman & Larusson, 2013). Because the learning framework for these two examples is different, the middle space will be different. The analytical methods might have some overlap, but some of the goals, results, and conclusions will be radically different.

The research reported on in this article analyzes learner-created content in two classes (107 and 50 students). The classes were both blended courses, with lectures twice per week, and ongoing participation in an online knowledge community as a requirement. While the amount of data is not as much as for institution-sized data sets or MOOCs (massively open online courses), the data is still significant because it represents the normal conditions for many learning environments. Despite the limits of the dataset size, important results can nevertheless be obtained.

In both classes presented, each as a case study, the students blogged throughout the semester, collectively producing much content each week; in one of the classes, students produced roughly 300 contributions per week totalling around 100,000 words, or ten times the length of this article! Without some learning analytic intervention, this amount of raw material for the students to navigate, and for the instructor to assess, is overwhelming.

This article explores the value of promotions in a collaborative online learning environment in a blended course. The students work in a loosely coordinated fashion — they work in a blogosphere — producing large amounts of content in a student-owned space. At least two issues emerge because of the number of posts; assessment can become overwhelming for the instructor, and finding the relevant content for the students. Promotions, as a feature, have the virtue of being peer-created content that does not increase the instructor's workload to produce. The issue concerns the value of promotion data: Does the promotion activity identify high quality with minimal instructor involvement?

2 PROMOTIONS AS A SOCIAL ACTIVITY

In a traditional classroom, a student typically only has access to verified high quality content like books, articles, and lectures supplied by the instructor. Online there might be videos, more articles, and tutorials by other professionals. Other sources of content are peers, but here the quality of the content is mixed. Offline peer content can be produced by study groups, group projects, and classroom discussions. Online peer content can be produced in a knowledge community that uses technologies to enable content creation and distribution online (e.g., wikis, blogs, forums, and chats). The higher quality peer-created content can serve as a valuable resource for learning, but traditionally separating the good from the bad is the province of the instructor, which, for large amounts of content, makes it prohibitively time-consuming and thus a prime target for learning analytics.

Giving students access to each other's contributions is the basis for a knowledge community (Scardamalia & Bereiter, 1994), with each contribution representing a different, but valid, viewpoint that provides scaffolding for others to build upon and improve. Individual perspectives are shared, developed (Stahl, 2003), and composed (Suthers, 2006). Different contributions and examples on the same subject are multiple representations, which are valuable for learning (Ainsworth, 2006), and can lead to higher order thinking (Ellison & Wu, 2008; Philip & Nicholls, 2009). Depending on the technology that mediates peer collaboration, there will be different benefits. For example, a blogosphere is especially valuable as a resource for co-reflection among the students, which is very helpful for learning (Deng & Yuen, 2011).

Students produce content of mixed quality; ideally, the best content is foregrounded in some way. One solution is to have the instructor and graders identify the quality content for the rest of the class. However, as the size of the class increases, the cost of finding good content, without any automated assistance, becomes quite labour intensive — perhaps prohibitively so. An alternate solution is to have students themselves locate and label important content as they read. In the studies presented in this paper, students can attach likes and badges to each other's contributions. Collectively these features can be called promotions: the reader considers the content noteworthy and wants to label it publicly as such for the community. If the social data is reliable, quality examples will be identified, which can subsequently be used as a basis for comparison and self-assessment by other students. This scheme potentially makes the large quantity of content being produced more manageable for the students. It also can provide valuable feedback, as it potentially both recognizes and confirms the value of an individual contribution.

In this article, we present a social feature (promotions) to filter for quality content. The purpose of this feature is to have students collectively identify content that is noteworthy and helpful when creating new material. These promotions are anonymous and persistent, making them very useful for later exploration by assisting navigation. The instructor can also use the promotions to classify students based on their reading and writing abilities. This leads to a more accurate sorting of the quality content, beneficial for both the students and the instructor. The main result is that analysis of the learner-created content verifies that student promotions can be used as a basis for identifying high quality material in the blogosphere.

The findings are several-fold: 1) students actively promote content; 2) promoted content tends to be of higher quality than non-promoted content; 3) students use the promotion data to navigate; 4) some students are reliably better at assessing quality than others, and they can be identified. Findings like these confirm the value of promotions as a learning analytic metric, adding a new dimension to the middle space generated by students engaged in loosely coordinated learning activities. The feature is simple to use, and easy to implement and understand.

3 BACKGROUND

The web has changed the way we do many things, Web 2.0 even more so (O'Reilly, 2007). It all began with a proposal on how to manage information more accurately, how to prevent the loss of information (Berners-Lee, 1989). This has direct impact on how the web is engineered and how useful it can be for education, which is in the business of creating and distributing information. Very soon researchers, educators, and science fiction writers started thinking about accessing human "knowledge" with notebook-sized computers (Kay, 1991) and creating computer supported (intentional) learning environments (CSILE) (Scardamalia & Bereiter, 1994).

One such learning environment is blogs, which are widely used for various educational purposes (Deng & Yuen, 2011; Ducate & Lomicka, 2005; Williams & Jacobs, 2004). Frequently blogs are used as an open journal (Zagal & Bruckman, 2007) or a tool of reflection (Deng & Yuen, 2011; Nardi, Schiano, Gumbrecht, & Swartz, 2004). Student blogging has been reported to have positive effects on, for example, on the following: higher order thinking skills (Ellison & Wu, 2008; Philip & Nicholls, 2009); knowledge sharing and reflection (Du & Wagner, 2006; Hong, 2008; Luehmann & MacBride, 2009); the learning process beyond the classroom (Ferdig & Trammell, 2004); sense of ownership (Hong, 2008); sense of community as measured by the community dimension (Deng & Yuen, 2011); and identity development (Luehmann, 2008). Some have cautioned that just having students blog does not automatically guarantee positive learning outcomes (Divitini, Haugalokken, & Morken, 2005; Krause, 2004).

There are two basic types of blogs: a community blog where the whole class posts to a single blog or a collection of individual, student-owned blogs. In the case studies presented in this article, each student has her individual blog within a single closed community, thus combining the two basic types. When viewing the front page of the class, it looks like a community blog where posts by all members are listed in reverse chronological order — this creates a sense of community. Alternately, the blogosphere can be accessed through the contributions of a single student: each student has her own profile page that lists all of her posts — this enables students to develop their identity and feel a sense of individual ownership.

When we say community, we must recognize the importance of the social aspect. Knowledge creation as a social product (Scardamalia & Bereiter, 2006; Scardamalia, Bransford, Kozma, & Quellmalz, 2012) is complicated and vast; it bounds both formal and informal learning environments as well as networks and communities of people. Ferguson and Buckingham Shum propose five categories of social analytics: 1) social network analytics, 2) discourse analytics, 3) content analytics, 4) disposition analytics, and 5) context analytics. Content analytics is classified as being "a broad heading for various automated methods used to examine, index, and filter online media assets for learners" (Ferguson & Buckingham

Shum, 2012). In this article, we examine learner-created promotions (i.e., content analytics) and propose that they can be used to filter for high quality content in a blogosphere.

Given the popularity of educational blogs, understanding some of the social features involved will have significant value. A knowledge-building community is not simply the text of posts and comments; there is so much valuable data for all involved in the learning process. Exploring how many posts students read has been found to be informative (Gunnarsson & Alterman, 2012). Meta-data (i.e., promotions, tags, et cetera) produced by the students can increase awareness and involvement. Counting comments, counting external links, login data, load time, and many other features (as listed in Kargar & Azimzadeh, 2009) can all help make sense of the online community.

Blogging is a simple activity requiring little technical skill (click buttons to write, publish, browse, read, or comment). Blogs are versatile and widely used for various purposes (Nardi et al., 2004). Different methods of measuring content quality have been applied to blogs (Huang, Cheng, & Huang, 2009; Ulicny, Baclawski, & Magnus, 2007), such as search and clustering algorithms (Ulicny et al., 2007); voting and automated methods based on various quality measures (Kargar & Azimzadeh, 2009); classification of different types of blog posts using machine learning techniques (Ni, Xue, Ling, Yu, & Yang, 2007); and using natural language processing (Rubin & Liddy, 2006). New features have been added to blogs attempting to enhance learning such as dynamic learning maps (Wang, Huang, Jeng, & Wang, 2008) and awareness graphs (Ferguson, Buckingham Shum, & Deakin Crick, 2011).

The blogosphere gives students the opportunity to collaborate outside of class in a loosely coordinated fashion (Alterman & Larusson, 2013). Multiple conversations emerge on a single topic, enabling greater coverage and diversity with reduced coordination costs. Students publish to the rest of the class, providing multiple examples of work in a commons of information (Benkler & Nissenbaum, 2006; Bruckman, 1998; Scardamalia & Bereiter, 1994). Each student has her own blog but the collection of individual blogs make up the knowledge community. The collaborative effort of the students is not jointly focused to solve a particular problem but is rather a loosely coordinated collection of representations that together form a collective solution. Multiple conversations emerge from a single topic without the need to converge. These conversations persist over the course of the semester giving students a variety of resources to reference and refine.

Blogging enables students to work together, even though they participate at different times from different places. Together the students explore the solution space more widely in their individual posts but each contribution is examined less closely. In a joint focus space, where students discuss each contribution as it is made, that mechanic of working together limits the number of contributions that can be considered, but increases the warranty of contributions that survive in the problem space. With a loosely coupled learning activity, like blogging, the situation is reversed. Larger numbers of contributions become a part of a persistent content space and additional work is required to provide warranties that rate the quality of a contribution. Promotions can potentially provide a crowd-sourcing solution to this problem: students' subjective assessments of high quality content can provisionally provide a cost-effective mechanism for creating warranties and input into a recommender system (Lü et al., 2012).

4 TWO CASE STUDIES

The two cases for data analysis come from two classes: the first of 107 students and the second of 50 students. There was a mix of undergraduate and master's students in both courses. The first class had

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

weekly writing assignments completed in a class blogging environment. There were three kinds of assignments: 1) to write an editorial based on a book read in class, 2) to write a review of the book, and 3) to write a reflection on cross-cutting themes between the assigned readings. Students were encouraged to post drafts before the deadline, so they could collaborate as they completed their work on each assignment. For this class, students could promote by giving a like or by assigning merit badges for various exemplary features of a post.

We used data collected from the second class (50 students) to confirm the key results of the first study; despite a smaller set of promotion types, the key results were verified. In the second class, the students had weekly assignments where they practiced the methods and techniques taught in the class. Again, students could post drafts and revise their own homework up until the deadline. For this class, there were no merit badges, but the students could give anonymous likes to the work of other students.

5 CASE STUDY ONE: WEEKLY WRITING ASSIGNMENTS IN AN INTERNET & SOCIETY COURSE

The first case study was done in a class on Internet & Society (I&S). During the semester, students read four books on the social impact of the Internet and had weekly writing assignments to be completed in the blogosphere. Each student had her own blog, composed of multiple posts she authored. Students were encouraged to leverage each other's work; to read freely throughout the semester in the blogosphere, joining conversations that struck their interest; and to use the blogosphere as a resource to improve their own posts. Just browsing in the blogosphere and looking at other students' work has tremendous educational value. While working on an assignment, reviewing the posted work of other students was allowed. It was also permissible to revise posts, and revise again, up until the deadline. In this manner, the blogosphere is a platform for *peer tutoring*, *peer assessment*, and *collaborative learning* (Topping, 2005). In addition to the blog assignments, the students also had weekly assignments to provide peer comments and assessment "officially" on the posts of other students.

For each book, the students had to write both an editorial post and a book review post. Each post was required to be between 600–750 words in length; in practice, many of the posts were longer than that. The editorials required the student to review an issue raised by the book and then take a position, either expanding on the argument of the book with examples or presenting counterarguments. Each review explicated the central argument of the entire book. Students explained the key points, referenced (at least three) editorial posts of fellow students as support for their analysis, provided additional examples, and argued for (or against) the core argument of the book. In the latter part of the semester, students were required to write three reflections synthesizing material from two or more books and referenced posts written on the books. A requirement for these reflections was to reference at least three editorials or reviews written by other students. The same comment and assessment process also applied to the reflections.

Each student was required to write two comments and give two peer assessments each week (400 words in total). After the assignment due date, two posts were randomly assigned to each student on which she would officially comment, and another two other posts for which she was responsible for giving a peer assessment using a 3-point questionnaire form. The peer comment was posted under the student's user name but the peer assessment was anonymous. The comment was public but only the author of the post had access to the results of the questionnaire; comments and assessments were to be both "thoughtful" and "judicious." To ensure quality, both forms of feedback were evaluated by a grader.

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

Students were also encouraged to make additional comments, respond to comments on their own posts, and anonymously promote posts by giving them likes and badges. The additional activities were not directly graded, no feedback was given about promotions being good or bad, but promotions were considered part of the participation metric in the overall grade.

Because of the specific requirements of the blogging assignment, the students got plenty of feedback on their contributions to the blogosphere. The comments and peer assessments produced by other students were one form of feedback. Grades and publicly displayed gold stars for excellent work, given by the TAs, were a second form of feedback, and the peer promotions (badges and likes) were a third kind of feedback.

The course assignments were designed to be used as reference material later in the semester. Students not only had reason to review their peer work in the current assignment but, because of the design of the review and reflection assignments, they also had a reason to review past assignments. This potentially meant each assignment was more valuable because it was not just deliver and forget; it had future value for the student and the class as a whole.

5.1 What is the Issue?

Students produced a huge amount of text (600–750 words) on a weekly basis. The official comments on the work of two other students after the deadline for the post were a minimum of 200 words. For a class of roughly 100 students, this meant 300 contributions per week (one post and two comments per student). Or to put it differently, each student was producing 1000 words per week, which meant a minimum of 100,000 words per week produced by the entire class.

Because students were collaborating as they wrote each post, the merits of finding better quality material are obvious. High quality content, by definition, is easier to understand, easier to learn from, and easier to expand upon. There was also additional value in peer promotions, the results were used to gauge the reading and writing ability of the student. Both the reviews and reflections required students to cross-reference other posts, meaning that the writer paraphrases or quotes from another student's post and provides a hyperlink to it. The reviews were required to refer to at least three editorials. The reflections were required to refer to three other posts, either reviews or editorials. Additionally, although not required, many students included cross-references in their editorials.

Two factors contribute to the selection of a post to cross-reference: 1) the student needs to find a post relevant to her argument in her own post (the search mechanisms provided by the blogging environment supported this activity); 2) even after the initial search, many posts are relevant and useful to a student's post — ideally, the student wants to find not just a relevant post, but also a good one. Thus, finding good content was important while writing any of the three types of posts. Do student promotions provided effective feedback to guide the selection of good content? The results indicate that they do.

For the instructor, the enormous amount of content being created is problematic for a number reasons. The advantage of a student-owned space is that it helps students to open up and explore the material among their peers. Ideally, some guidance is provided to help students orient themselves towards the better content in the blogosphere, thus improving their overall peer discussion. Given the amount of material, depending on the instructor to provide all of this guidance is problematic because the work it would entail is enormous. For example, one mechanism for the instructor to provide direction is to give

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

gold stars to the best work for each assignment. The gold stars come after the completion of the assignment, providing a standard to which all students in the class can compare themselves. But identifying the gold star material in a class of 100+ students is not a trivial task. Can the analysis of the peer activity data, the promotions of the students, provide some of this guidance? The results from this study say yes.

5.2 More about the Class: Grading

There were six graders for the I&S class; this amount of grading support cannot be assumed for a typical class. Posts were graded differently than comments and peer assessments. Three graduate students graded the posts; a third of the posts were randomly assigned to each of the graders after the due date of each homework assignment. The grading was done by filling out a 6-item questionnaire. Each homework type (editorial, review, and reflection) had different questions. For example, the grading questionnaire for an editorial post was:

1. The issue is clearly explained.
2. The opinion is interesting and substantial.
3. The references are relevant to the argument.
4. The post is well written.
5. The post demonstrates understanding of the subject of the book.
6. Overall grade for this post?

Each question could be given a value from 0 to 3. Zero means that part was “not completed” (for example no references in the post would yield a 0 for that part of the questionnaire), 1 is “not good,” 2 is “good work,” and 3 “exceeds expectations.” Students were told to expect a 2 to be a good passing grade. Grading was done inside the blogosphere but in a special grading view where only the questionnaire and the text of the blog post were visible. The graders could not see the blog post’s comments, likes, or badges without intentionally browsing to that blog post in the standard view. This “blind review” was done to minimize any bias the post’s comments and promotions could have on the grader.

The comments were graded by three undergraduate students. The comment and peer review forms were simpler — just asking if the comment or peer review met expectations — the graders could give a 0 for “not completed,” 1 for “not good” and 2 for “good work.” Grading did not overlap between graders within the same assignment. Posts, however, were randomly assigned to a grader so that over the semester each student had assignments graded by all graders.

Periodically the head TA would give out gold stars to the high quality posts for each assignment. The grades of gold star posts varied a bit, although all were of high quality (graded from 14 to 18) some provided exceptional examples even if the post as a whole did not warrant the highest possible grade. The purpose of the gold stars was to identify high quality content “officially” for the community as a whole so that students could compare it to their own. Sometimes the gold stars were given out shortly after the assignment due date and at other times much later. A total of 80 gold stars were given throughout the semester.

The average number of person-hours spent grading each week was 31 hours; over the entire semester, the total was roughly 350 hours. For each assignment, the instructor would meet weekly with all six graders to give an overview of how to grade the most recent assignment. These meetings typically lasted around an hour (totalling ten person hours per week). From interviews with the graduate graders (who graded the posts), we learned that the average time spent on grading posts was around four hours per grader (twelve hours per week total). The average time spent grading comments and assessments was about three hours per grader (nine hours per week total). Clearly, this is an enormous amount of work and many courses will not have sufficient resources to commit to this level of grading. Being able to simplify some of the process of identifying good content, providing feedback, and evaluating student work would have tremendous value.

5.3 More about the Class: The Technology

The blogging environment was developed at Brandeis University over a number of years; the current version is a recent rebuild by the first author of this paper, which has already been used in several classes. We refer to this type of blogging as social blogging because students blog collaboratively, as individuals within a group blogging system, and the blog includes several social features, like promotions.

Students may browse content in the blogosphere using several different views. Most of the time, students would start a session by viewing the front page, which lists of the most recent posts. From there, a student could switch to an alternate view. Examples of available views include the following: all posts on a specific book, all posts by an individual student, all posts with a specific type of promotion, all posts receiving a gold star from the TA, and all posts the student had commented upon. It is also possible to do a keyword search to retrieve a list of all posts containing the search term. Regardless of the view, posts are always listed in reverse chronological order, with the most recently added posts at the top of the page. For each post in any view, its entry includes the name of the author of the post, the title of the post, the type of post (editorial, review, or reflection), the book(s) the post discussed, the number of comments the post had accrued, and the number of promotions it had received. Students could preview the first paragraph of a post by hovering over its title; they could then click on the title to view the post in its entirety, as well as any comments made on the post.

5.4 Student Promotions in the Blogosphere

While browsing in the blogosphere, students could promote posts by assigning them merit badges and likes. Liking a post means exactly that — the student supposedly liked it. Merit badges are a more specific type of a post promotion. Instead of clicking a button to say you like the post, you can click a button to say that you liked it because it was, for example, “nicely written.” Students could only give each post one like and one badge of each type. Each student could, in practice, give a post one like and six different badges. The badges were:

- Good example: The post contains an interesting example or case as a basis for its argument.
- Good question: The post raises an interesting issue.
- Nicely written: The post is well written.
- Good argument: The post makes a persuasive argument.
- Good references: The references are interesting and relevant.
- Good summary: The post provides an accurate and succinct summary of an issue or the book.

There was incentive to identify high quality posts in the blogosphere — earlier posts were referenced in later assignments. By finding high quality posts that clearly explained a topic in one of the books, the student could more easily build her argument when reviewing or reflecting on the central issue of the book. Reading poor quality posts would not be as helpful. Searching for good quality posts takes work, but also depends on a student's ability to identify what is a good post. The data will show that not all students were good at assessing quality.

5.5 Evaluation: The Data

The last assignment of the semester is left out of the data, as it is an outlier. Student participation in the last assignment was very different from the previous ten. This is most likely due to the fact that the semester was ending and students were busy finishing up their work in many classes.

The data compiles the promotion statistics of 92 students. The 15 students filtered out did not participate fully in the blogosphere and there were several possible reasons for excluding their data: only completing the first one or two of the 11 assignments, incomplete assignments, not participating in the blogosphere, or finishing most of the work after the end of the semester. For each of these students, their data was removed from the set.

5.6 Evaluation: Overview

All of the posts were graded, during the semester, on a scale of 0 to 18. For the purposes of this evaluation, posts with a score of 14 to 18 are high quality posts, 9 to 13 are average quality posts, and below 9 are poor. An outline of the evidence presented in the evaluation is shown in the *Summary of Evaluation* below. The utility of student promotions depends on several factors. For this scheme to work, students must actively promote during the semester, and the content they promote must in general be of high quality (evaluations 1&2). This information is useful for students only if they use promotions to help navigate in the blogosphere (evaluation 3). If promotions can be used to identify better content, they can potentially be used for highlighting content — as support or replacement for the identification of gold star material — and therefore, improving the reliability of promotions can greatly improve their functionality. How can this be done? One way is to separate the good promoters from the bad promoters. The evaluation presented explores whether promoters whose evaluations are reliable can be differentiated from those whose evaluations are more random (evaluation 4). The evaluation also explores how promotions and the content quality changes during the semester (evaluation 5).

5.6.1 Summary of Evaluation

Evaluation 1. Do students promote posts as they read in the blogosphere?
Evaluation 2. Are the promoted posts, on average, of higher quality than the average blogosphere post?
Evaluation 3. Do students use promotions to navigate in the blogosphere?
Evaluation 4. Are some students more reliable evaluators of quality content than others?
Evaluation 5. Promotions and post quality during the semester.

The data shows that students produce many promotions, many posts do not receive promotions, and the posts that receive promotions vary in how many promotions they receive. The average grade of promoted posts is significantly (p-value < 2.2e-16) higher than that of non-promoted posts. Some students were better than others at promoting, and as the semester progressed, the quality of the content in the blogosphere increased.

