Title of Study: Beyond Surveys and Data Mining: Searching for New and Potentially More Useful Evidence of Disciplinary Engagement

Principal Investigator Name and Department: Daniel Hickey, Counseling and Educational Psychology

Year/Semester Awarded: 2014-2014

Number of graduate students who were subjects of your study: 40

Executive Summary: An exploratory analysis of an innovative open online courses was conducted on data from discussion threads in order to identify potentially useful evidence of productive student engagement. Social network analysis was used to identify patterns of interaction and engagement between types of users. The graphical representations of these interactions were used to identify central actors within the courses and to yield compelling evidence of engaged participation to include in web-enabled microcredentials offered to students.

Narrative: Please see attached paper.
Beyond Surveys and Data Mining: 
Searching for New and Potentially More Useful Evidence of Disciplinary Engagement

Daniel T. Hickey 
Joshua D. Quick 
James E. Willis III

Indiana University 
Center for Research on Learning and Technology

Abstract
An exploratory analysis of an innovative open online courses was conducted on data from discussion threads in order to identify potentially useful evidence of productive student engagement. Social network analysis was used to identify patterns of interaction and engagement between types of users. The graphical representations of these interactions were used to identify central actors within the courses and to yield compelling evidence of engaged participation to include in web-enabled Microcredentials offered to students.
Beyond Surveys and Data Mining: Searching for New and Potentially More Useful Evidence of Disciplinary Engagement

Surveys of student engagement in higher education have become quite well known and widely cited in recent decades. In particular, the National Survey of Student Engagement (NSSE) has become a widely accepted index of engagement. However, NSSE and other such measures rely on self-report items asking about students’ activities (e.g., meeting with faculty, engaging with classmates, extra-curricular work, etc.) and student satisfaction with those activities. While response rates tend to be quite low (generally around 20%, Herzog & Bowman, 2011) and explained variance even lower (usually less than 20%, Coates & Ainley, 2007), the relativistic nature of Likert scale items and mass survey distribution yields statistically significant difference of all manner. Despite the limitations, these difference are widely cited in the news and deeply scrutinized by institutions. While there is certainly evidence that this information leads to institutional efforts that increase the availability of and satisfaction with surveyed activities, there is apparently little evidence that survey-driven changes have any impact on engagement in learning.

In ways that are both similar and different to survey studies, the field of Educational Data Mining (EDM) has grown in prominence in recent years (e.g., Baker & Yacef, 2009) and now features its own journal and conference. Outside of educational contexts, data mining is often referred to as Knowledge Discovery in Databases (KDD, emphasis added). What distinguished data mining is the use of sophisticated algorithms and powerful computers to “discover” meaningful patterns in large data sets. Not surprisingly, many EDM researchers came to the field as Computer Scientists.

Like survey research, EDM is relatively agnostic about the nature of learning. Just as survey methods don’t distinguish between what students are discussing when they meet in reportedly-engaging class discussion groups, EDM is typically agnostic about the content of student interactions that are being analyzed. Put differently, both survey research and EDM are “naturalistic” methods that are primarily intended to gather a detailed picture of the current state of affairs. This means that neither are “interventionist” methods that are primarily concerned with gathering useful information that directly informs educational practices (as articulated by Lagemann, 2002).

Significantly, the field of EDM is now clearly distinguishable from the related field of Learning Analytics, which also features its own journal and conference. While the fields overlap significant, there is an important distinction. Rather than data discovery, Learning Analytics research tends to start with conjectures about learning if not actual innovations of specific practice. It then examines educational data to answer questions about that learning or innovation. In this sense, Learning Analytics is a relatively interventionist approach to educational data.

Beyond Surveys and Educational Data Mining in the Learning Sciences

Learning Sciences is an inherently interventionist field. It is also an interdisciplinary field. Generally speaking, learning scientists aims to bring new insights to bear on enduring problems in education that have eluded resolution. Learning scientists generally engage in design-based research (DBR, Cobb, et al, 2003), aiming to build useful “local” theory in the
context of efforts to reform practice. In doing so they reject the conventional distinction between “basic” and “applied” research (as embodied in the distinction between cognitive psychology and educational psychology). Many learning scientists embrace Lagemann’s (2002) argument that the most useful theories for reform in educational practice are theories that are refined and validated in the context of efforts to reform practice. Needless to say, many learning scientists question the ultimate