5.7 Evaluation 1: Promoting Content

Under what conditions are student promotions useful? If the promotion feature is rarely used, they risk getting lost. If there promotions are too many, they become meaningless. For the entire class, 89.7% of

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

students used the promotion feature at least once. In the first assignment, many students used the feature, but the number of students who promoted declined in later assignments. During the whole semester, the average student promoted 6.77 posts from 3.75 assignments. For each assignment, 37.8% (an even third if the first assignment is excluded) of the students gave out a promotion. During the first assignment 78.2% of students gave a promotion and during other assignments between 50% and 16.3% (in the last assignment) gave promotions.

The proportion of the number of promotions per assignment follows a similar trend, each of the ten assignments received, on average, 10% (8.5% excluding the first assignment) of the total number of promotions given throughout the semester. The first assignment received 23.7% of all promotions with other assignments getting between 11.9% and 4.9%.

Around 80% of promotions were made before the official commenting period ended (and thus before the assignment was graded). This makes them very useful while students are working on the same assignment and helpful when students are searching for quality content to reference in later assignments. Quite a few promotions were available before the assignment deadline, giving students access to higher quality content, on average, to help understand the problem. Later, when reviewing prior assignments, the search function can be used to find relevant content and then the promotions quickly indicate the best posts.

Later assignments were less likely to be used as reference material because they were designed as reflections on the first eight assignments. Students are also busy with exams at the end of the semester. Both of these aspects could have a reducing effect on promotions. Despite the outliers, the first and last assignments, we believe both the rate and number of promotions given throughout the semester were good.

5.8 Evaluation 2: Quality of Promoted Content

This evaluation seeks to answer the core question of this article: Does the promotion activity identify high quality with minimal instructor involvement? In order to answer this question we ask what was the quality, in terms of grades, of the posts that received a promotion? We answer this question in three parts:

- Higher quality posts got more promotions. (See: *Number of promotions per post*)
- Higher quality posts were more likely to be promoted. (See: *Promotion hit rate*)
- All types of promotions were useful (See: *Different types of promotions*)

If higher quality posts get more promotions and if higher quality posts are more likely to be promoted, then the subset of promoted posts is of higher quality. How much we answer in the following evaluations.

5.8.1 Number of promotions per post

Figure 1 shows the total number of promoted posts and the total number of posts not promoted, both grouped by grade. It also shows the number of total promotions for posts of each grade. The data shows an interesting increase in the number of promotions on content with high grades (see the shaded region in Figure 1). In other words, all posts can receive promotions, but posts with high grades get significantly more promotions than posts with a lower grade (p-value = 0.002045) (see Figure 2).

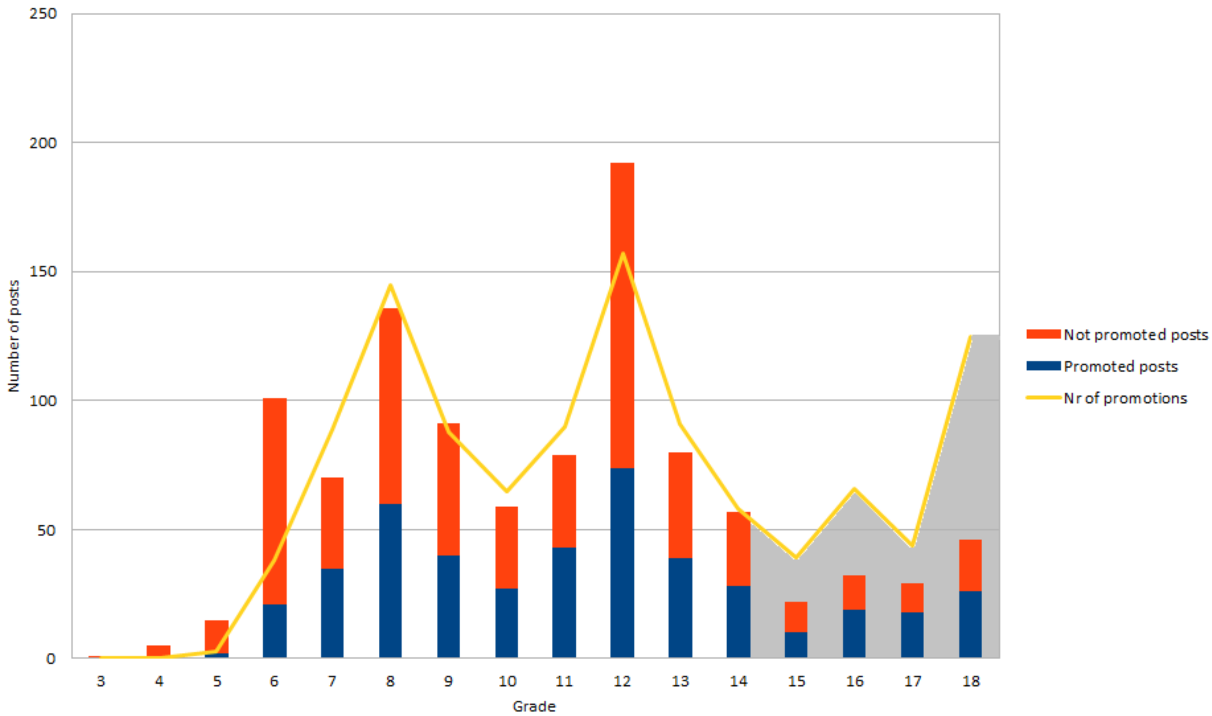

Figure 1. Number of posts and promotions per grade

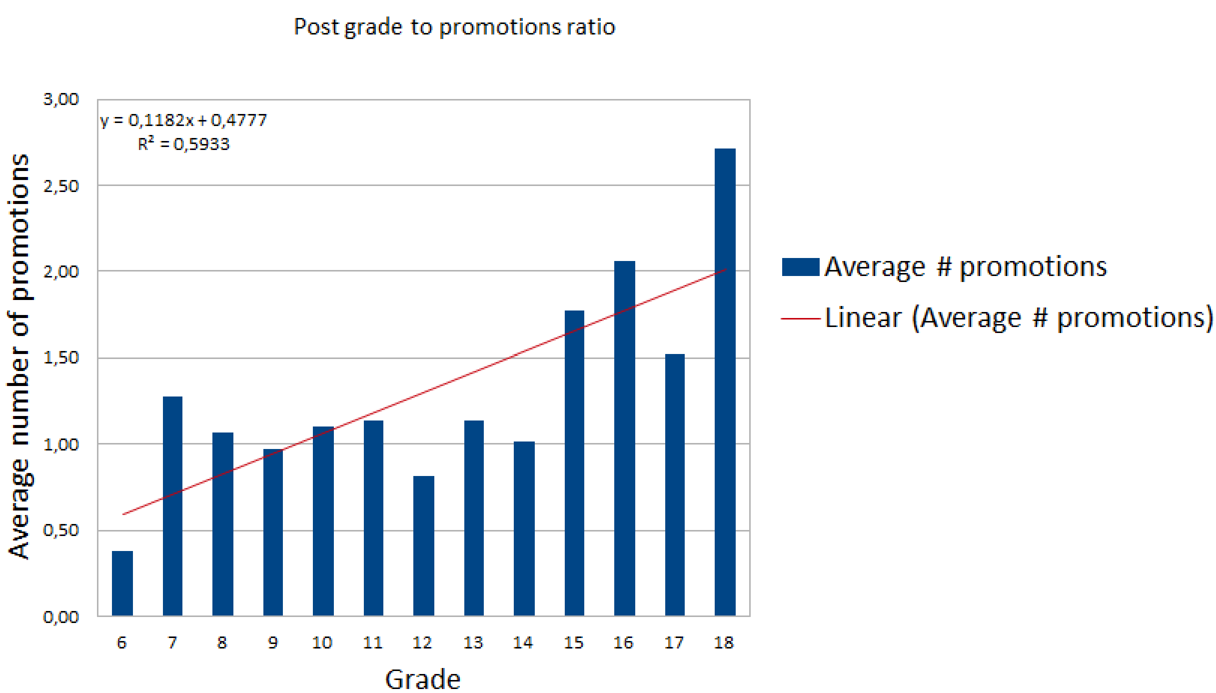

Figure 2. Average number of promotions per grade

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

5.8.2 Promotion hit rate

Were high quality posts more likely to be promoted than lower quality posts? Only 36.2% of all low quality posts (graded 8 or lower) were promoted. This means low quality posts have a 36.2% *hit rate*. The hit rate increased to 47.3% for posts of average grade (graded between 9 and 13), and high quality posts (graded 14 or higher) had a hit rate of 54.3%. The difference in hit rate average is only significant between poor and high (p-value = 0.033) but the distinction between poor and average is higher (p-value 0.11) than between average and high (p-value = 0.2165); students seem to identify average posts over poor posts more easily than high quality posts over average quality posts.

Comparing the average grade of posts with and without a promotion paints a clearer picture. The distribution of promotions between high and low quality posts is significantly different (see below), indicating that students can, on average, identify high quality peer content.

5.8.3 Different types of promotions

As mentioned before, we named the activity of liking and giving a badge as *promoting*. We analyzed the data for each type of promotion (see Figure 3). The average grade for all posts during the semester was 10.7 (see horizontal line for average class grade). With one exception, all types of promoted content had higher grades than the average class grade.

There were six different types of promoted posts in the blogosphere: 1) posts that were both liked and badged; 2) posts that were liked and could have badges too, 3) posts that were liked and had no badges, 4) posts that got badged and could also have likes; 5) posts that were only badged and had no likes; and 6) posts that received no likes or badges. Comparing the five groups that had promotions to the group (6) that had no promotions:

1. Posts that had 1+ likes and 1+ badges had a significantly higher grade (p-value < 2.2e-16).
2. Posts that had 1+ likes and 0+ badges had a significantly higher grade (p-value < 2.2e-16).
3. Posts that had 1+ likes and 0 badges had significantly higher grades (p-value = 8.181e-09).
4. Posts that had 0+ likes and 1+ badges had significantly higher grades (p-value = 0.006887).
5. Posts that 0 likes and +1 badges did not have significantly higher grades (p-value = 0.3264).

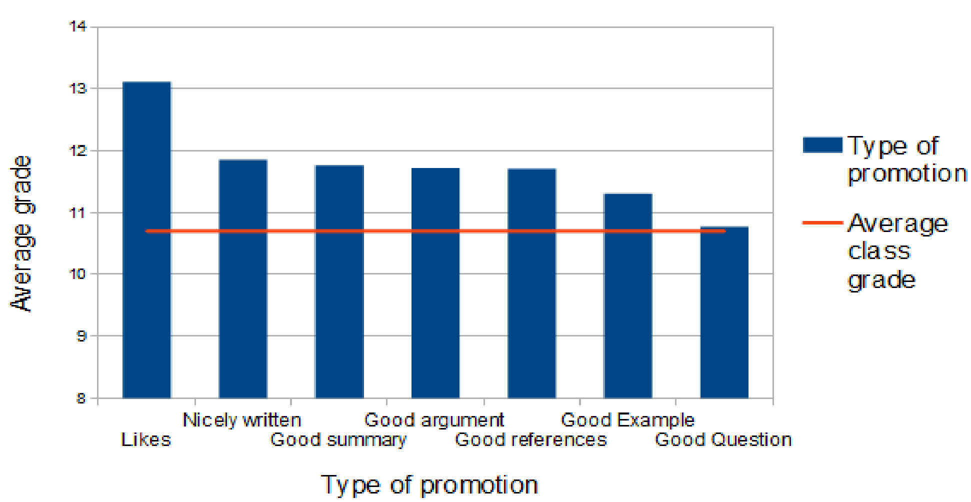

Figure 3. Average grade per type of promotion

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

Badges do not add anything to the predictive power of likes, statistically. The numbers, however, are different for each badge, as shown in Figure 3. Badges can prime posts for likes or provide additional meta-information of value.

5.8.4 Summary

Combining all of these results (keeping in mind that 80% of all promotions happen before assignment deadline) — high quality posts get more promotions, the promotion hit rate is higher for quality content, and all types of promotions are useful — this means promotions are a powerful method for identifying quality material: students can successfully identify high quality content.

5.9 Evaluation 3: Navigating Promotions

On average, each post was read 22.9 times during the semester; gold star posts were read, on average, 44 times. The promoted posts were, on average, of higher quality, but were the students guided by the promotions to read those posts more frequently than the average post? Posts without any promotion feedback were read 16.8 times. Promoted posts got 33.5 reads — twice as many! Not only are students promoting quality content but they also act on the promotions. The community as a whole steers itself towards quality content using the promotions feature.

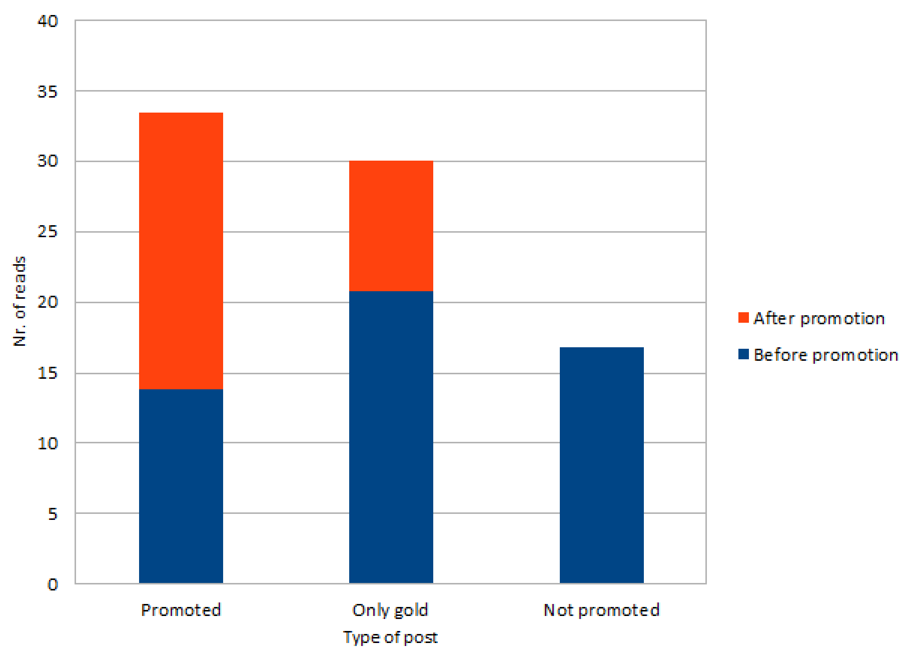

Figure 4. Average number of reads before and after promotion.

Some posts only get likes, some posts only get badges, other posts (25 out of 80 total gold stars) get only gold stars, and finally many posts get no promotions or gold stars. All posts get about the same average number of reads (~15) before any promotions or gold stars. After a post gets a promotion its' exposure is greatly increased and it accrues a lot more reads from students (see Figure 4). Gold stars are given after the deadline, shortly after grading (usually about a week or two after deadline), and do increase the number of reads a post receives. Why a gold star only post has more reads before it gets a gold star is interesting, a possible reason is because they are cited in other students' posts (see *cross-links* below).

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

These posts also get fewer average reads than promoted posts but that is most likely because gold stars are awarded later.

How students use promotions can be seen in students' *cross-links* (links to other students' posts within their own post). On average a cross-linked post has 1.93 promotions whereas a post that was not cross-linked only has 0.88 promotions, on average. The quality, measured by grade, of the cross-linked posts is also significantly higher (p-value = 4.778e-12). Students clearly use the promotions to navigate and use them as a source to improve their own work.

5.10 Evaluation 4: Promotion Predictions

Are some students better at promoting quality material than others? We wanted to know which students were reliable: both among students who consistently promoted high quality content and students who tended to promote poor content. Figure 5 shows that some students promote high quality posts, other students mostly promote low quality posts, but the majority of students are less predictable. A third (33) of the students regularly promote posts (5 or more assignments); a quarter (23) of all students consistently promote high quality posts (labelled good promoters in Figure 5). Out of those 23, seven students are also regular promoters, averaging 16.7 promotions throughout all assignments.

In Figure 5, good promoters refers to students who on average promoted posts that received a good grade (on or above the top line), and poor promoters are the students who promoted posts that received a poor grade (on or below the bottom line). Obviously, the posts promoted by students who reliably promote high quality material can be used to feature content in the blogosphere. In terms of assessment, the users classified as poor promoters present an interesting case. First of all, getting them to stop promoting content will immediately increase the average quality of promoted content. Secondly, this information can be used by the instructor as a teaching opportunity — these users lack an important critical reading skill.

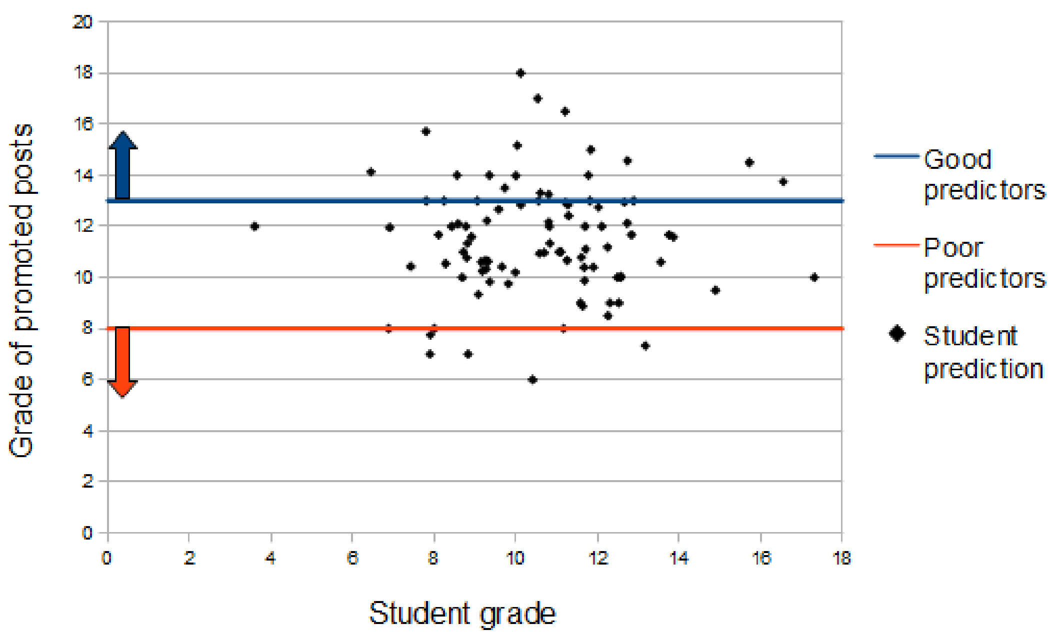

Figure 5. Average grades vs. average promoted grades

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

The distribution is interesting in that the grades of students and the grades of the posts they promote do not correlate. One might expect that students who produce high quality content to be the same ones who can identify quality content by promoting. However, some students who get high grades only promote average or low quality content. Perhaps, even more surprisingly, there are students who receive low grades on the posts they produce but are very good at identifying and promoting high quality content. This suggests that students are developing two different skills (that could be classified by drawing a line up and along the average grade (10.7) on both axes): 1) writing and 2) identifying quality content. A student can be good at both, either, or neither. One of the benefits of the promotion data for instructors is that it can be used to measure the writing and critical reading abilities of each student. We consider this an important finding, one that can make promotions a very useful feature in educational blogs.

5.11 Evaluation 5: Promotions and Post Quality During the Semester

The distribution of grades and accuracy of promotions did not stay the same throughout the semester. The number of good predictors increased slightly and the number of poor predictors decreased by more than half. At the same time, the number of poor quality posts decreased and the number of average posts increased. What this tells us is that the average quality of the content is increasing and that the students are becoming more accurate at promoting. Analysis of the data is not that conclusive though.

A three-factor ANOVA test of assignment number, likes, and badges indicated a negative trend in student promotions. While there was a slight increase in mean assignment grade over the course of the semester, it was not significant (F = 0.983, p-value = 0.452). There was a high negative correlation in the difference of marginal means (Pearson correlation -0.8, p-value 0.005). This can be explained by the change in the distribution of grades (see Figure 6), making it more difficult to identify high quality posts.

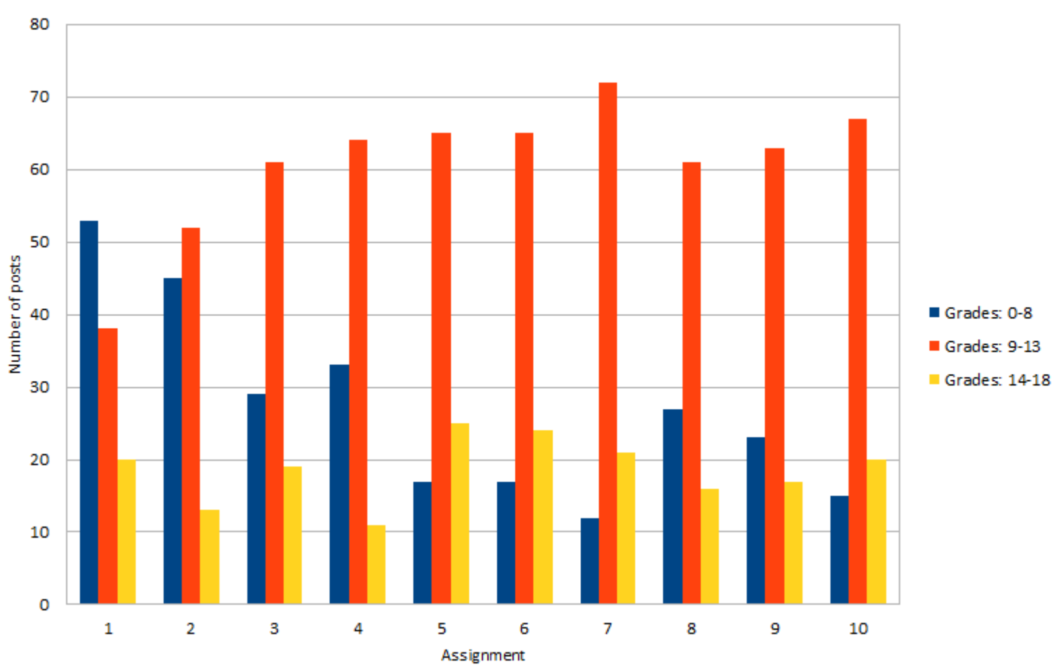

Figure 6. Number of posts per grade and assignment.

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

Remember from the grading rubric, a grade of 12 was considered a good post; high quality posts are considered to "exceed expectations." During later assignments, having a much higher number of average quality posts compared to poor quality posts means that it is much harder to differentiate between high quality and poor quality. The main difference between the group of students who promote high quality and poor quality is the reliability of good promoters. Figure 7 shows the number of people who could be called "consistent promoters"; at least half of their promotions were either of high quality or poor quality.

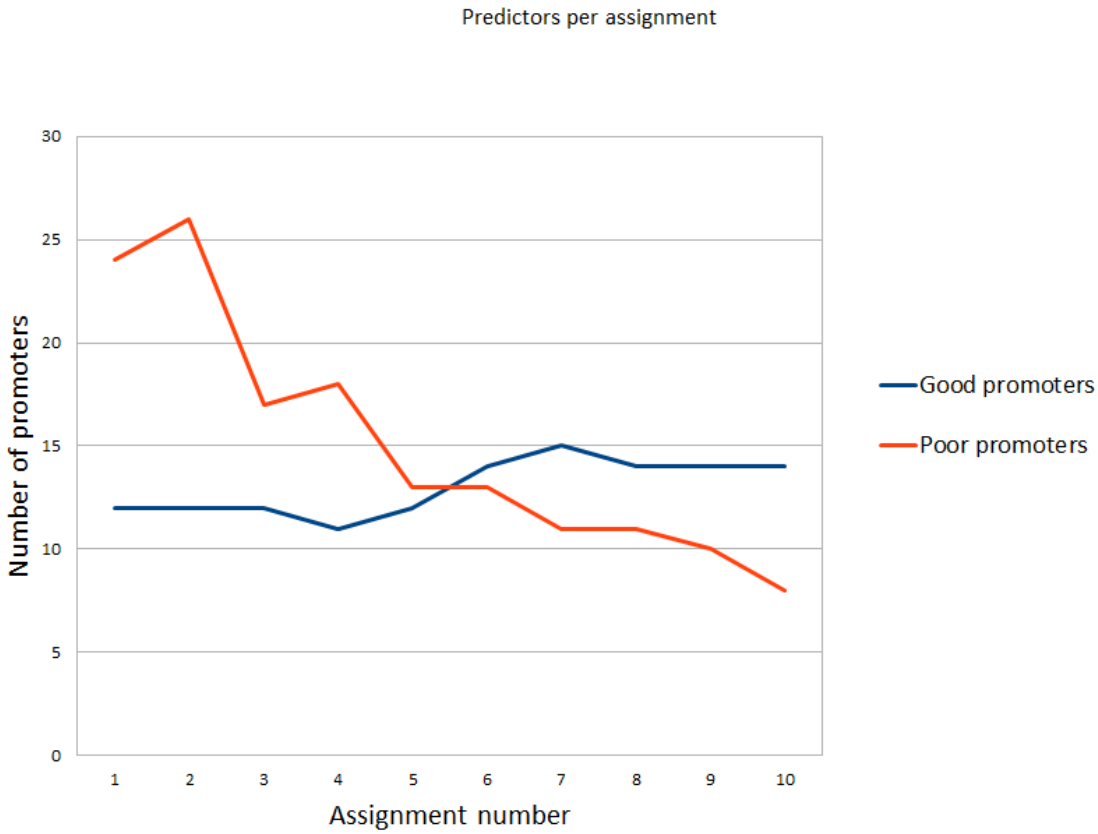

Figure 7. Number of consistent promoters.

Ideally, students improve their promotion skills over time, meaning that as the semester progresses students become even more accurate at identifying high quality content. Unfortunately, the results are mixed for this claim. The number of promotions went down over the semester while the overall quality of content went up. Students became less accurate with their promotions but their hit rate was slightly better.

There was considerably less poor content in later assignments and fewer students consistently giving poor promotions. However, the number of high quality posts was about the same as poor quality posts in later assignments, yet the number of reliably good promoters remained steady and exceeded the number of people who promoted poor posts in later assignments, even though the number of good and poor posts was about the same. This is an echo of the hit rate in evaluation 2: higher quality content has a higher promotion hit rate than lower quality content.

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

In an environment of contrasting content, in terms of quality, differentiation is easy. Exploring the hit rate between assignments provides an insight. During the promotion-stable part of the semester (assignments 2 through 9), there was a noticeable drop in the promotion hit rate for poor and average quality posts between the first half of the assignments and the second half. The poor quality hit rate dropped from ~35% to ~30% and the average quality hit rate dropped from ~50% to just below 45%. The high quality hit rate was the same at around 56%. While the overall quality of content improved, the hit rate of lower quality posts declined.

Early in the semester, the difference between high quality posts and poor quality posts was much larger than later in the semester. Notice in Figure 6, because of the pattern of assignment (editorial, review, editorial, review, et cetera), the improvement began to emerge with the second assignment of each type. With the overall improvement in the quality of posts, the discrimination between high quality and poor quality becomes more difficult to make, but also less significant; in other words, the average promoted post was always of higher quality than one that was not promoted. The change in overall quality is a good thing for the students, because reading better material is more productive. The goal of having students read better content is achieved with the overall improvement in content and by providing a smaller subset of higher quality posts through the promotion feature.

5.12 Summary of Analysis

Evaluation 1. Students promote posts as they read in the blogosphere.
Evaluation 2. Students accurately identify high quality content.
Evaluation 3. Students use promotions to navigate.
Evaluation 4. Students can be classified by the content they promote.
Evaluation 5. As the semester progresses, students have access to better content.

So, does this mean that the promotion activity identifies high quality with minimal instructor involvement? Because students actively promote, because the promotions are useful, and because we can identify good predictors and poor predictors, high quality content can be predicted based on who promotes it. Yes, the content students promote is of considerably higher quality than posts not promoted. In addition, the analysis provides hints that the promotion feature can, to some extent, be used to measure a student's critical reading and writing skills.

6 CASE STUDY TWO: WEEKLY ASSIGNMENTS IN A HUMAN COMPUTER INTERACTION COURSE

Data from another case study confirms some of the results of the initial study concerning the quality of promotions; that students actively use promotions and that they accurately identify high quality content.

The basis of the second study was a course on Human Computer Interaction (HCI) taught in a Computer Science department and cross listed in Psychology (Alterman & Gunnarsson, 2013). There was a mix of 50 undergraduate and master's students in the course, with weekly homework assignments. Homework was done in a blogging environment. The weekly assignments were blog posts where students either implemented or applied HCI techniques to various assignments. Each week there was an assignment deadline and a comment deadline. Students could, at any time, read, comment, and give likes. For this class, there were no merit badges.

Students were free to post drafts of their homework before the deadline, read each other's drafts, and revise their assignments up until the deadline. After the completion of each assignment, there was an official three-day commenting period; students were required to post substantial comments on the randomly assigned posts of two other students. As the semester progressed, some of the earlier work of the students became relevant to the term project. Thus students were to free browse, comment, and like the work of other students. The data shows that many students actively read in the blogosphere throughout the semester (Alterman & Gunnarsson, 2013); they regularly read each other's posts while working on each assignment and continued to read them after the completion of each assignment. The term project was purposefully designed in such a way that the content created in previous assignments could be used as source material. While writing the term project, students frequently browsed previous posts.

The same blogging environment was used in both classes, the I&S class had been upgraded to use badges and some minor UI (user interface) changes had been made. For the HCI class, the front page listed the most recent posts in reverse chronological order. The students could filter the posts by assignment, by author, or by using a search mechanism. Regardless of how the posts were filtered, the list of posts generated was always in order of the most recent, with each post summary including a count of the number of likes and comments. The content of a post could be previewed by hovering with the mouse over the title.

Grading of posts and comments was done by three graduate students. The grading scheme was simpler than that employed for the I&S class. Where for the I&S course the parts of the post were graded, for this class a single grade, on the scale of 0–3 was given to each assignment:

- 0 means that part was "not completed"
- 1 means "not good"
- 2 means "good work"
- 3 means "exceeds expectations"

Comments were graded on the 0–2 scale where the numbers meant the same as for post grades but comments could not exceed expectations. Each post was graded by a single grader; posts were randomly assigned to each grader for each assignment. The instructor and the graders had weekly meetings, before the grading for an assignment began. During the meeting, several assignments were graded collectively to coordinate the grading rubric for the assignment.

Students actively used the promotion feature; the average quality of liked posts was higher than posts that were not promoted (see Figure 8, which shows grades scaled between 0 and 1).

Table 1 compares some of the key numbers in this second study to those in the I&S study. The students in the I&S class did more promoting, but those promotions also include badges; liking an entire post is a more stringent criteria than liking some feature of the post, for example, if it had "good references." While a student might not be ready to give a post a like, she might consider giving the post a badge. For both classes, high quality material was more likely to be promoted than poor quality material, as measured by the hit rate. Some differences between hit rate numbers are undoubtedly a result of the difference between the grading and promotion schemes in the two classes. To repeat: in the HCI study, students could only "like," there were no merit badges, and grades were on a simpler scale, from 0 to 3. The merit badges in the I&S course were less accurate predictors of high quality content than the likes but they do contribute to a greater hit rate of high quality content. The simpler grading scheme and the lack of merit badges probably explain the lower hit rate of poor quality posts in the HCI class data set. So while the high quality hit rate in the HCI study might not look as impressive as in the I&S one, the low poor quality hit rate is impressive — it makes the subset of promoted posts less "polluted" by poor quality content than the promoted I&S content.

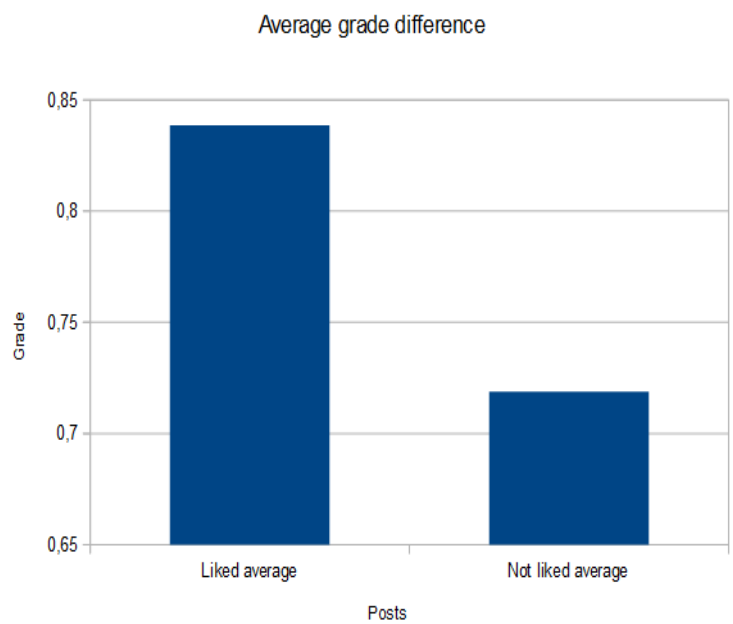

Figure 8. Difference between the average grade of posts that were "liked" and those that were not. Grades are scaled between 0 and 1.

Table 1. Comparing data from the two case studies.

	I&S	HCI
Promotions per user	6.77	4.1
High quality hit rate	54.3%	45.5%
Poor quality hit rate	36.2%	18%

To summarize, the data from the second case study confirms some of the key results of the I&S case study:

1. The students actively promoted content throughout the semester.
2. The promoted posts were, on average, of higher quality than the posts that were not promoted.
3. High quality content was more likely to be promoted and thus had a higher hit rate.

7 DISCUSSION

In our class, students get feedback through grades, comments, peer assessments, and promotions. Future work will evaluate the effectiveness of the promotion activity without the rigorous grading done in this class. If instructors were to grade fewer posts, would student promotion improve over the course of the semester as the data has shown? What if the only posts graded were the promoted ones?

Consider, as a possible grading scheme, for the instructor to grade only promoted posts. This reduces work for the instructor. Under this grading scheme, it would be possible to ascertain which students are good predictors, which students are less predictable, and which students do not participate. This level of effort on the instructor's part is sufficient to develop some kind of highlighting mechanism, possibly a replacement for the gold star feature. The obvious downside of this scheme is that some students would not be receiving feedback or grades from the instructor (or the graders). If needed, posts that were never promoted could be randomly sampled for grading purposes. With this additional work, the instructor would be able to monitor the progress of students from assignment to assignment, both as readers and contributors. Regarding feedback, students would be receiving regular feedback from the instructor (but not every week). They would also be receiving peer feedback in the form of comments and promotions, and have access to highlighted content to compare to their own work or to build upon. With this scheme, some of the posts that were not promoted but were of high quality would be discovered.

In a blogosphere, as a learning environment without the promotions described in this article, there is a possible bias for work posted either early or late. Do early posts garner disproportionately more attention and feedback? Do poor quality early posts have a negative impact on the discussion? Do later contributions suffer from the opposite problem: too little feedback and promotions from peers? Is the peer assessment requirement the driving force of promotions, without which there wouldn't be many promotions? The hope is that the quality of contributions gradually increases towards the due date as students read and update previous contributions based on more recent and collected content, and that the early post bias does not slow down this progress.

The blogosphere, with the promotions described in this article, potentially compounds the early and late bias. Based on the results, the promotions may have a positive impact on the early post bias by identifying the high quality early posts, hinting that those not promoted may need updating. In the early bias case, promotions may be the solution. The problem in the other direction — overlooking good quality posts contributed late — may require some scheming to remedy. One possibility is to leverage the random assignment of required commenting. During the commenting period students can, and do, promote content. If these late promotions are made by good promoters then they become prime candidates for further review as gold star contributions. But in either direction, the question of bias requires further study.

Suppose the blogosphere has only a few users that produce little content. Would promotions work differently with this number of students versus this content-amount dynamic? Since there is so little content, students who want to collaborate or need help can just read everything. Also, would a more intimate setting of only a few users trigger everyone to promote everything? In this case, the promotions are working differently as mostly a positive feedback mechanism supporting student engagement.

One part of the class was peer assessments. Generally, students did not give likes to the posts they were assigned to assess (only 10 likes from seven students); students were more likely, however, to give badges during assessments (58 different students on 213 occasions giving badges to 196 different posts). This amounts to just fewer than 6% of all likes and 21% of all badges awarded. Students only awarded a post with a promotion if they had assessed it to be of high quality; however, a brief exploration of the peer assessments indicated that they were not reliable. The interaction between peer assessments and promotions could be a topic of another analysis. For this article, assessments may have had a negative impact on the quality of badges, the difference in average grade between badges awarded during assessment and those awarded at other times was significant (p-value = 0.02434). The difference in means, while statistical, is only half a grade point. We cannot conclude that, without the assessments, badges would have been of a higher quality.

While badges do not add much statistical value to likes, they are numerous and as such of social value. Would likes be as predictive, for example, without badges? The difference in grade schemes in the HCI and I&S courses makes it difficult to say. The badges may add a level of assessment between a like and no like, thus making the difference between a like and no like greater than it would be without badges.

Do these results generalize beyond blogging? Suppose the students are tweeting (many contributions from either many or a few students). Does the notion of promoting make sense under these conditions? Perhaps, but for this kind of learning activity the middle space would be somewhat different, and consequently the analytics would clearly have to be adjusted. For various kinds of loosely coordinated learning activities, promotions hold promise, but the devil is in the details. Whether and how to convert the promotions into assessment and feedback would necessitate further study.

Overall, the results of this research are very promising but we do not assume that the promotion feature can be automatically applied to any learning environment and be a success, as explained by the middle space. However, the possibilities the promotion feature enables for learning analytics are of interest.

8 CONCLUSION

In this article we have explored the use of peer promotions as a tool to filter for high quality content. The data shows that students willingly use the tool and that the application of the tool provides the desired results — the promoted content is of significantly higher quality than content that is not promoted, and content that is repeatedly promoted is of higher quality than content that has fewer promotions. These results have been verified by two different case studies. Other results show that good and poor promoters can be identified. Both classifications of promoters have value: by focusing on good promoters, the reliability of quality assessment can be improved; by focusing on poor promoters, the instructor is in a better position to identify students who may be struggling.

Navigating a huge amount of peer-generated content in a knowledge community on a weekly basis is a lot of work. Not all content is of sufficient quality to be considered useful for all students. Being able to filter the content for quality would be invaluable. The social activity of promoting each other's material, if properly applied, enables students collectively to identify content that is interesting or of particular value. What the number of promotions might mean for a post will vary from one setting to another, but the promotion feature will always have social value.

REFERENCES

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198.

Alterman, R., & Gunnarsson, B. (2013). *The blogosphere as representational space*. *CSCL 2013 Conference Proceedings Volume 1: Full Papers & Symposia*, pp. 26–33. International Society of the Learning Sciences.

Alterman, R., & Larusson, J. A. (2013). Participation and common knowledge in a case study of student blogging. *International Journal of Computer-Supported Collaborative Learning*, *8*(2), 149–187. doi:10.1007/s11412-013-9167-2

Arnold, K. E. (2010). Signals: Applying academic analytics. *Educause Quarterly*, *33*(1), n1.

Baker, M., Hansen, T., Joiner, R., & Traum, D. (1999). The role of grounding in collaborative learning tasks. *Collaborative learning: Cognitive and computational approaches*, 31–63. Oxford, UK: Pergamon, Elsevier Science, 1999.

Benkler, Y., & Nissenbaum, H. (2006). Commons-based peer production and virtue. *Journal of Political Philosophy*, *14*(4), 394–419. doi:10.1111/j.1467-9760.2006.00235.x

Berners-Lee, T. (1989). Information management: A proposal. Retrieved 9 June 2014 from http://www.w3.org/History/1989/proposal.html

Bruckman, A. (1998). Community support for constructionist learning. *Computer Supported Cooperative Work (CSCW)*, *7*(1–2), 47–86. doi: 10.1023/A:1008684120893

Dawson, S. (2009). "Seeing" the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology*, *41*(5), 736–752.

Deng, L., & Yuen, A. H. K. (2011). Towards a framework for educational affordances of blogs. *Computers & Education*, *56*(2), 441–451.

Divitini, M., Haugalokken, O., & Morken, E. M. (2005). Blog to support learning in the field: Lessons learned from a fiasco. *Fifth IEEE International Conference on Advanced Learning Technologies — ICALT,* 5–8 July 2005, Kaohsiung, Taiwan.

Du, H. S., & Wagner, C. (2006). Weblog success: Exploring the role of technology. *International Journal of Human-Computer Studies*, *64*(9), 789–798.

Ducate, L. C., & Lomicka, L. L. (2005). Exploring the blogosphere: Use of web logs in the foreign language classroom. *Foreign Language Annals*, *38*(3), 410–421.

Ellison, N., & Wu, Y. (2008). Blogging in the classroom: A preliminary exploration of student attitudes and impact on comprehension. *Journal of Educational Multimedia and Hypermedia*, *17*(1), 99–122.

Ferdig, R. E., & Trammell, K. D. (2004). Content delivery in the "blogosphere." *T.H.E. Journal*, *31*(7), 12–20.

Ferguson, R., & Buckingham Shum, S. (2012). Social learning analytics: Five approaches. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge,* 29 April–2 May 2012, Vancouver, BC, Canada. New York: ACM Press. doi:10.1145/2330601.2330665.

Ferguson, R., Buckingham Shum, S., & Deakin Crick, R. (2011, December 14). EnquiryBlogger: Using widgets to support awareness and reflection in a PLE Setting. In W. Reinhardt & T. D. Ullmann (Eds.), *Proceedings of the 1st Workshop on Awareness and Reflection in Personal Learning Environments,* 11 July 2011. Southampton, UK**.** In conjunction with the PLE (Personal Learning Environment) Conference 2011. Retrieved from http://oro.open.ac.uk/30598/1/arple-2011_(006904v1).pdf

Gunnarsson, B. L., & Alterman, R. (2012). Predicting failure. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge: LAK '12* (p. 263), 29 April–2 May 2012, Vancouver, BC, Canada. New York: ACM Press. doi:10.1145/2330601.2330665.

Hong, W. (2008). Exploring educational use of blogs in us education. *US–China Education Review, 5*, 34–38.

Huang, T. C., Cheng, S. C., & Huang, Y. M. (2009). A blog article recommendation generating mechanism using an SBACPSO algorithm. *Expert Systems with Applications*, *36*(7), 10,388–10,396.

Kargar, M. J., & Azimzadeh, F. (2009). A framework for ranking quality of information on a weblog. *World Academy of Science, Engineering and Technology*, *56*, 690–695.

Kay, A. (1991). Computers, networks and education. *Scientific American*, *265*(3), 138–148.

Krause, S. D. (2004). When blogging goes bad: A cautionary tale about blogs, email lists, discussion, and interaction. *Kairos*, *9*(1). Retrieved 9 June 2014 from http://english.ttu.edu/kairos/9.1/praxis/krause/

Luehmann, A. L. (2008). Using blogging in support of teacher professional identity development: A case study. *The Journal of the Learning Sciences*, *17*(3), 287–337.

Luehmann, A., & MacBride, R. (2009). Classroom blogging in the service of student-centered pedagogy: Two high school teachers' use of blogs. *THEN: Technology, Humanities, Education, & Narrative*, *6*, 5–36.

Lü, L., Medo, M., Yeung, C. H., Zhang, Y. C., Zhang, Z. K., & Zhou, T. (2012). Recommender systems. *Physics Reports*, *519*(1), 1–49.

Nardi, B. A., Schiano, D. J., Gumbrecht, M., & Swartz, L. (2004). Why we blog. *Communications of the ACM*, *47*(12), 41–46.

Ni, X., Xue, G. R., Ling, X., Yu, Y., & Yang, Q. (2007). Exploring in the weblog space by detecting informative and affective articles. *Proceedings of the 16th International Conference on World Wide Web,* 8–12 May 2007, Banff, AB, Canada.

O'Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*, *1*, 17.

Philip, R., & Nicholls, J. (2009). Group blogs: Documenting collaborative drama processes. *Australasian Journal of Educational Technology*, *25*(5), 683–699.

Rubin, V. L., & Liddy, E. D. (2006). Assessing credibility of weblogs. *Proceedings of the AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs (CAAW),* 27–29 March 2006, Stanford University, CA, USA.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, *3*(3), 265–283.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97–115). Cambridge, UK: Cambridge University Press.

Scardamalia, M., Bransford, J., Kozma, B., & Quellmalz, E. (2012). New assessments and environments for knowledge building. *Assessment and Teaching of 21st Century Skills*, 231–300. Netherlands: Springer. doi 10.1007/978-94-007-2324-5_5

Siemens, G. (2011). Learning analytics: A foundation for informed change in Higher Education. http://www.educause.edu/library/resources/learning-analytics-foundation-informed-change-higher-education

Smolin, D., & Butakov, S. (2012, April). Applying artificial intelligence to the educational data: An example of syllabus quality analysis. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 164–169). New York: ACM Press.

(2014). Peer Promotions as a Method to Identify Quality Content. *Journal of Learning Analytics*, *1*(2), 126-150.

Stahl, G. (2003). Meaning and interpretation in collaboration. *Designing for change in networked learning environments: Proceedings of the International Conference on Computer Support for Collaborative Learning (CSCL'03).* Bergen, Norway

Stahl, G. (Ed.). (2009). *Studying virtual math teams*. New York: Springer.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge, UK: Cambridge University Press.

Suthers, D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 315–337. doi:10.1007/s11412-006-9660-y

Suthers, D., & Verbert, K. (2013). Learning analytics as a "middle space." *Proceedings of the Third International Conference on Learning Analytics and Knowledge — LAK '13* (p. 1). New York: ACM Press. doi:10.1145/2460296.2460298

Topping, K. (2005). Trends in peer learning. *Educational Psychology*, *25*(6), 631–645. doi:10.1080/01443410500345172

Ulicny, B., Baclawski, K., & Magnus, A. (2007). New metrics for blog mining. *Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security 2007*; SPIE Volume 6570, 65700I, 9 April 2007. Orlando, FL, USA.

Wang, K. T., Huang, Y.-M., Jeng, Y.-L., & Wang, T.-I. (2008). A blog-based dynamic learning map. *Computers & Education*, *51*(1), 262–278. doi:10.1016/j.compedu.2007.06.005

Williams, J. B., & Jacobs, J. (2004). AJET 20: Exploring the use of blogs as learning spaces in the higher education sector. *Australasian Journal of Educational Technology, 20*(2), 232–247.

Zagal, J. P., & Bruckman, A. (2007). GameLog: Fostering reflective gameplaying for learning. *Proceedings of the 2007 ACM SIGGRAPH Symposium on Video Games,* 5–9 August 2007, San Diego, CA, USA.

JOURNAL OF LEARNING ANALYTICS
(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

Analyzing Engineering Design through the Lens of Computation

Marcelo Worsley and Paulo Blikstein
Stanford University

Abstract: Learning analytics and educational data mining are introducing a number of new techniques and frameworks for studying learning. The scalability and complexity of these novel techniques has afforded new ways for enacting education research and has helped scholars gain new insights into human cognition and learning. Nonetheless, there remain some domains for which pure computational analysis is currently infeasible. One such area, which is particularly important today, is open-ended, hands-on, engineering design tasks. These open-ended tasks are becoming increasingly prevalent in both K–12 and post-secondary learning institutions, as educators are adopting this approach in order to teach students real-world science and engineering skills (e.g., the “Maker Movement”). This paper highlights findings from a combined human–computer analysis of students as they complete a short engineering design task. The study uncovers novel insights and serves to advance the field’s understanding of engineering design patterns. More specifically, this paper uses machine learning on hand-coded video data to identify general patterns in engineering design and develop a fine-grained representation of how experience relates to engineering practices. Finally, the paper concludes with ideas on how the specific findings from this study can be used to improve engineering education and the nascent field of “making” and digital fabrication in education. We also discuss how human–computer collaborative analyses can grow the learning analytics community and make learning analytics more central to education research.

Keywords: Engineering design, design thinking, machine learning analytics, expertise

1 INTRODUCTION

Over the past three decades, technology has had a significant impact on education (see Koedinger & Corbett, 2006; Lawler & Yazdani, 1987; Papert, 1980; Resnick, 2002; U.S. Department of Education, 2010; Wilensky & Riesman, 2006 for examples). From the observed transition from chalk and blackboard to whiteboards to overhead projectors to PowerPoint presentations to online videos to cognitive tutors to virtual learning communities. Through these developments, it is apparent that instructional approaches have gradually incorporated new technologies. But innovations were not only in the delivery of information: more recently, technology has clearly altered elements of teaching and learning. Technological innovations have also allowed us to capture and process much more extensive traces of how people learn in digitally monitored settings. Access to this expanse of data has been central to the development and proliferation of both the learning analytics and educational data mining communities (Baker & Yacef, 2009; Siemens & Baker, 2012; Bienkowski, Feng, & Means, 2012). Furthermore, the use of these technologies has enabled researchers to tackle and study educational challenges at scale and in novel ways. Despite all of the affordances, a number of challenges remain outside of the current capabilities of traditional learning analytics and educational data-mining approaches. As we consider learning analytics as a middle space, we would like to propose that computer-based analysis, by itself, is insufficient for answering many important research questions in education. Domains with a wide variety of possible solutions and learning pathways represent a challenge for purely automated analyses. For example, fields where students are challenged to invent or create hardware or software solutions typically necessitate a level of human interpretation that can be difficult for a computer to infer. Similarly, situations where the design constraints may involve carefully weighing social and cultural concerns in conjunction with traditional engineering requirements may also require intensive and subtle human interpretation. While technological advances will undoubtedly expand the capabilities of pure computational analysis to a larger array of learning activities, we argue that we can address some of these challenges by combining analytics techniques with human coded data analyzed through qualitative approaches. This methodological intersection creates hybrid systems in which computer analysis is employed to study human labelled data. While we have been using these types of approaches in the nascent field of multimodal learning analytics (e.g. Blikstein, 2013; Worsley, 2012; Worsley & Blikstein, 2013), to date, we know of very few instances of Learning Analytics research that takes human-labelled data and exhibits how computational analysis can mirror, and extend, approaches and results achieved through traditional education research. However, a reality is that this type of qualitative research is what many education researchers are pursuing. By demonstrating the existence of robust computational methods that can be used to streamline traditional education research analyses, the field of Learning Analytics can more squarely enter the fold of the learning sciences. Such collaboration will serve to improve the quality and scalability of current education research, and increase the impact of Learning Analytics.

To help advance these goals and further the fields understanding of engineering design practices, we present two examples from an engineering task that demonstrate how combining elements of traditional qualitative analysis with machine learning can 1) help us identify patterns in engineering strategies and 2) allow us to garner a more fine-grained representation of how engineering practice varies by experience level.

2 LITERATURE REVIEW

This study is informed by prior research from engineering education and the learning sciences, and a largely distinct body of literature on artificial intelligence techniques that can potentially be used for studying open-ended learning environments. Our research bridges these communities by showing that each domain has a strong contribution to make in advancing the field’s understanding of learning, especially in constructionist learning environments (Papert, 1980; Harel & Papert, 1991). In what follows, we highlight key studies from these paradigms and describe how their work informs the current study.

2.1 Engineering Education Research

The area of Engineering Education has received a great deal of attention recently. There have been various efforts to bring project-based learning to the forefront of engineering education and an equally strong call for curriculum to emphasize process instead of *product*. As an example of these changes, professors and researchers have been redesigning both first year and capstone design projects with the hope of helping students develop greater fluency with the tools and methods that they will need as practicing engineers confront. Traditionally, work in engineering education and project-based learning has involved developing new approaches for assessing learning and knowledge. Typically, studies from this body of research focus on qualitative analyses of student language (Atman & Bursic, 1998; Dym, 1999; Russ, Scherr, Hammer, & Mikeska, 2008), student artifacts (Dong & Agogino, 1997; Lau, Oehlberg, & Agogino, 2009), or the combination of language and artifacts (Atman & Bursic, 1998,; Worsley &

Blikstein, 2011) created in the process of designing, building and/or inventing. We look to contribute to the body of engineering education research by analyzing these practices at a very fine-grained scale.

2.2 Research on Expertise

Within the engineering education community, and beyond, considerable research has been undertaken in the study of expertise (for examples see Chi, Glaser, & Rees, 1981; Cross & Cross, 1998; Ericsson, Krampe, & Tesch-Römer, 1993). More specifically, a collection of researchers has investigated design patterns on engineering tasks through think-alouds (Atman & Bursic, 1998; Ericsson & Simon, 1980; Russ et al., 2008). When considering expertise in the engineering context, many of the constructs discussed have been cast under different names: computational thinking (Resnick et al., 1998; Wing, 2006; Guzdial, 2008), designing thinking (Dym, 1999; Dym, Agogino, Eris, Frey, & Leifer, 2005), and mechanistic reasoning (Russ et al., 2008). Because each of these constructs could easily be the subject of an entire review, we will only mention them in passing as to indicate that these ideas have contributed to the analyses in this paper we will focus on a single body of literature by Atman and her collaborators (Adams, Turns and Atman, 2003; Atman & Bursic, 1998; Atman, Chimka, Bursic, and Nachtmann, 1999; Atman et al., 2007; Atman, Kilgore, & McKenna, 2008) and which is representative of the state of the field, and is directly related to our analyses. Atman and Bursic (1998), Atman et al. (1999), and Adams, Turns, and Atman (2003) investigate engineering design language and practices by comparing engineering practices between freshmen and senior engineering students. In Atman et al. (2007), they compare expert engineers to college engineering students. The comparisons examined how the respective groups were rated in terms of time spent along several dimensions: Problem Scoping, Information Gathering, Project Realization, Total Design Time, and Considering Alternatives in the following activities: problem scoping, information gathering, project realization, total design time, considering alternatives, and solution quality. In conducting this comparison, the authors expected to find that experts (a) do a better job at gathering information (b), spend more time in the decision making process (c), spend more time in the project realization process (d), consider fewer design alternatives, and (e) spend more time transitioning between the different types of design activities. They employed basic quantitative measures to keep track of the number of times a given action was taken and the amount of time devoted to each action, and found that while a handful of their hypotheses were correct, the most insightful finding had little to do with the quantitative differences between the groups. Instead, the true findings had more to do with the overall pattern that different experts followed. While the group had previously identified iteration as an important component to engineering design, Atman et al. (2007) describe the expert design process as being like a cascade. These cascades were seldom present among novices. To identify cascades, Atman, Diebel, and Borgford-Parnell (2009) focused on three different representations of the students' design activities and stages. These representations include a timeline plot, which shows the presence or absence of a given action at each increment in time; a cumulative time plot, which captures the amount of time spent in each activity (y-axis) relative to the total amount of time (x-axis); and progress time plots, which is the same as the cumulative time plot, except that the x-axis is the percentage with respect to each individual action, as opposed to the overall time for all activities. Using the progress time plots, Atman, Diebel, and Borgford-Parnell (2009) define a cascade as being a design process typified by considerable time doing project scoping at the onset, and project realization at the end. Embedded within the ways that Atman et al. identified iterations and cascades is the importance of temporality. Simply looking at the number of times, or amount of time individuals spent in a given action was not predictive. Instead, the authors needed to look at the entire sequence of actions taken and the context in which each action appeared.

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

In the same way, we are interested in studying the overall patterns of the engineering design process and doing so in a relatively automated fashion. However, traditional approaches for conducting this type of research are limited to human-based video analysis, which can be quite laborious and time-consuming. Other strategies, such as sequence mining techniques, tend to remove the individual or groups of segments from the context in which they appear. Nonetheless, we have identified a set of approaches from machine learning that inform the computational aspects of this study.

2.3 Machine Learning Analysis of Computer Programming Behaviour

Central to this paper is a desire to study human actions on relatively open-ended tasks. When considering automated analysis of open-ended tasks, much of the previous work relates to studying computer-programming behaviour. Blikstein (2011), Piech, Sahami, Koller, Cooper, and Blikstein (2012) Blikstein, Worsley, Piech, Sahami, Cooper, and Koller (in press) are examples of this work. The three papers describe a similar strategy of gathering snapshots of students' computer programs. Blikstein used the snapshots to examine the differences between expert and novice programmers. There he identified prototypical styles and patterns that students used over the course of a multi-week assignment. Piech (2012) and Blikstein et al. (in press) used the snapshots as the basis for identifying a set of generalizable states that students enter while completing a given assignment, or set of assignments. These states were determined through clustering and used to construct models of student learning pathways. In Piech et al., the authors build a Hidden Markov Model (HMM) of the different student paths. The transition probabilities from the HMM were used to compare individual students and ultimately cluster them into three groups. The clusters identified in their study aligned with final examination performance at a higher level of accuracy than could be achieved by using the midterm examination grade as a predictor. In Blikstein et al. (in press), the authors examine learning pathways across an entire computer science course, and show how progressions in students' tinkering and planning behaviour correlates with student grades.

From these three studies, it becomes apparent that the tools of computational analysis hold significant promise, especially when faced with large datasets. In the case of Piech et al. specifically, we find that using machine learning and probabilistic graphical models can be invaluable in developing representations of the students' data from which we can learn. In our analysis, we follow a similar approach but in the domain of hands-on engineering design tasks. Additionally, where Piech et al. computes student similarity based on the HMM transition probabilities, we chose not to make the Markov assumption which is only concerned with the immediately preceding state. This allows us to maintain the context for each student's action.

Other work by Berland et al, Martin, Benton, Ko, and Petrick-Smith (2013) uses clustering to study prototypical program states among novice computer programming students. They used these prototypical programming actions as the basis for studying how students transition between different actions. In so doing, they found that the data could be used to identify three general patterns: tinkering, exploring, and refining. These categories extend previous work on tinkering and planning behaviours in computer programming with a more complex representation of student programming practices (Turkle & Papert, 1991). Our Analysis 1 from this paper follows a similar paradigm by identifying general building patterns among our population of participants.

Taken together, a primary affordance of computational analysis is the high level of resolution one can achieve. While our analysis does not employ a big data science in the traditional sense of having thousands many participants, we do look at participant actions at a level of granularity that would be

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

hard to replicate by purely human analysis. Because of this, we are able to identify otherwise undetectable patterns in behaviour.

The remainder of this paper is divided into five sections. In the next section, we describe the dataset and the coding scheme. This is followed by an introduction to the basic machine learning techniques used in the two analyses that we present. We then transition into studying the first set of questions: 1) What are the prototypical building strategies used among engineers of varying levels of experience? 2) In what ways do these building practices relate to prior literature? 3) What new insights can we garner about engineering through these prototypical building practices? After addressing these questions, we move to the second set of questions that are specifically related to correlations between student actions and prior experience: 1) What building actions distinguish individuals of different levels of experience? 2) How do these building actions align with, contribute to, or differ from prior research in this field?

3 METHODOLOGY

3.1 Data

Data is drawn from thirteen participants. Each participant was given everyday materials, and asked to build a tower that could hold a small mass (< 1 kg). Participants were also challenged to make the mass sit as high off the ground as possible. The task was designed to examine how successfully students are able to take their intuitions about mechanics and physics and translate them into a stable, well-engineered structure. We expected students to use knowledge about forces, symmetry, and the affordances of different geometric shapes, to enable them to complete the task. The additional challenge of making the structure as tall as possible was introduced to push all students to the limits of their ability, regardless of prior experience.

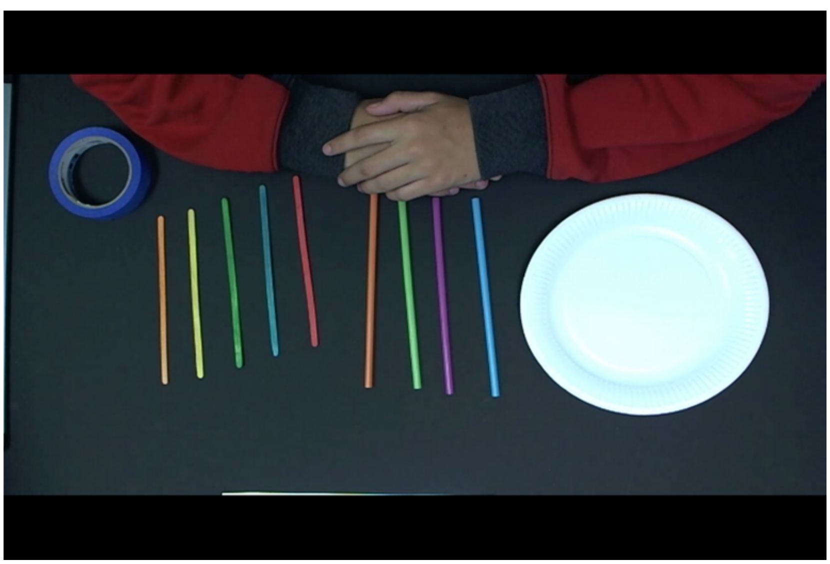

Figure 1. Initial Materials

Students were given four drinking straws, five wooden Popsicle sticks, a roll of tape and a paper plate (Figure 1) and were told that they would receive ten minutes to complete the activity. In actuality, they were permitted to work for as long as they wanted. Average participation time was approximately 25

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

minutes (SD=13minutes.) Three sample structures are depicted in Figures 2, 3, and 4 to give the reader a better idea of the task.


Figure 2. Sample Structure 1

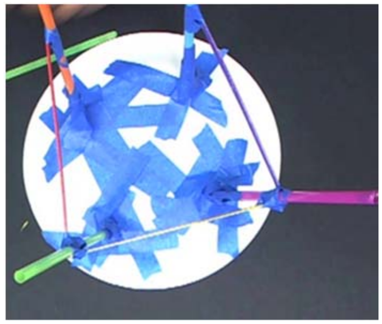
Figure 3. Sample Structure 2

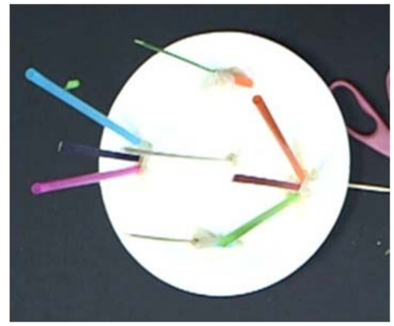
Figure 4. Sample Structure 3

Audio was used to capture meaningful utterances made by the participants, though students were not required to engage in think-alouds. Audio was also captured of each student's metacognitive analysis of his or her building approach. A video camera place above the students, pointing vertically down to the work area, captured the movement of objects as students progressed through the task (Figure 1). Gesture data, which consisted of twelve upper-body parts from a Kinect sensor, recorded the students' physical actions. While we only focus on the video data for this paper, Worsley and Blikstein (2013) contains a preliminary analysis of how the gesture data may provide an automatic channel for predicting expertise based on the frequency of two-handed actions.

3.2 Defining Experience

Prior to the study, students were classified based on their level of experience in the domain of engineering design based on two dimensions. The first dimension pertains to the amount of formal instruction students had received in engineering. Individuals who had completed undergraduate or graduate degrees in engineering were labelled as relative experts. Individuals who had not completed degree programs in engineering answered interview questions about their prior experience. These interviews, in conjunction with teacher-based ratings, were used to label the relative level of experience of each participant. To provide some additional context, the teachers worked with the students for more than two-hundred hours in an engineering and digital fabrication class, over four weeks. Student experience labels were assigned only when all researchers agreed. This labelling process resulted in a population of three experts, two high experience students, five medium experience students, and three low experience students.

3.3 Coding

In order to establish a basis for comparing students, we developed a set of action codes (Table 1). The process we followed in developing these codes mirrors that commonly undertaken in grounded theory-based research. An initial set of codes was identified through open coding of a sample of the videos. After individually developing a set of codes, the research team came together to discuss those codes and agree upon which ones to include in the final codebook. Once those codes had been defined and agreed upon, a graduate research assistant coded each video.

Table 1. Fine-Grain Object Manipulation Codes

Code	Description
BUILDING	Joining objects by tape or other relatively permanent means.
PROTOTYPING MECHANISM	Seeing if putting two (or more) objects together will work. This may include acting out a mechanism with the materials.
TESTING MECHANISM	Testing a subsection of the overall system.
UNDOING	Taking the structure apart to make a change to a previous build.
SINGLE OBJECT EXAMINATION	Pressing on or bending an object to explore its properties.
THINKING WITHOUT AN OBJECT IN HAND	Surveying the pieces but not touching anything or actively doing anything.
THINKING WITH AN OBJECT IN HAND	Holding one or more objects but not manipulating them.
SYSTEM TESTING	Putting force on a collection of relatively permanently affixed pieces to see if they will hold the mass.
ORGANIZING	Repositioning the raw materials but not actually building, examining, or prototyping.
BREAKING	Breaking apart sticks, bending straws, or ripping the plate.
ADJUSTING	Repositioning an object slightly, or applying more tape to reinforce or correct portion of the structure.

Similar to Atman's "Design Stages," we developed a scheme of higher-level object manipulation classes. These include *Realization, Planning, Evaluation, Modification*, and *Reverting.* The mapping between Object Manipulation Classes and Object Manipulation Codes can be found in Table 2. For the analyses presented in this paper, we will focus on examining patterns at the Object Manipulation Class level.

Table 2. General Object Manipulation Action Classes

Class	Codes
REALIZE	• Building and Breaking
PLAN	• Prototyping mechanism • Thinking with or without an object • Single object examination • Organizing and Selecting materials
EVALUATE	• Testing a mechanism • System testing
MODIFY	• Adjusting
REVERT	• Undoing

3.4 General Algorithm

3.4.1 Sequence Segmentation

The analytic technique begins by segmenting the sequence of action codes every time an EVALUATE action occurs. Our assumption is that we need to have a logical way for grouping sequences of user actions and each time a user completes an EVALUATE action, they are signalling that they expect their

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

previous set of actions to produce important, actionable information and feedback, which may be in the form of their current structure succeeding or failing.

3.4.2 Segment Characterization

Each segment is recorded based on the proportion of the five object manipulation classes (REALIZE, PLAN, EVALUATE, MODIFY, REVERT) that took place during that segment. Put differently, we now have a five dimensional feature vector for each segment, where each dimension corresponds to one of the object manipulation action classes. As an example, consider the following set of codes:

PLAN, REALIZE, EVALUATE, MODIFY, REVERT, REALIZE, EVALUATE

This sequence of eight codes would be partitioned into two segments. The first segment would be PLAN, PLAN, REALIZE, EVALUATE; the second would be MODIFY, REVERT, REALIZE, EVALUATE. These two segments would then be used to construct two feature vectors based on the proportion of each of the action classes. In the case of the first segment of the example sequence — PLAN, PLAN, REALIZE, EVALUATE — we see that there are two PLANs, one REALIZE, one EVALUATE, zero MODIFY, and zero REVERT. Thus, the proportion of the segment occupied by PLAN is one-half, or 0.50. The proportion of the segment occupied by REALIZE is one-fourth, or 0.25 and the proportion of the segment occupied by EVALUATE is also one-fourth. Following this same procedure for both of the segments yields the results in Table 3.

Table 3. Sample Segmented Feature Set

Segment	MODIFY	REALIZE	PLAN	EVALUATE	REVERT
1	0.00	0.33	0.33	0.33	0.00
2	0.25	0.00	0.00	0.25	0.25

3.4.3 Segment Standardization

After constructing all segments for all participants, each column (MODIFY, REALIZE, PLAN, EVALUATE, REVERT) of the feature set is standardized to have unit variance and zero mean. This step is taken in order to ensure no biases when we perform clustering in the next step.

Table 4. Sample Segmented Feature Set after Standardization

Segment	MODIFY	REALIZE	PLAN	EVALUATE	REVERT
1	-1	1	1	1	-1
2	1	-1	-1	-1	1

3.4.4 Segment Clustering

Following standardization, the segments are clustered into four or ten clusters using the k-means algorithm (the selection of four and ten as the number of clusters will be discussed in more detail later.) The clustering process uses all of the students' action segments in order to develop a set of generalizable action segments.

Each of the resultant clusters contains several of the segments, and can be characterized by the cluster centroid. This cluster centroid represents the cluster's average value along the five dimensions. As an example, if segment 1 and segment 2 defined a cluster, their cluster centroid would be zero along all dimensions (Table 5).

Table 5. Hypothetical Cluster Centroid from Sample Feature Set

Segment	MODIFY	REALIZE	PLAN	EVALUATE	REVERT
1	-1	1	1	1	-1
2	1	-1	-1	-1	1
Centroid	0	0	0	0	0

3.4.5 Segment Re-labelling

Each segment for each student is replaced with the generalizable action segment that it is most similar to (recall that an action segment is characterized by a five-dimensional vector that reports the proportion of that segment spent in each of the five Object Manipulation Classes: REALIZE, PLAN, EVALUATE, MODIFY, REVERT). Following from the example, above, the two segments would be re-labelled using the cluster centroid values and label (Table 6).

Table 6. Hypothetical Segment Re-labelling

Segment	Cluster
1	0
2	0

3.4.6 Dynamic Time Warping

Once each student's sequence has been re-labelled using the cluster centroids from above, "dynamic time warping" is used to compute the minimum distance between each pair of participants. Dynamic time warping minimum distance can be seen as a constrained form of the minimum edit distance or Levenshtein distance. It differs from edit distance in that when computing the distance between two sequences, a given sequence can only undergo item insertion. The inserted item must either be a repetition of the preceding item or the subsequent item. Item insertion is only used if it will reduce the overall distance between two students' sequences; otherwise, a simple numerical value is computed based on the Euclidean distance between the two vectors. As a very simple example, if we were computing the distance between two sequences: A) 1, 2, 0 and B) 1, 2, 2, 2, 1; we would extend sequence A to be 1,2, 2, 2, 0, such that the second value is repeated in order to produce the maximum alignment between sequences A and B. The reason for using dynamic time warping is that we are interested in looking at the overall design patterns that participants are using and are less interested in the amount of time spent in the respective stages. Dynamic time stretches the different students' vectors based on minimizing the differences between them and in no way alters the order in which actions appear. This computation yields an n-by-n matrix of minimum distances.

3.4.7 Participant Clustering

Finally, the n-by-n matrix from the dynamic time-warping calculation is standardized along each column, before being used to construct the final clustering, again with the k-means algorithm.

3.4.8 Algorithm Summary

In summary, this algorithm takes the full sequence of actions from each student and splits them in smaller segments every time a student explicitly evaluates, or elicits feedback from, his or her structure. The proportions of actions in the different segments are used to find representative clusters, which are subsequently used to re-label each user's sequence of segments. Finally, we compare sequences across participants and perform clustering on the pair-wise distances in order to find natural groupings of the participants. Figure 5 provides a visual representation of the overall process. In the following sections,

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

we show how this general algorithm can be used to 1) study prototypical building patterns as well as 2) identify those building patterns that differentiate individuals of differing levels of experience.

Figure 5. Summary of General Algorithm

4 ANALYSIS 1: PROTOTYPICAL BUILDING STRATEGIES

The goal of Analysis 1 is to identify prototypical building strategies among the research participants. More specifically, we answer the question of how building patterns can be used to understand engineering practices better. To address our question, we proceed by first discussing the types of patterns that we expected to see among our participants. We also describe the specific instantiation of the general algorithm that we applied for this portion of the analysis and why this approach is feasible. We then present three different representations of our findings: one quantitative, one based on video analysis, and one qualitative. Finally, we conclude the analysis with a discussion of how these findings can be used for studying the learning of basic engineering skills and used more broadly in the learning analytics community.

4.1 Hypotheses

Based on prior literature, one hypothesis is that students will use cascades (Atman et al., 2007; Atman Deibel, & Borgford-Parnell, 2009). During such cascades, students pay particular attention to PLAN at the beginning of the task and gradually decrease the proportion of time spent in PLAN as a greater proportion of time is spent in REALIZE. They are also constantly in the process of considering alternative designs. Another design process pattern to look for is iterative design. Atman and Bursic (1998) found that iterative design was important for creating effective solutions. As the individual begins to engage in the realization process, he or she is constantly updating the design, and perhaps even returning to PLAN actions in order to refine the product iteratively. In our building context, we expect to see cascades manifested as different amounts of iterative design. While Atman et al. typically attribute this to be being an expert-like behaviour, they do indicate that it is not limited to experts. Instead, they found that it merely occurred more frequently among their more experienced research participants.

Connected with the above hypothesis about design process is one of quality. Prior research found that as the amount of iterative design or cascading increased, so did the quality of the artifacts produced (Atman et al., 2007). Accordingly, an additional hypothesis is that the prototypical building strategies will have some correlation with the quality of the products.

4.2 Algorithm Implementation Specifics

We use the methodology described in the General Algorithm section and cluster the data into four clusters during the Segment Clustering step, as well as during the Participant Clustering step. The number of clusters was set to four during segment clustering based on the silhouette score (Rousseeuw, 1987). In the case of participant clustering, four clusters were used in order to ensure some variation between clusters, while also avoiding clusters with only one participant.

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

4.3 Object Manipulation Generalizable Segments

Recall that the approach we use involves k-means clustering at two different times. The first instantiation of clustering is intended to identify a set of generalizable action clusters. Each of these action clusters is defined by the percentage of time spent in Planning, Realizing, Modifying, Evaluating, and Reverting.

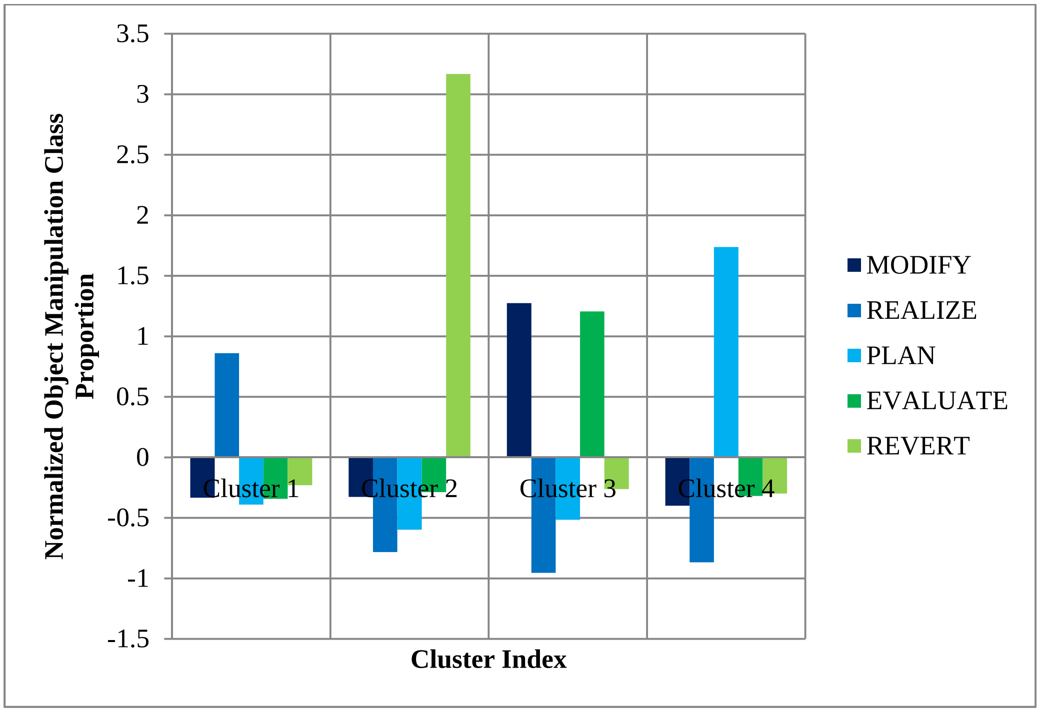

Figure 6. Object Manipulation Class Proportions by Cluster

In Figure 6, we report the Object Manipulation Class proportions for the different cluster centroids. From this, it is clear that Cluster 1 primarily aligns with REALIZE, Cluster 2 with REVERT, Cluster 3 with MODIFY and EVALUATE, and Cluster 4 with PLAN. In order to simplify discussion of these in the following sections, we will refer to the generalizable clusters as G-REALIZE, G-REVERT, G-MODIFY-EVALUATE, and G-PLAN. In this way, the user will not be confused between our discussion of the Object Manipulation Classes and the Generalizable Segment Labels.

4.4 Participant Cluster Centroids

The second stage of clustering occurs among the participants and is based on the similarity of their dynamically time warped Object Manipulation Sequences. From that clustering, four participants were assigned to Cluster A, three to Cluster B, two to Cluster C, and four to Cluster D. To simplify the naming, we will always refer to clusters of object codes (from Segment Clustering) using numbers, and clusters of participants (from Participant Clustering) using letters.

In order to better understand the nature of these different clusters and explore how their characteristics relate to prior research, we present three representations of the clusters. However, as a first indication that the clusters are differently, we treat each participant action segment as independent of all others. This obviously is not true, but provides a means for a quick comparison via Chi-Squared analysis. The

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

Chi-Squared analyses suggest that each cluster used the Generalizable Action Segments with markedly different frequencies (Table 7).

Table 7. Pair-wise Chi-Square Analysis of Generalizable Action Segment Usage

Group 1	Group 2	Chi-Square Statistic	Probability
A	B	57.6	0
A	C	42.48	0
A	D	33.35	0
B	C	69.03	0
B	D	44.7	0
C	D	64.78	0

4.5 Time-Based Graphical Representation of Participant Clusters

Having established that the four clusters are different, we now examine the nature of those differences. The first representation that we employ is a comparison of the time spent in the different Object Manipulation Classes.

Figure 7. Relative Time Spent in Each Object Manipulation Class by Cluster

As we compare the clusters in Figure 7, we clearly observe that Clusters A is characterized by having the most PLAN, and the least amount of EVALUATE. Cluster B has the least amount of REALIZE, and the most amount of REVERT. Cluster C (green) stands out for having the most time planning. On the other extreme is cluster A, which also spent considerable time in REALIZE and the most EVALUATE. It also has the least amount of PLAN, REVERT, and MODIFY as compared to its peers. Finally, Cluster D falls in the middle along all five of the dimensions.

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

4.6 Proportion-Based Graphical Representation of Participant Clusters

One drawback of the normalized time plot (Figure 7) is that it does not take into account the total amount of time participants took on the task. Accordingly, we present a graph of the proportion of time spent in each of the Generalizable Action Segments by cluster (Figure 8).

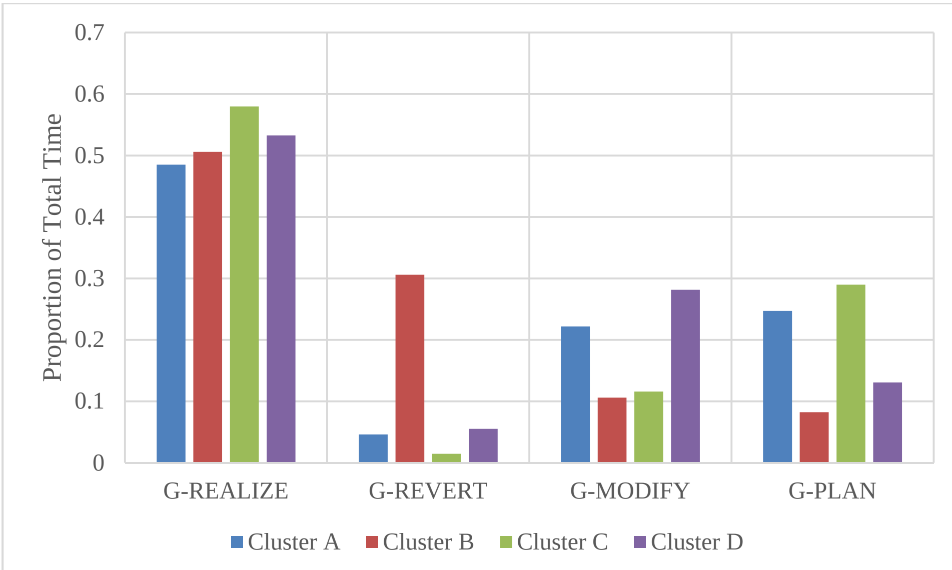

Figure 8. Proportion of Time Spent in Each of the Generalizable Action Segments by Cluster

When we move to this representation, we find that Cluster C, who spent the least amount of time planning in the previous graph, spends the largest proportion of time in G-PLAN. Hence, it is not that Cluster C participants did not plan; it is simply that their total time planning was less than their peers. Proportions for G-REVERT and G-MODIFY are at the lower end of the spectrum for the different clusters. Looking at the proportions also informs our understanding of Cluster A, which spends a large proportion of time in G-MODIFY despite spending little (absolute) time in MODIFY and EVALUATE (Figure 6). Apart from these, this representation appears to be analogous to what we observed in Figure 7. Thus using the proportion of time helps to better describe some of the nuances of each group's behaviour, while confirming many of the observations from the absolute time spent doing each object manipulation type. Furthermore, these two representations describe the noted differences within the Chi-Squared analyses.

4.7 State–Transition Representation of Participant Clusters

The first representation focuses on aggregate time spent in the different Object Manipulation Classes and Generalizable Action Segment types. These have been used within the literature as ways of studying engineering design patterns (e.g. Adams, Turns & Atman 2003; Atman et al., 1999, 2007, 2008). However, one goal of this paper is to go beyond this and look more closely at the patterns of actions that students take. The literature has suggested that examining the rate of transition among different actions can be informative for studying design patterns (Atman et al., 1999, 2007). To consider these, we construct a state machine of student actions within each cluster. Moreover, we can construct a transition probability table that examines the frequency with which individuals of a given cluster transitioned between different Generalizable Action Segments. Putting the data in this form deviates

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

from the dynamic time-warp analysis that we completed on the entire sequence of user actions, but still offers some insights into what characterizes each of the clusters.

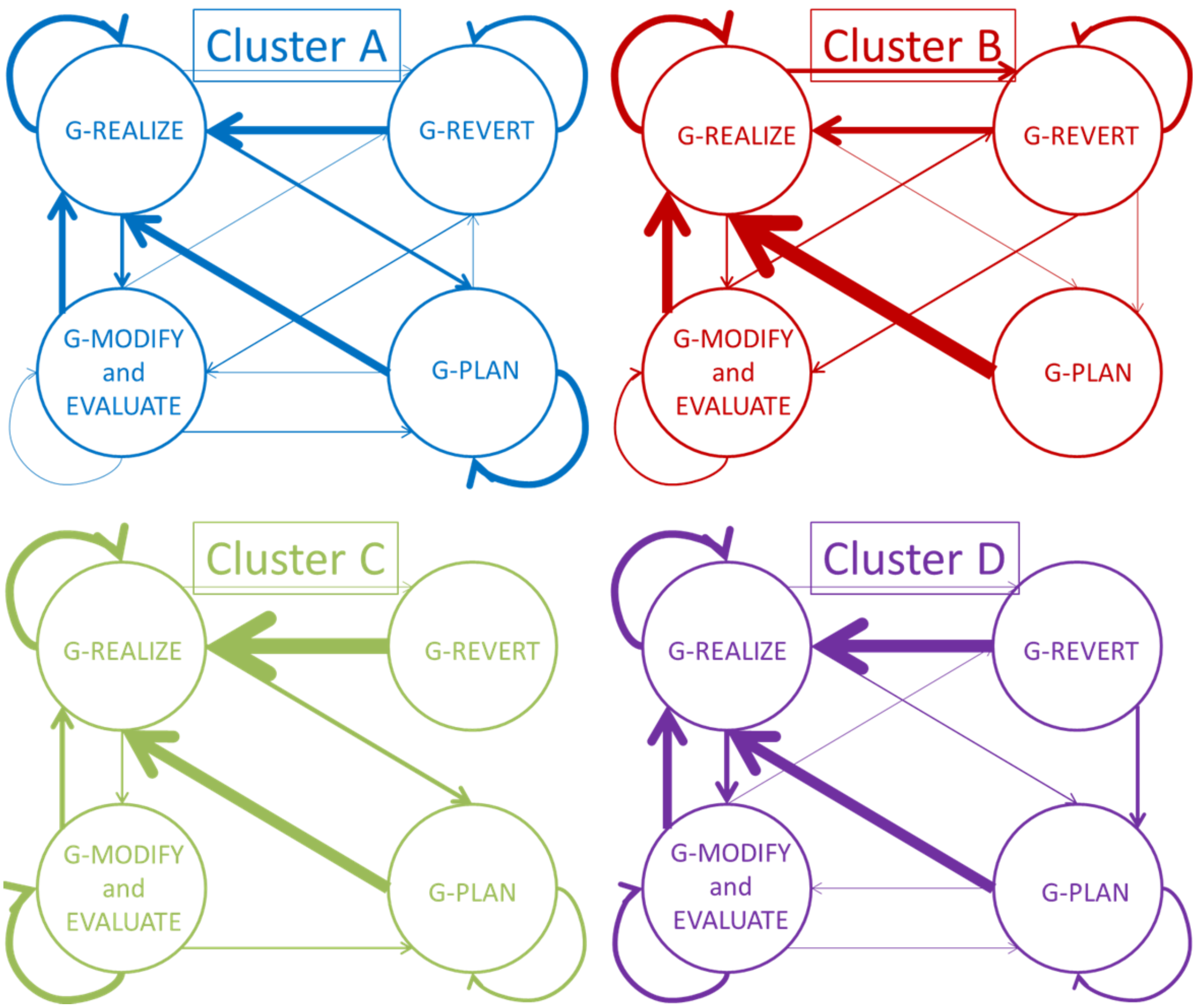

Figure 9. Transition Diagrams by Different Clusters. The size of each line corresponds to the probability of that transition with thicker lines indicating higher probability than thinner lines

Figure 9 shows the state machine diagram for all four clusters. Before diving into the specifics of each group's transition patterns, we present pair-wise Chi-Square analyses of the transition probabilities across all pairs of states. From

Table 8 we again see that all of the groups significantly differ from one another in their transition behaviour.

Table 8. Pair-wise Chi-Square Analysis of Transition Probabilities

Group 1	Group 2	Chi-Square Statistic	Probability
A	B	100.08	0
A	C	127.66	0
A	D	89.12	0
B	C	53.60	0
B	D	68.76	0
D	C	72.20	0

4.7.1 Cluster A

Cluster A is typified by planning behaviour, which appears to be sustained and frequent. Cluster A also records a relatively reduced transition probability for building. As a point of comparison Cluster A participants spend relatively less time transitioning to G-REALIZE than any other cluster. Approximately 50% of Cluster A's actions consist of transitions to G-REALIZE, whereas the value is roughly 65% for the other clusters. Instead of transitioning into G-REALIZE, Cluster A is frequently transitioning in and out of G-PLAN. Moreover, unlike many of the other clusters, Cluster A is more likely to engage in sustained planning, meaning that they will return to G-PLAN immediately after completing a G-PLAN segment.

4.7.2 Cluster B

Cluster B is typified by a lack of planning, and a prevalence of reverting. As evidence for this categorization, Cluster B seldom transition into G-PLAN. Furthermore, after completing a G-PLAN segment, the group always transition into G-REALIZE. Hence there are no instanced of sustained planning, as was the case for Cluster A. Apart from frequently transitioning to G-REALIZE and seldom transitioning to G-PLAN, Cluster B, differs from Clusters C and D in how they transition into G-REVERT. Namely, Cluster B is more likely to enter into a G-REVERT state than Clusters C and D.

4.7.3 Cluster C

From the transition probabilities, Cluster C appears to be largely focused on building. Of all of the clusters, Cluster C engaged in the most sustained G-REALIZE activity. The probability of staying in G-REALIZE was 0.55, whereas for the other clusters this valued ranged from 0.46 and 0.49. Additionally, Cluster C seldom transitioned into G-REVERT, and would always follow a G-REVERT segment with a G-REALIZE segment.

4.7.4 Cluster D

Cluster D is typified by being at the middle of the pack along all four measures. Cluster D is very focused on building, but also makes frequent use of G-REVERT.

4.8 Qualitative Representation and Discussion

Thus far, we have focused on using quantitative data to study each cluster's characteristics. In what follows, we synthesize data from the two previous representations, and combine it with some qualitative analysis in order to solidify and summarize the four prototypical groups that we identified. During this section, we use progress time plots. For all of these plots, purple corresponds to G-PLAN, blue corresponds to G-REALIZE, red corresponds to G-REVERT, and green corresponds to G-MODIFY.

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

4.8.1 Cluster A — PLAN, REALIZE and MODIFY

In our estimation, Cluster A represents a group of students that is exhibiting a robust design process and high quality ideas. Through the two graphical representations of time spent in each Object Manipulation Class and Generalizable Action Segment, we saw that this group exhibited a large amount of planning behaviour. Furthermore, as we turned to the state machine representation we observed the sustained planning behaviour that this group followed, in which they would repeatedly undertake G-PLAN actions. This is corroborated by qualitative observations made from the dataset. All of the individuals in this cluster built in a modular and iterative fashion. They started by planning, and then got a portion of their structure to achieve stability.

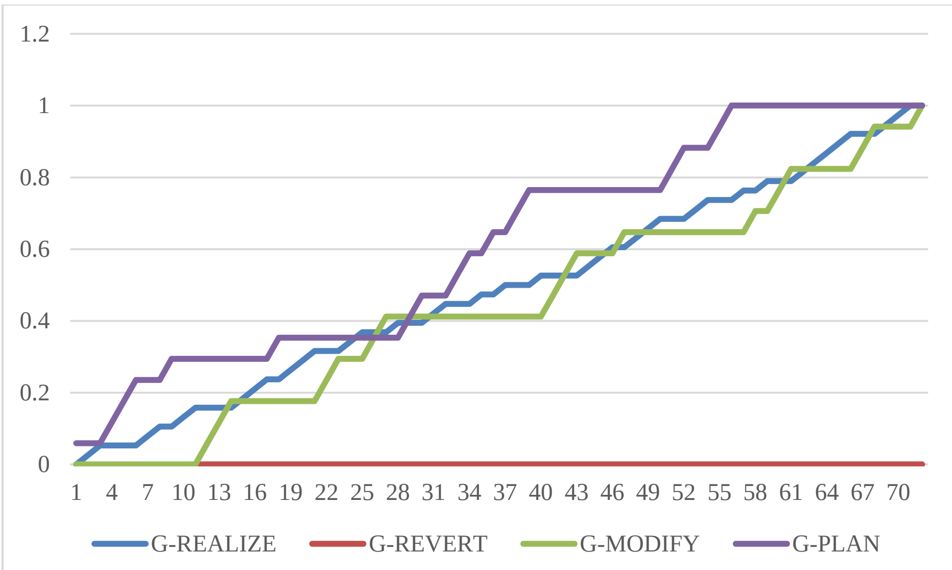

Figure 10. Cluster A Sample Progress Time Plot

After getting one part stable, they would return to planning and make another addition to their structure. This process would repeat itself until the participants were satisfied with their design or until all materials were used. Additionally, in their post-task metacognitive analysis, these students described their process as being iterative in nature, as well as involving some unexpected modifications. An example of this iterative approach can be seen in Figure 10, which depicts a progress time plot for one of the members of Cluster A. One can see from the plot that the blue and the purple lines are extensively intertwined. This is because the student alternated in using G-PLAN and G-REALIZE at different portions of the task. Finally, knowledge about engineering structures was evidenced in how the student talked about using triangles to reinforce the various supports in the structure.

4.8.2 Cluster B — REALIZE and REVERT

At the other end of the spectrum from Cluster A is Cluster B. From the aggregate time and state machine representations, we saw that Cluster B was characterized by G-REVERT actions and a lack of planning. In Figure 11, we see this represented in the purple and red lines with the purple line depicting G-PLAN. In this case, the line is flat, meaning that the student did all the planning at the beginning. The red line, indicating G-REVERT actions, steadily climbs throughout the process of the task and largely dominates all other activities. This was a common practice for this group. All of the individuals in this group had to undo their structures at one or more points during the task. Another key point of distinction that we observed qualitatively was that this cluster tended to use excessive amounts of tape in order to reinforce connections or as the actual support mechanisms in their structure, which means that they were less likely to use a variety of engineering strategies.

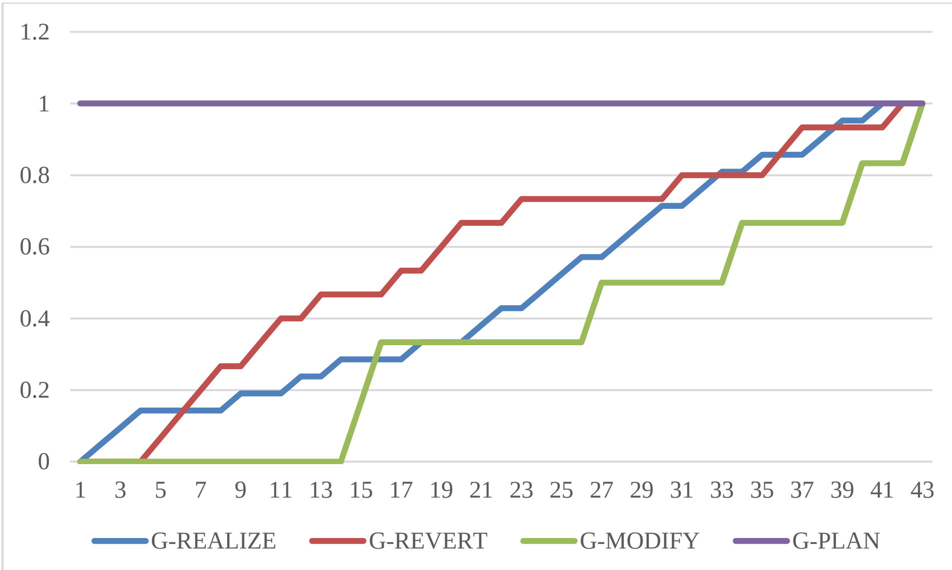

Figure 11. Cluster B Example Progress Time Plot.

4.8.3 Cluster C — REALIZE

In contrast to Cluster B, Cluster C consists of students who spent very little total time planning, relative to their peers, though they did spend a considerable proportion of their time planning. Interestingly, however, is that whereas Cluster A engaged in G-PLAN throughout the process, for Cluster C, planning was concentrated in just some moments. Comparing the rate of increase for planning instances in Figure 10 and Figure 12Figure 12, we see that the proportion of planning increases in much larger chunks for Cluster C, than for Cluster A.

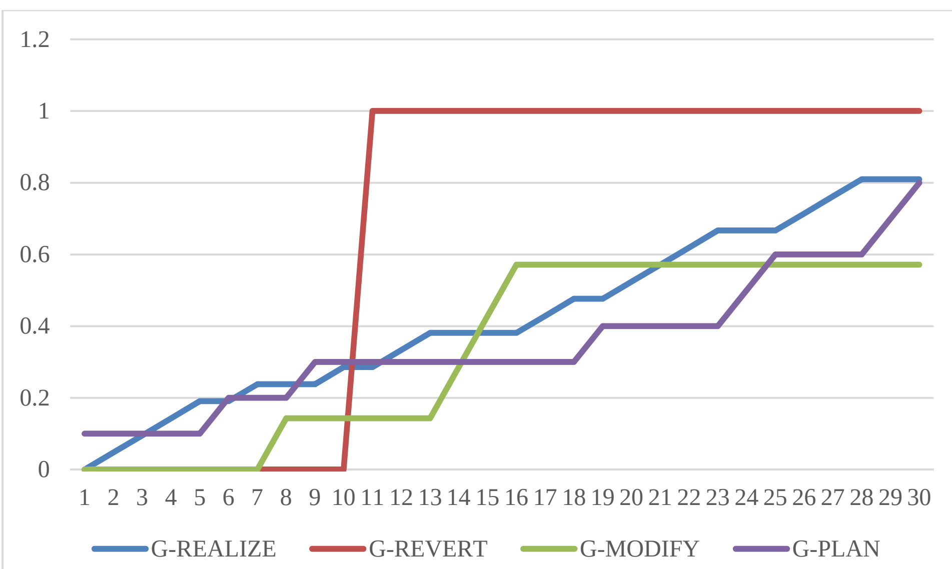

Figure 12. Cluster C Example Progress Time Plot. Purple is G-PLAN and blue is G-REALIZE

From the aggregate time plots and state machine representations, Cluster C appears to be typified by building process and a lack of undoing. This suggests that they generated ideas during the initial stages of the task that were sufficient to support the mass. As expected, in the qualitative analysis of these students, we saw a very streamlined process. Students would prototype a mechanism to make legs for the structure, test that prototype, and then repeat the process in order to make enough legs for the entire structure. One member of this cluster also found a way to use the roll of tape in the physical structure constructed (Figure 13).

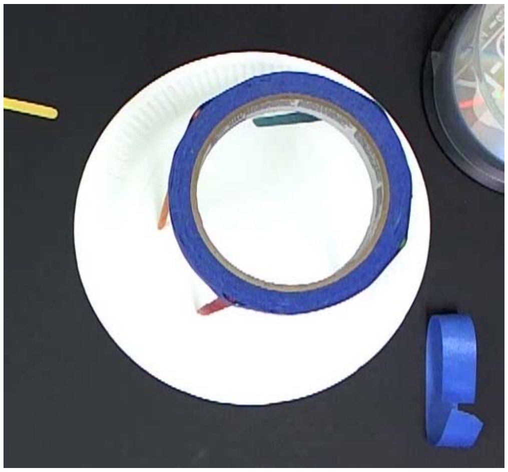

Figure 13. Clever use of roll of tape in design of structure

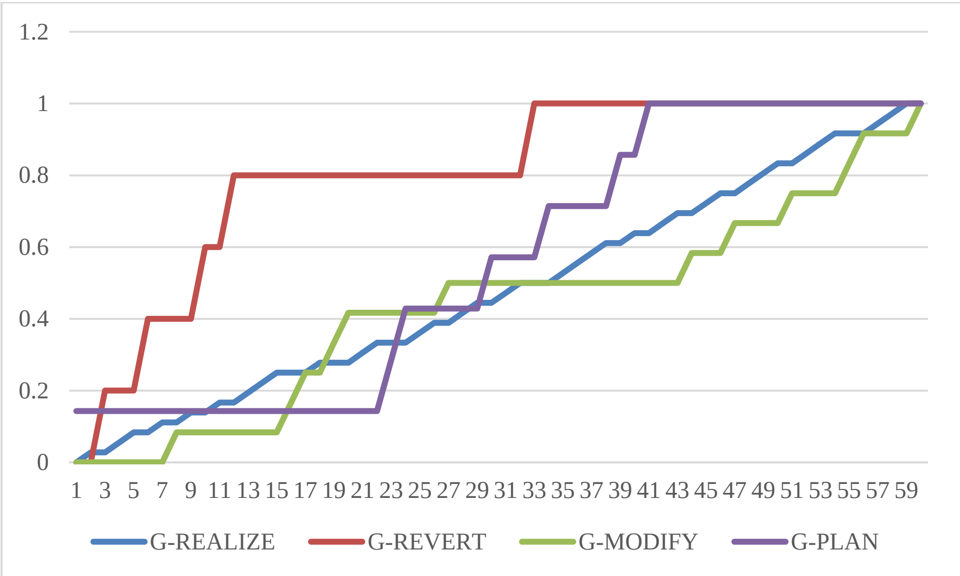

Figure 14. Cluster D Example Progress Time Plot. Red is G-REVERT, purple is G-PLAN and blue is G-REALIZE

4.8.4 Cluster D — PLAN, REALIZE, MODIFY and REVERT

Cluster D remained in the middle of the spectrum across all of the dimensions that we analyzed. Their distinctive characteristic is tied to starting in planning, and then subsequently iterating between G-REALIZE and G-PLAN, and later having that iterative process be disrupted by G-REVERT actions. For

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

example, they transition into G-REALIZE more frequently than Cluster A, but less so than the other clusters; and less frequently into G-REVERT than Cluster B, but more so than the other clusters. As can be seen in Figure 14 they begin by planning, and follow a process of iterating between G-REALIZE and G-PLAN. However, their pattern is also marked by frequent G-REVERT actions.

When we examine the behaviours of this cluster qualitatively, we confirm that it shares behavioural elements with each of the other clusters. For example, many of its members follow an iterative design process, in which they repeatedly prototype different aspects of their design and gradually test along the way. In this regard, they share the iterative building characteristic of Cluster A. However, they differ from Cluster A in the level of success of their ideas. Despite following a relatively sound design process, their structures lacked the appropriate engineering principles. For example, some of the students failed to reinforce the legs of their structures, causing them to fall over immediately.

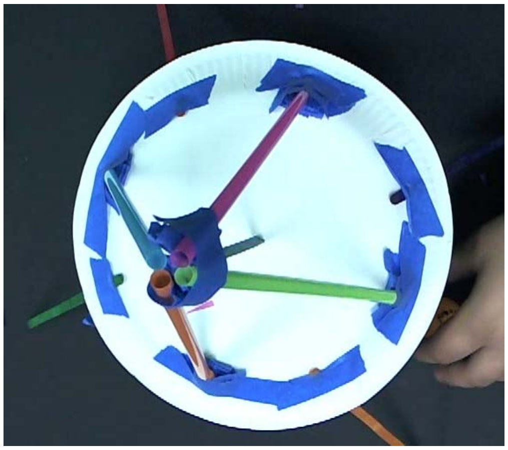

Figure 15. Cluster D Failed structure

Another student not only failed to reinforce the legs of the structure, but also assumed that the mass could balance on a circular surface without any reinforcement (Figure 15). Upon encountering such problems, students would often undo portions of the structure, not realizing the source of the structural failure. Thus, we would hypothesize that students who started with an iterative, systematic design had to resort to G-REVERT actions because of their lack of engineering knowledge. This difficulty in knowing how to debug their problems may have caused these students to share characteristics with the other clusters, despite being relatively systematic.

4.9 Dimensions of Analysis

As our initial hypothesis suggested, the approach that we used largely aligns with the quality of the design process and the quality of the engineering intuitions. Clusters A and D appear to be high on the quality of design process axis, as they follow an iterative design approach. Clusters B and C appear to be low on the scale of design process. Along the axis of engineering principles, clusters A and C appear to outpace clusters B and D. In this sense, the clustering has broken the participants into four quadrants of performance. Figure 16 shows the approximate placement of the clusters along these axes. One could posit that the clusters differ along a single dimension. However, from our qualitative analysis, quantitative analyses this does not seem to be the case. For example, based on the pair-wise Chi-Square analyses there is no way to reconcile the pair-wise Chi-Square statistics. However, the values can easily be reconciled when representing the clusters along two dimensions.

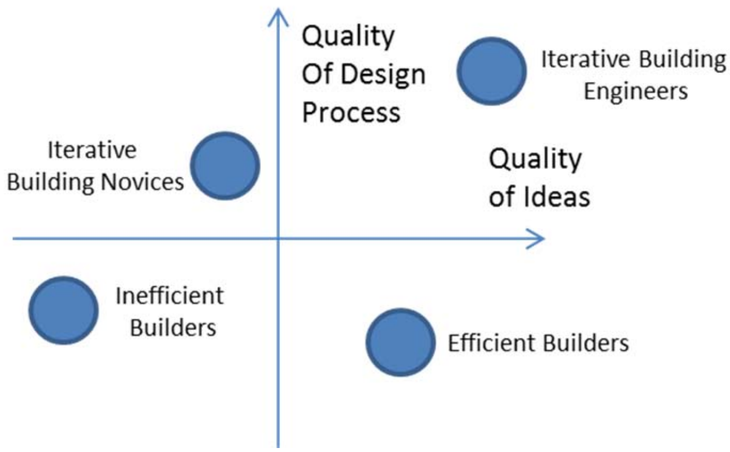

Figure 16. Quality of Design Process and Quality of Ideas Framework

4.10 Analysis Summary

From this analysis, we were able to generate prototypical building patterns that can be viewed as aligning to two dimensions. The two dimensions — design process and idea quality (Atman et al., 2007) — have been referred to in prior literature as being salient for engineering design tasks. However, unlike previous work, our analysis was completed using a computational analysis. Specifically, our result was achieved by using a dynamic time-warping algorithm, in conjunction with human coding and two iterations of clustering. The point of this analysis is not to suggest that these findings could not have been achieved through purely qualitative analysis, or that clustering will always produce these results. Qualitative analysis could have attained similar findings, but would have involved a much more labour-intensive effort. For example, it could have been quite challenging to 1) develop a systematic way for analyzing the data that would highlight these differences, 2) figure out an appropriate number of action segments to consider, and 3) determine a way to cluster that explicitly takes process into account (simply looking at the proportion of time each individual spent would overlook many of the nuances in the students' building patterns). These are all affordances that computational analysis can provide. At the same time, however, relying on purely computational approaches would not have been sufficient because of the challenges in coding the action segments. While our previous work has found some predictive power in using purely automated labelling of actions (Worsley & Blikstein 2013), here we are suggesting that combining qualitative analysis with data mining can enable education researchers to study complex learning environments more easily. Hence, as was the case with this study, they can provide novel frames for understanding the interaction between multiple dimensions.

In the second analysis, we move into answering more targeted questions about the nature of hands-on building practices. Instead of looking for an opportunity to explore the data and better understand the general patterns of behaviour, we enter with the goal of specifically identifying how to differentiate individuals of various levels of experience.

5 ANALYSIS 2: DISTINGUISHING BETWEEN DIFFERENT LEVELS OF EXPERIENCE

Under the previous analysis, we were primarily interested in identifying prototypical actions of students as they participated in an engineering design challenge. In this section, we are specifically interested in

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

what it means to have more experience. Moreover, since each student was classified based on his or her level of prior experience, we are interested in understanding how those differences in experience are manifested in the students' building practices. For this section, we again begin by considering what types of distinguishing practices we would expect to see among participants, and then discuss the specifics of the algorithm that we used. In discussing the algorithm, we also provide justification for using this approach over other alternatives. The discussion of the algorithm is followed by three representations of the findings. As before, we conclude with a discussion of how these findings relate to prior literature, and how this technique may be more widely applied.

5.1 Hypotheses

From prior literature, there are a number of hypotheses to consider about what will distinguish individuals with varying levels of experience. These hypotheses relate to the dimensions of planning, project realization, solution quality, and rate of transitioning. First, one would expect more experienced engineers to spend a greater proportion of time in project scoping or planning (Atman et al., 1999, 2007, 2009) and Adams et al. (2003). Furthermore, this additional planning behaviour should be evidenced both at the onset of the task, and throughout the activity, as the experts utilize a cascading approach. Secondly, one would expect more experienced engineers to spend more time in the project realization phase than those with less experience. Thirdly, the quality of the solutions may not differ very much among the different levels of experience, but this remains to be seen. This hypothesis is based on the divergent results presented in Adams et al. (2003) and Atman et al. (2007). Fourthly, in the terms of rate of transitioning between different activities, prior literature would suggest no significant difference between the different populations.

To the above, we also add the conjecture that experts will spend less time reverting and less time adjusting than their less experienced counterparts, but that they will test their structures more often as part of their iterative design process.

5.2 Algorithm Implementation Specifics

We use the algorithm described in the General Algorithm section, organizing our data into ten clusters in the Segment Clustering step and four during the Participant Clustering step. The number of clusters was set to ten during segment clustering on the basis that this provided the best result for distinguishing among individuals of varying levels of experience. More specifically, when we compared the accuracy of the results from different cluster counts, we found that ten clusters produced the best differentiation between experience levels. Because our objective is to develop a model that helps us understand the differences between the different populations, we are not concerned with over-fitting or confirmation bias. Put differently, our goal is not to create a classifier meant to apply to another set of students. Instead, it is to study this population of students and identify patterns or characteristics that vary by experience level. Furthermore, after we have identified these characteristics we will use qualitative analysis to validate the reliability of the approach. In the case of participant clustering, four clusters were used to align with the four different levels of experience present in our sample population. However, we again note that the objective of this approach was less about making a classifier to predict experience, as it was about understanding the nature of expert experience.

5.2.1 Justification for Approach

As mentioned in the prior literature section, others have employed different approaches for analyzing this type of process data (e.g. Adams 2003; Atman et al., 1999, 2007, 2008). However, most of these approaches have not maintained the temporality of the data, and have instead looked at each student's process in chunks, or in aggregate. Because we are looking for iterative cycles, it seemed fitting to compare entire sequences of actions, as opposed to looking at subsequences, as would be done in sequence mining. Additionally, in previous work, when we explored using non-process-based metrics, we found them far less successful in describing the role of experience in our dataset (Worsley & Blikstein, 2013).

5.3 Object Manipulation Generalizable Segments

During the first clustering phase, we identified the ten Generalizable Action Segments. Figure 17 highlights these differences along the five General Object Manipulation classes. Looking at the figure, there is one cluster for EVALUATE, one cluster for MODIFY, five clusters that represent REVERT, and three clusters related to different combinations of PLAN, REALIZE and MODIFY. To make naming of the clusters easier to follow, each cluster will be given a title that characterizes its primary constituents.

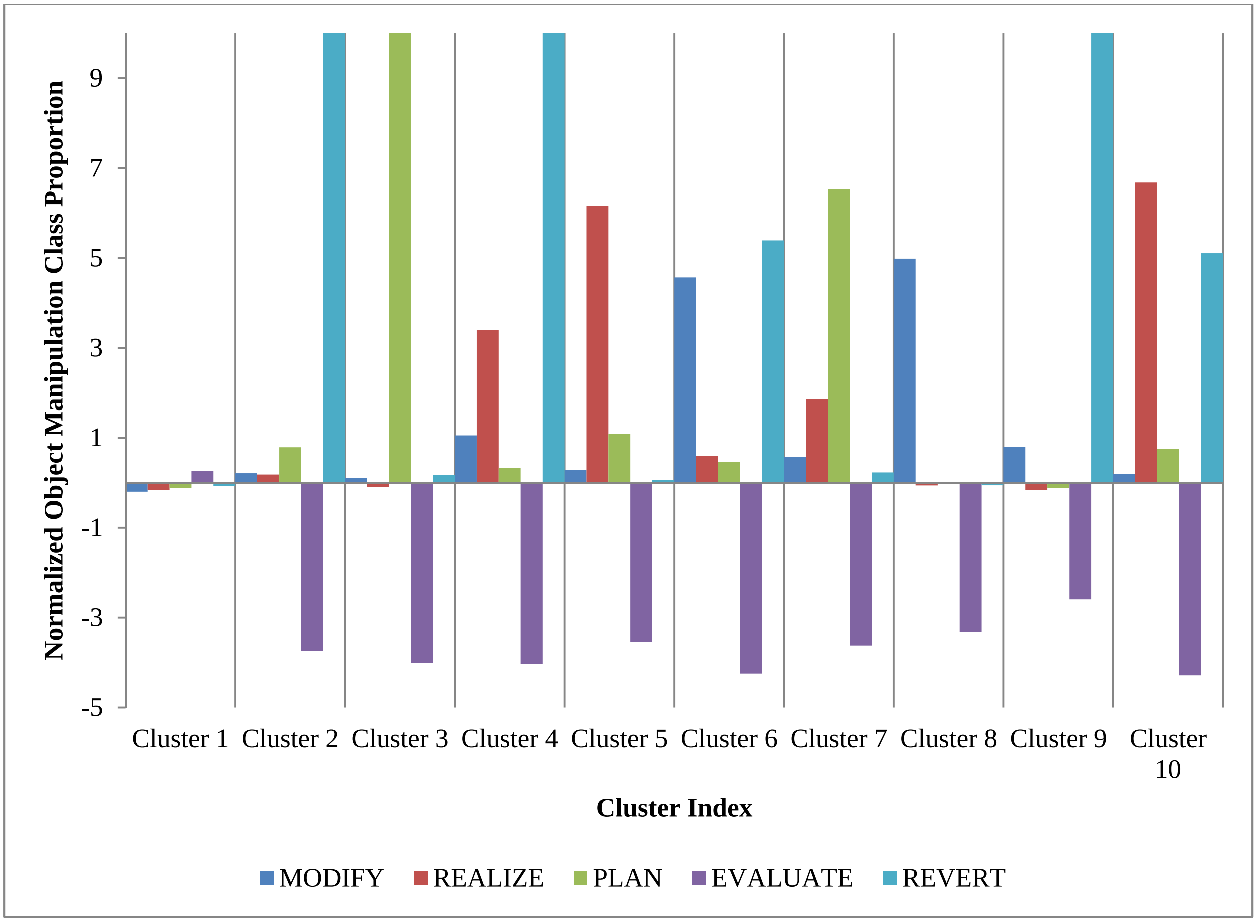

Figure 17. Object Manipulation Class Proportions by Cluster

5.3.1 EVALUATE Cluster

Cluster 1 represents the EVALUATE action, and was used for segmenting the sequence of actions. Accordingly, we expect this to be small in magnitude, and for all of the other clusters to include below average EVALUATE action proportions.

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

5.3.2 REVERT Clusters
Clusters 2, 4, 6, 9 and 10 (Figure 16) correspond to a large amount of REVERT. This suggests that undoing is an important behaviour to pay attention to when studying experience. However, simply looking at REVERT by itself is not sufficient. Instead, one needs to observe what other actions are taking place in the context of the REVERT action. In the case of cluster 2 (aka G-REVERT), the user is performing significant REVERT actions in the absence of any other action. This is in contrast to cluster 4 (G-REVERT-REALIZE), for example, where the user is completing a large number of REVERT actions, but is also doing several REALIZE actions. From this perspective, cluster 2 seems to correspond to doing a sustained REVERT, without any building. An example of this would be a student completely deconstructing the structure. Cluster 4, on the other hand, is more akin to undoing a few elements of one's structure with the intent of immediately modifying it. Put differently, cluster 4 may correspond to microscopic REVERT actions, whereas cluster 2 consists of more macroscopic REVERT actions. Clusters 6 (REVERT-MODIFY-REALIZE) and 9 (G-REVERT-MODIFY) appear to be characterized by a combination of REVERT and MODIFY actions. In this case, the user is undoing, not to make large structural changes to the design, but to make small adjustments. Clusters 6 and 9 are not identical, however. Cluster 6 also contains REALIZE actions. Cluster 10 (REALIZE-REVERT-PLAN) differs from the other REVERT clusters, in that REALIZE is the primary component, followed by REVERT and PLAN.

5.3.3 MODIFY Cluster
Cluster 8 (G-MODIFY) stands alone as a primarily MODIFY cluster, with below average values for all other actions.

5.3.4 PLAN, REALIZE, MODIFY Clusters
The remaining clusters — 3 (G-PLAN-MODIFY), 5 (G-MODIFY-REALIZE-PLAN), and 7 (G-PLAN-REALIZE-MODIFY) — can be characterized as different combinations of PLAN, REALIZE and MODIFY. Cluster 3 was dominated by PLAN actions, whereas clusters 5 and 7 include REALIZE and MODIFY actions.
In summary, we see that five of the cluster centroids emphasize REVERT actions, and the context in which they appear, while the remaining five are aligned with different proportions of EVALUATE, PLAN, REALIZE, and MODIFY actions. We can anticipate that there are distinguishing factors about how each of these are used that will help us as we examine the impact of experience on engineering practice.

5.4 Participant Clusters

Clustering students based on their pair-wise dynamic time-warped distances results in the precision and recall values presented in Table 9. Precision refers to the proportion of items identified for a certain class that actually belong to that class. Precision of 0.5 means that half the students placed into the low experience cluster were actually of low experience. Recall refers to the proportion of items belonging to a certain class that are correctly identified. Recall of one, means that all of the students of low experience were included in the low experience cluster.

Table 9. Precision and Recall for Cluster to Experience Alignment

Experience	Precision	Recall
Low	0.50	1.00
Medium	1.00	0.60
High	0.67	1.00
Expert	1.00	1.00

From Table 9, we see that the algorithm worked best at uniquely clustering Expert behaviour. It also attained recall of 1 for individuals of Low experience. Again, a recall of 1 means that all Low experience individuals were properly assigned to a single cluster. For individuals of intermediate experience, the algorithm was less accurate. Nonetheless, we reiterate that our primary objective is to understand better the patterns that distinguish "relative" experts from "relative" novices. We refer to the students as "relative" experts and novices because we did not employ a universal standard of expertise, but instead based their expertise on the amount of expert experience that they had. Hence, the majority of this analysis will be on examining how this representation of student actions was able to delineate between different levels of experience.

As a first indication that student behaviour differs by experience, we performed pair-wise Chi-Squared analyses (Table 10). Once again, in order to use Chi-Squared we treat every action of each participant individually. Making this simplification has limitations, provides a quick means for comparing across levels of experiences. From Table 10 it appears that all groups differed from one another in terms of usage of the different Generalizable Action Segments. Additionally, based on the Chi-Square statistics, we see that the pair-wise relationships follow the expected trend, with the most similar pairs (Expert-High, High-Medium, Medium-Low) having lower Chi-Square statistics than more dissimilar pairs (Expert-Medium, Expert-Low, High-Low). In order to pinpoint the nature of these differences we return to the three representations used in Analysis 1.

Table 10. Pair-wise Chi-Squared Analysis of Generalizable Action Segment Usage

Group 1	Group 2	Chi-Square Statistic	Probability
Expert	High	20.49911	0
Expert	Medium	47.26303	0
Expert	Low	260.6675	0
High	Medium	43.7318	0
High	Low	81.84317	0
Medium	Low	26.14816	0

5.5 Proportion-Based Graphical Representation of Participant Clusters

As before, we begin with a graphical representation of time spent in different activities. Among this first set of Generalizable Object Manipulation Segments (Figure 18) we see that G-REVERT is only used by individuals of Low experience, and G-REVERT-REALIZE is used more frequently among lower levels of experience. G-REVERT-MODIFY-REALIZE is only observed among individuals of Low and Expert experience. Finally, G-REVERT-MODIFY is relatively high for Medium experience individuals.

From Figure 19, we see that G-MODIFY is used extensively by individuals of all experience levels. We also observe that G-MODIFY-REALIZE-PLAN accounts for a larger proportion of user actions as experience level increases.

From Figure 20, we see that G-PLAN-MODIFY accounts for a larger proportion of time for High experience and Expert individuals than for Low and Medium experience individuals. G-PLAN-REALIZE-MODIFY also appears to follow this trend among individuals of Low, Medium, and High experience.

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

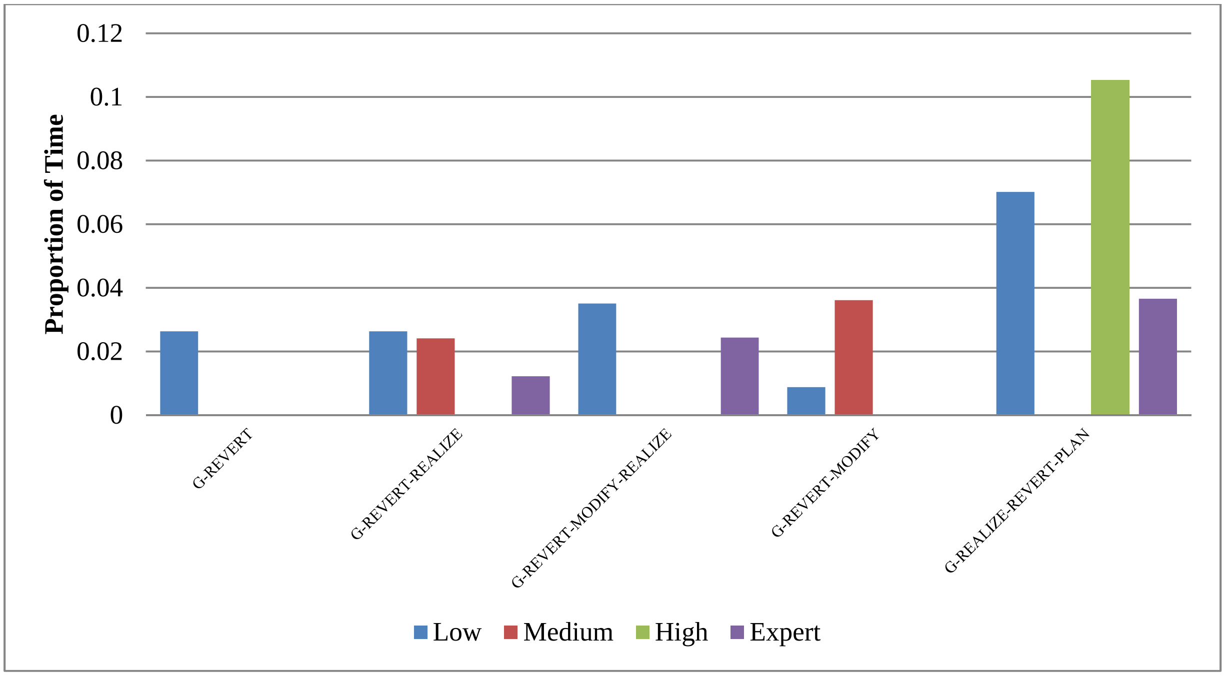

Figure 18. Proportion of Time Spent in each REVERT Action Segment by Experience

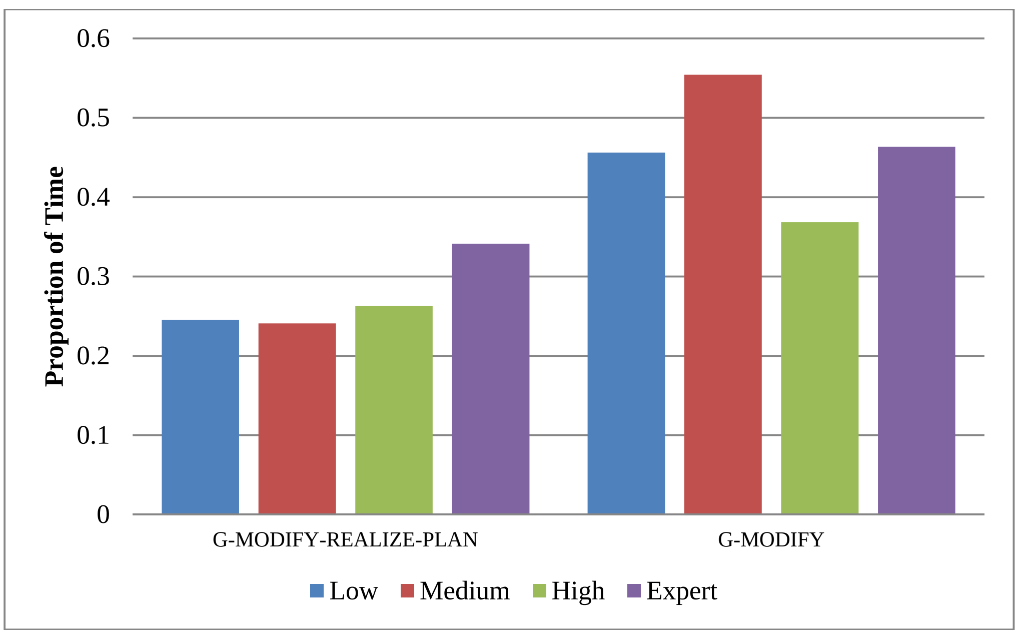

Figure 19. Proportion of Time Spent in each MODIFY Action Segment by Experience

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

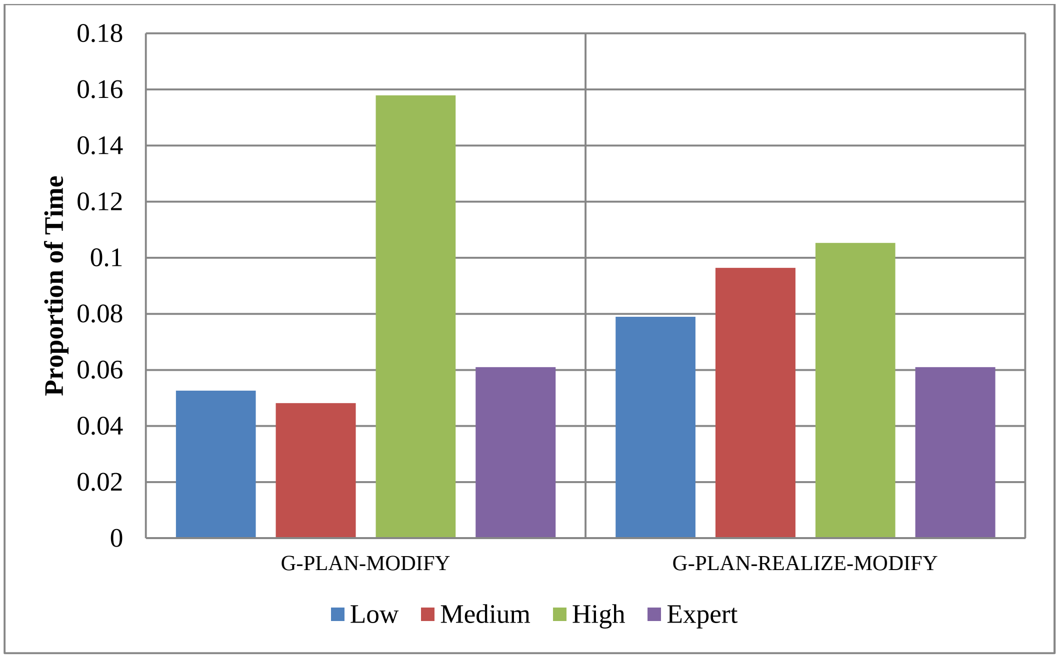

Figure 20. Proportion of Time Spent in each PLAN Action Segment by Experience

As we step back from the specifics of each generalizable segment, we observe that the Low experience population uses all five G-REVERT segments, Medium uses two of the five, High uses one of the five, and Expert uses three of the five. High experience and Expert individuals only used G-REVERT segments that also included REALIZE, with those of High experience having the additional constraint of only using G-REVERT segments that included both REALIZE and PLAN. This provides some initial indication that as experience increases, both PLAN and REALIZE become more central to the building process, which aligns with two of the hypotheses.

5.6 State–Transition Representation of Participant Clusters

In order to study how experience relates to behaviour more closely, we will again turn to a state–transition probability representation (Figure 21). As before, we compute the frequency of transitions between the different generalized segments and examine differences among our population of research participants. Through a pair-wise Chi-Square analysis of transition probabilities (Table 11) we see that all the transition behaviour of Expert does not significantly differ from that of High or Medium, but that significant differences exist among all other pairs. The lack of significant differences between Expert-High and Expert-Medium may initially seem odd, but when one considers that significant differences remain among the overall usage of individual behaviours, this becomes less problematic and may offer some meaningful insights into how experience impacts behaviour. However, we withhold the remainder of this discussion for a later section.

Because of the large number of states, we will only construct the diagrams for the six states associated with significant differences when comparing individuals of Expert experience to people of lower experience. These include the five REVERT states and G-PLAN-REALIZE-MODIFY. This observation in itself corroborates the idea that the frequency and context of REVERT actions is important when studying the role of experience on engineering practice.

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

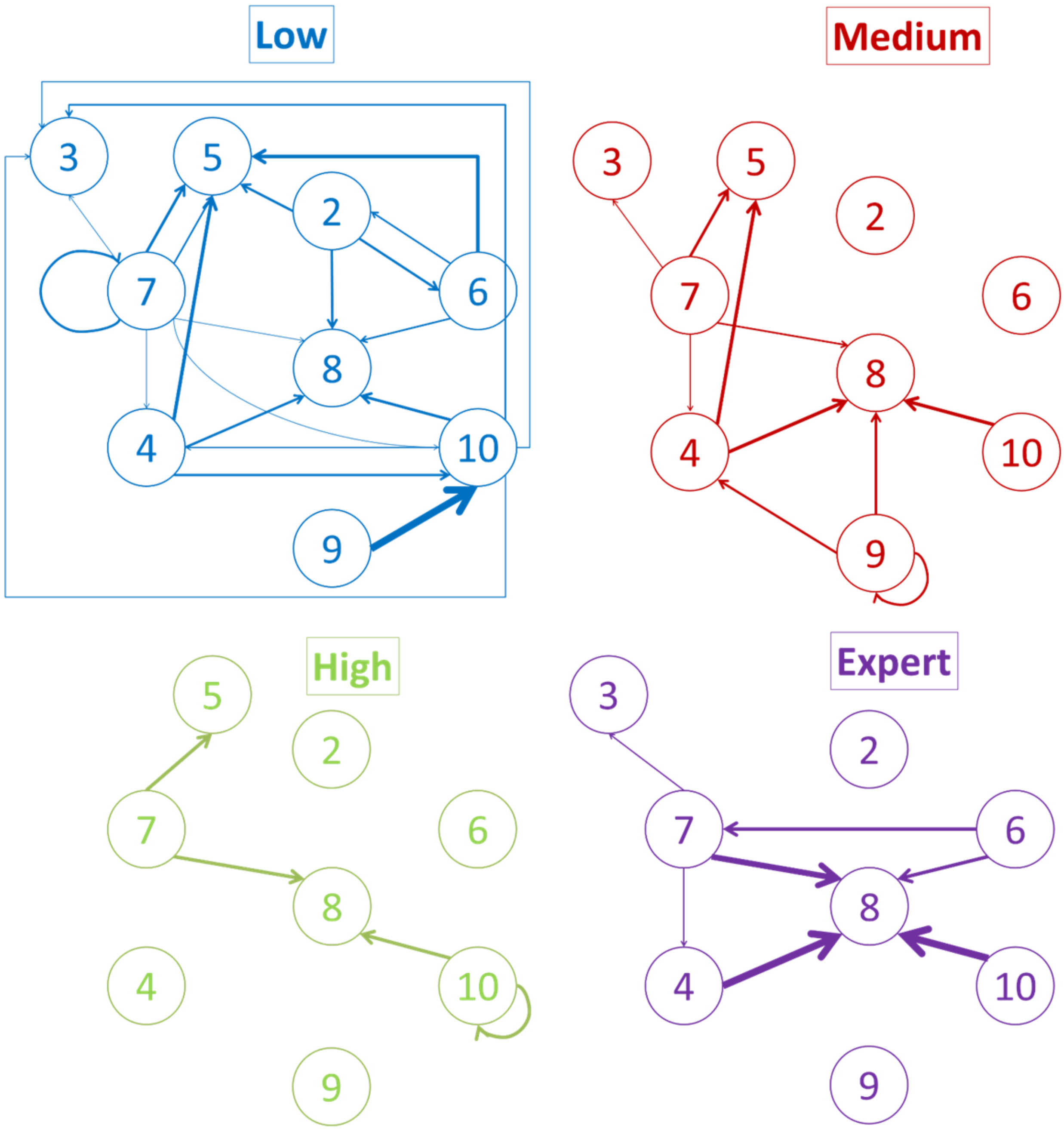

Figure 21. State–Transition Diagrams for Clusters of Different Experience Levels

5.6.1 Expert-Low Comparison

Expert and Low demonstrated differences in the nature and extent of structural undoing, as well as in what prompted them to adjust their structures. The Expert experience group typically engaged in modifications only after REALIZE actions. The Low experience group resorted to modifications through a much larger variety of previous actions.

Table 11. Pair-wise Chi-Square Analysis of Transition Probabilities

Group 1	Group 2	Chi-Square Statistic	Probability
Expert	High	90.25028	0.723537
Expert	Medium	108.9014	0.233152
Expert	Low	626.605	0
High	Medium	354.1774	0
High	Low	1071.301	0
Medium	Low	190.4553	0

These two groups differed in how they transitioned into six of the ten generalized segment: G-REVERT, G-REVERT-MODIFY, G-MODIFY, G-REALIZE-REVERT-PLAN, G-REVERT-REALIZE, and G-REVERT-MODIFY-REALIZE. From the Figure17, we observed that the Expert group never used the G-REVERT or G-REVERT-MODIFY states. They were less likely to transition into G-REALIZE-REVERT-PLAN and less likely to transition into G-REVERT-REALIZE. The Expert group would only transition into this state from G-MODIFY, whereas the Low group would transition into G-REVERT-REALIZE from three different states. This pair also differed in how they transitioned into G-MODIFY with Expert more likely to transition into G-MODIFY from previous states that included REALIZE actions.

5.6.2 Expert-Medium Comparison

Where Expert and Low demonstrated differences in the sequencing of building and modifying, Expert and Medium demonstrated differences in the context in which undoing actions were used. Specifically, the Expert group typically used more complex REVERT actions, meaning that the REVERT was used amidst several other actions.

The Expert group demonstrated differences from the Medium group. These differences were recorded in transitions into G-REVERT-REALIZE, G-REVERT-MODIFY, G-REVERT-MODIFY-REALIZE, and G-REALIZE-REVERT-PLAN. The Medium group never used the G-REVERT-MODIFY-REALIZE or G-REALIZE-REVERT-PLAN actions. On the other hand, the Expert group never used the G-REVERT-MODIFY action. Finally, for the G-REVERT-REALIZE state, the Expert group is more selective in its use, and only does so from G-MODIFY, whereas the Medium group only does so from G-PLAN-REALIZE-MODIFY and G-REVERT-MODIFY.

5.6.3 Expert-High Comparison

High and Expert groups differ in the nature of their planning behaviour. The Expert group is more likely to engage in planning behaviour that is in conjunction with project realization, whereas the High experience group was more likely to enter explicit and dedicated planning sessions.

Here we see differences in four classed: G-REVERT-REALIZE, G-REVERT-MODIFY-REALIZE, G-PLAN-REALIZE-MODIFY, and G-REALIZE-REVERT-PLAN. The High experience group never uses G-REVERT-REALIZE, or G-REVERT-MODIFY-REALIZE. They are less likely to transition into G-PLAN-REALIZE-MODIFY and do so from a smaller number of prior states. Finally, the High experience group is more likely to transition into G-REALIZE-REVERT-PLAN. In addition to these statistically significant differences, we also observed a trending difference on G-PLAN-MODIFY, which mirrors the observation from Figure 20. High experience individuals were more likely to transition into this state than Experts were.

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

In summary, then, the distinguishing factors of the Expert group include:

- Their REVERT actions are always exercised in the context of REALIZE actions. This means that they would be more likely to undo a structure while completing a larger objective that might involve adding a new part of modifying an existing component.
- They use an iterative strategy that involves returning to planning, in the midst of building. For example, students in this group would complete a portion of their design and then enter another stage of planning. In contrast, students from other groups might engage in planning, but then simply move forward towards realizing their design without ever going back to planning.

5.7 Discussion

As we move into the qualitative analysis portion, we will see how many of the differences observed quantitatively are corroborated through video analysis.

5.7.1 Revert Action Context

A key observation made from the quantitative analysis is that Expert individuals complete REVERT actions in the context of REALIZE actions. In order to make this more evident, consider the structure in Figure 22. The user has just added two pieces of green tape to support the structure, and is about to test the strength of her structure.

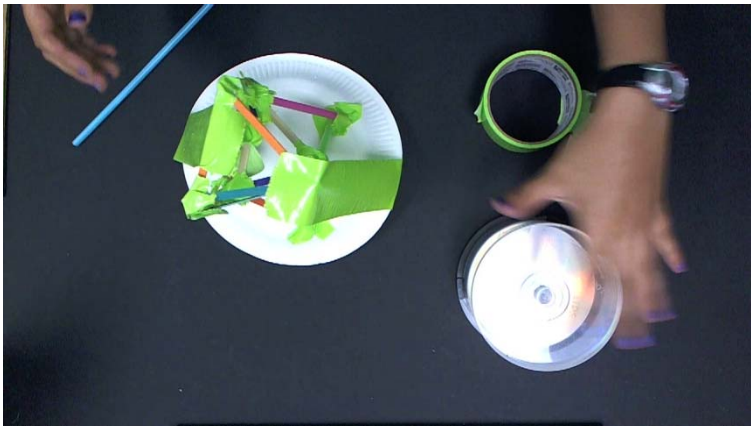

Figure 22. Example of Structure *Before* Undo

After testing the structure, however, she finds that it is not sufficiently stable and requires additional support. She therefore removes the two pieces of tape and employs the light blue straw instead (Figure 23). In this way, we see that while she did revert her design, it was not a matter of completely undoing the structure. Instead, she needed to find a better way to distribute the mass across the structure and correct for weaknesses.

5.7.2 Interspersing PLAN and REALIZE actions

The second observation is that the Expert individuals return to PLAN activities throughout the design process. One way to express this is through a timeline (Figure 24). The nodes on the graph correspond to different Generalizable Action Segments. For the sake of readability, this has been simplified by merging two of the segments that contain PLAN actions into segment number 1. Segment number 2 on the Y-axis corresponds with generalizable states associated with building and adjusting. Thus, we see that this expert began in a planning stage and then transitioned into building and modifying. After completing that cycle of planning, building, and adjusting, the individual returns to planning again around time step 60 on the x-axis, and repeats the process. If we examine the video content at this point (Figure 25), we see that the individual has managed to complete the base of their structure and is now considering what to do next. In fact, one observes in the image that the user is testing the material again while reasoning about the next steps.

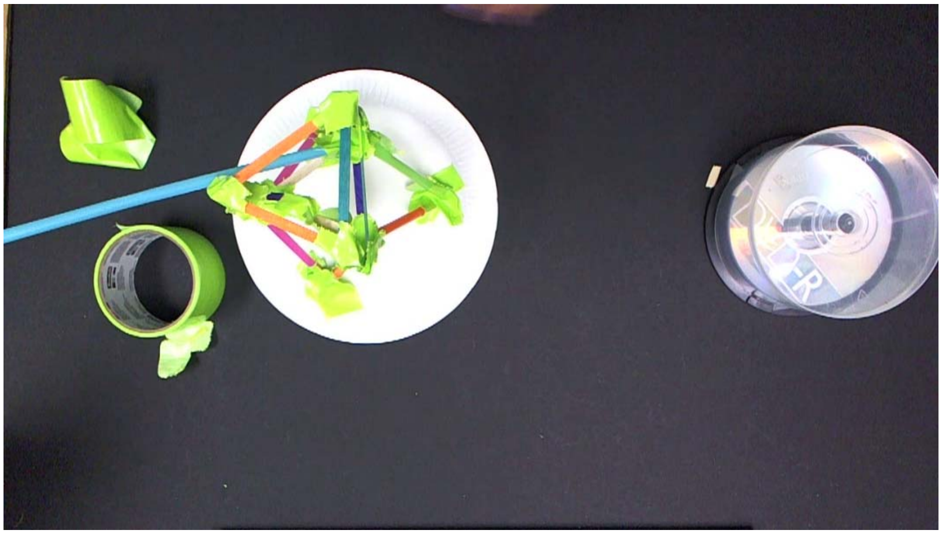

Figure 23. Example of Structure *After* Undo

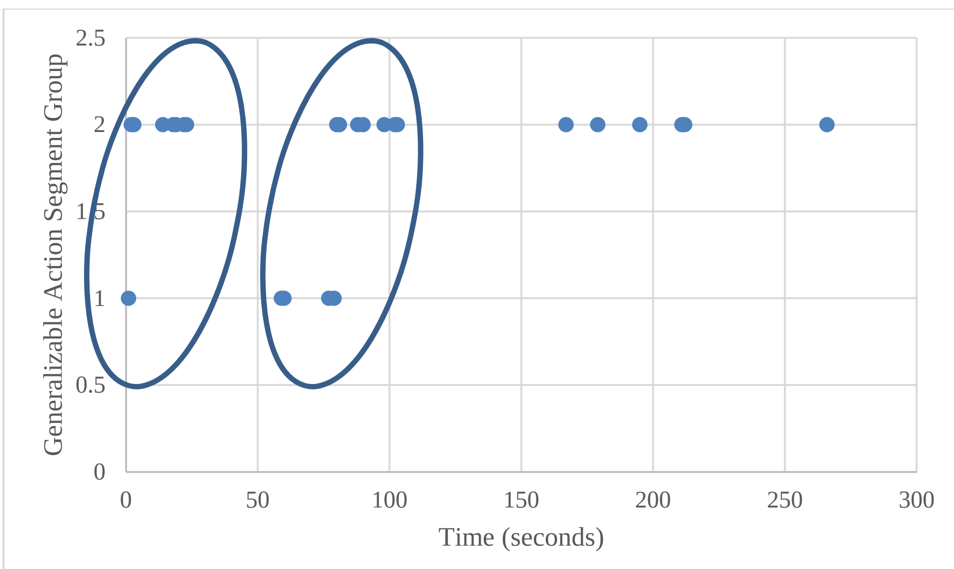

Figure 24. Sample Expert Timeline (1 = PLANNING, 2 = BUILDING and ADJUSTING). Typical Plan and then Build and Adjust cycles are enclosed within the ovals.

(2014). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2), 151-186.

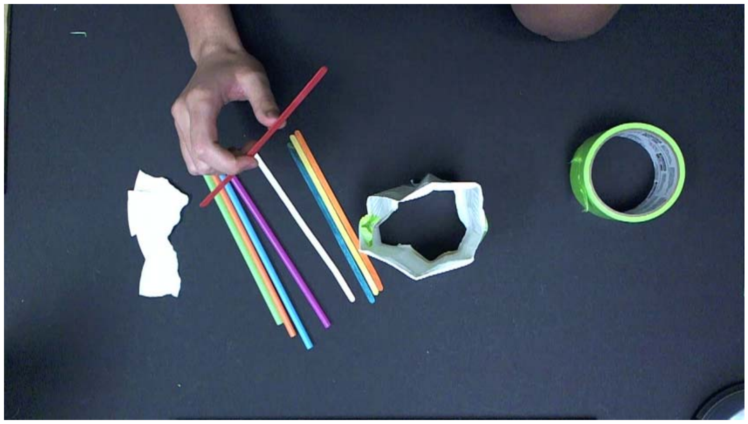

Figure 25. Expert Structure after first iteration of planning, realizing, and adjusting

Later on, we see that the student has made several new additions to the structure (Figure 26), but these additions were only conceptualized during the second iteration of planning and building.

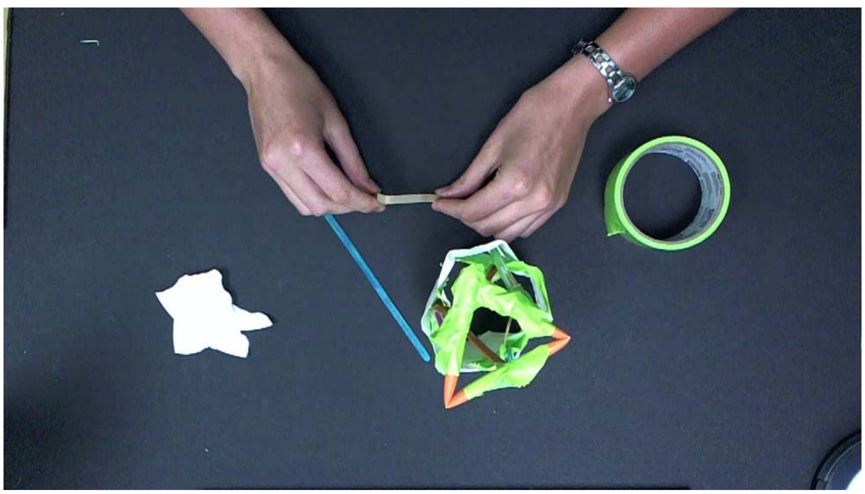

Figure 26. Expert Structure after Second Iteration of Planning, Realizing and Modifying

To corroborate this approach further, the participant described the process as being iterative and requiring flexible planning.

> "I thought that I was going to make this my base [pointing at the folded paper plate in [Figure 25]. And then when I realized that I had extra materials left over, I decided to try and add more height underneath, so that was unexpected. I thought I was going to have this be the base, plus struts coming up."

Once again, there are elements of this analysis that could have been completed using purely qualitative analysis. However, garnering these results, and at a level of granularity that explicates the different types of REVERT actions and the context in which REALIZE and PLAN actions are completed, would have been quite challenging. Through the affordances of computational analysis, we are able to focus the human analysis component on the output of the different algorithms being used.

6 DISCUSSION

In the first analysis, we used machine learning in conjunction with human coding to cluster students based on their action sequences. Those sequences were first transformed into generalized sets of action segments as determined through k-means clustering. The transformed sequences were then used to compute pair-wise distances between all study participants. This process-based analysis generated four clusters of students that differed along two dimensions. The first dimension relates to how students engaged in the design process. Consistent with prior literature, we identified a group of students that employed iterative design practices (e.g. Atman & Bursic, 1998) and others who tended to follow less systematic approaches (failing to iterate, or use incremental development strategies). While we did not focus on students' level of experience in that analysis, we did observe that the majority of individuals in the iterative design group had Expert-level experience with engineering. At the same time, being among the iterative designers did not necessarily mean that an individual had Expert-like experience. This is consistent with the prior literature in this domain, which states that novices can also use iterative design strategies.

We also observed that the clustering analysis differentiated students on the axis of idea quality. We saw that students who spent considerable time undoing previous actions oftentimes overlooked key engineering intuitions that would have made their structures more stable. Without knowledge of how to correct their problems, students employed noticeably different building patterns. What this means for engineering education, especially at the K–12 level, is that we cannot focus only on the engineering design process. Instead, we have to ensure that students also find ways to develop their knowledge about deep engineering intuitions. Examining the extent to which employing iterative design helps students investigate engineering principles in order to develop more accurate intuitions about how their structures will behave. In particular, activities based on pure "trial and error" are likely to be inefficient at helping novices effectively decipher sound engineering principles. This is significant because in many engineering and "maker" programs at the K–12 level, there is a popular belief that letting students tinker will eventually generate more advanced knowledge about engineering and computer programming. However, it is quite possible that many of these students are learning about the engineering design process, without truly gaining insights into engineering principles.

Additionally, using this type of process-based analysis can help educators gain a deeper understanding of a student's conceptual challenges. For example, students who are unsuccessful at an assigned task may be dealing with challenges in the design process, in engineering intuitions, or in basic engineering principles, or any combination of those. Furthermore, even for those students who are successful in completing a given task, identifying where the student falls on these two dimensions can help instructors streamline their learning interventions, and reduce the likelihood that teachers provide students with non-applicable suggestions.

In the second analysis, we performed a more targeted comparison of how more experienced students tackle engineering design challenges. Using the same dataset and the same overall approach as the first analysis, we again identified a set of generalizable action segments. These generalized segments are more fine-grained than those from the first analysis, and consisted of five variants of undoing (e.g. G-REVERT-MODIFY), a variant of modifying (G-MODIFY), and three variants of the combination of planning, realizing, and modifying (e.g. G-PLAN-MODIFY). The level of specificity in these segments helped us to realize the different ways that the five basic actions (EVALUATE, REVERT, MODIFY, REALIZE, PLAN) are used. For example, some REVERT actions were completed in the absence of any other actions, while others were completed in conjunction with different proportions of REALIZE, MODIFY and PLAN. At a higher level of comparison, the second analysis also demonstrated how individuals with different levels of experience use the five basic actions. Even when the same action is used, it often may be employed in a different context. For example, Experts had a tendency to return to planning midway through the process, while other groups would do batch planning. Hence the scope of a given PLAN action differed by experience, with Experts using the PLAN step to address shorter-term goals and objectives. In this way, the second analysis helped disambiguate among similar actions, and suggested ways that experience impacts how students approach engineering design and how it evolves over time and practice.

7 LIMITATIONS AND FUTURE WORK

A key portion of this analysis was the human labelled video data. This provided a coarse-grain sequencing of each participant's actions. One future direction for this research is to leverage the hand-labelled data, in conjunction with computer vision and the gesture data, to label student actions automatically. However, to date, the open-ended nature of the task means that individuals go about enacting each of the possible actions in very different ways. This has been the primary hindrance to training a classifier to detect the different actions accurately. Nonetheless, as the field continues to improve the sophistication of multimodal analysis techniques, automatically extracting data from open-ended tasks will become increasingly feasible.

Another area for future work is to examine the extent to which employing iterative design helps students investigate engineering principles. Analysis 1 presented quality of design and quality of idea as orthogonal dimensions. However, it may be that the two dimensions are related to one another.

8 CONCLUSION

With the expansion of "making" in education, complex hands on learning environments are receiving a lot of attention without having a significant research base. The analyses reported in this paper were motivated by a desire to study complex hands on learning. Furthermore, the goal was to identify some keys insights into understanding how study develop and demonstrate proficiency within the hands-on learning context. Traditionally, analyzing video from hands-on learning has been extremely difficult, labour intensive, and hard to describe in discrete quantitative terms. It is also challenging to see the evolution of subtle patterns that might not reveal themselves using traditional statistical approaches. However, contrary to prior research in this area, our primary data source was not student speech but student actions. Because of the lack of computational approaches for extracting this data automatically, we relied on human coding to provide time-stamps of when students started and stopped different object manipulation actions. We then took this data and performed a sequence of machine learning algorithms in order to study general building practices and to highlight manifestations of relative expertise in engineering design actions. The first analysis showed how we can garner similar results to prior qualitative research, but by using computational clustering techniques and dynamic time warping. It also highlighted that idea quality and design process are two prominent dimensions through which to compare and contrast student development. In the second analysis, we showed how a similar approach could also be used to better identify the characteristic behaviour differences between individuals with various levels of prior experience. Specifically, we showed that both the intent and context of engineering practices is closely related to students' level of experience. For both studies, we validated our findings through visual and qualitative representations that helped us distill what the different clusters of action segments and different clusters of participants might tell us about engineering design patterns.

As we close, we want to take a step back and consider the larger implications of this work, beyond improving our understanding of engineering practices. There is a tremendous opportunity for Learning Analytics to intersect with qualitative research methods to tackle questions that do not lend themselves to easy data extraction. Embarking on work that lies at this intersection will continue to help bridge the learning community and the analytics community. To date, we know of very few instances of Learning Analytics research that takes human labelled data and exhibits how computational analysis can mirror, and extend, approaches and results achieved through traditional education research. If we can show them robust methods for streamlining their analyses, while also permitting them to remain in their current areas of specialization, we have the potential to bring Learning Analytics more squarely into the fold of education research. This will serve to improve the quality and scalability of current education research, and thus increase the impact of Learning Analytics.

REFERENCES

Adams, R. S., Turns, J., & Atman, C. J. (2003). Educating effective engineering designers: The role of reflective practice. *Design Studies, 24*(3), 275–294.

Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering design processes: A comparison of students and expert practitioners. *Journal of Engineering Education, 96*(4), 359–379.

Atman, C. J., & Bursic, K. M. (1998). Verbal protocol analysis as a method to document engineering student design processes. *Journal of Engineering Education, 87*(2), 121–132.

Atman, C. J., Chimka, J. R., Bursic, K. M., & Nachtmann, H. L. (1999). A comparison of freshman and senior engineering design processes. *Design Studies, 20*(2), 131–152.

Atman, C. J., Deibel, K., & Borgford-Parnell, J. (2009). The process of engineering design: A comparison of three representations. *Proceedings of the 17th International Conference on Engineering Design (ICED '09).* August 24–28, Palo Alto, CA, USA (pp. 483–494).

Atman, C. J., Kilgore, D., & McKenna, A. (2008). Characterizing design learning: A Mixed-Methods study of engineering designers' use of language. *Journal of Engineering Education, 97*(3), 309–326.

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.

Berland, M., Martin, T., Benton, T., Ko, P., & Petrick-Smith, C. (2013). Using learning analytics to understand the learning pathways of novice programmers. *Journal of the Learning Sciences*, 22(4), 564–599.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *Washington, DC: SRI International*.

Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. *Proceedings of the First International Conference on Learning Analytics and Knowledge* (LAK '11). February 27–March 1, Banff, AB, Canada (pp. 110–116). New York: ACM.

Blikstein, P. (2013). Multimodal learning analytics. In Dan Suthers, Katrien Verbert, Erik Duval, and Xavier Ochoa (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)* (pp. 102–106). New York: ACM.

Blikstein, P., Worsley, M., Piech, C., Sahami, M., Cooper, S. & Koller, D. (in press). Programming pluralism: using learning analytics to detect patterns in novices' learning of computer programming. *Journal of the Learning Sciences.*

Chi, M. T., Glaser, R., & Rees, E. (1981). *Expertise in problem solving.* Learning Research and Development Center, University of Pittsburgh.

Cross, N., & Cross, A. C. (1998). Expertise in engineering design. *Research in Engineering Design, 10*(3), 141–149.

Dong, A., & Agogino, A. M. (1997). Text analysis for constructing design representations. *Artificial Intelligence in Engineering, 11*(2), 65–75.

Dym, C. L. (1999). Learning engineering: Design, languages, and experiences. *Journal of Engineering Education, 88*(2), 145–148. doi:10.1002/j.2168-9830.1999.tb00425.x

Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. *Journal of Engineering Education, 94*(1), 103–120.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, *87*(3), 215.

Guzdial, M. (2008). Education paving the way for computational thinking. *Communications of the ACM, 51*(8), 25–27.

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. *The Cambridge Handbook of the Learning Sciences* (pp. 61–78)*.* Cambridge: Cambridge University Press.

Lau, K., Oehlberg, L., & Agogino, A. (2009). Sketching in design journals: An analysis of visual representations in the product design process. *Engineering Design Graphics Journal, 73*(3), 23–29.

Lawler, R. W., & Yazdani, M. (1987). *Artificial Intelligence and Education: Learning environments and tutoring systems* (Vol. 1). Bristol, UK: Intellect Books.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas.* New York: Basic Books.

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012). Modeling how students learn to program. *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education,* (SIGCSE '12) (pp. 153–160). New York: ACM.

Resnick, M. (2002). Rethinking Learning in the Digital Age. In *The Global Information Technology Report: Readiness for the Networked World*. Edited by G. Kirkman. Oxford University Press.

Resnick, M., Martin, F., Berg, R., Borovoy, R., Colella, V., Kramer, K., & Silverman, B. (1998, January). Digital manipulatives: New toys to think with. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 281–287). New York: ACM Press/Addison-Wesley.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education, 92*(3), 499–525.

Siemens, G., & Baker, R. S. D. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (pp. 252–254). New York: ACM.

Turkle, S., & Papert, S. (1991). Epistemological pluralism and the revaluation of the concrete. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 161–191). New York: Ablex Publishing Corporation.

U.S. Department of Education, Office of Educational Technology (2010). Transforming American Education: Learning Powered by Technology. National Education Technology Plan 2010. Washington, DC.

Wilensky, U., & Reisman, K. (2006). Thinking Like a Wolf, a Sheep, or a Firefly: Learning Biology through Constructing and Testing Computational Theories—An Embodied Modeling Approach. *Cognition and Instruction*, *24*(2), 171–209.

Worsley, M. (2012). Multimodal learning analytics: enabling the future of learning through multimodal data analysis and interfaces. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (pp. 353–356). New York: ACM.

Worsley, M., & Blikstein, P. (2011). What's an expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis. *Proceedings of the Fourth Annual Conference on Educational Data Mining (EDM 2011).* Eindhoven, the Netherlands (pp. 235–240).

Worsley, M., & Blikstein, P. (2013). Towards the development of multimodal action based assessment. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 94–101). New York: ACM.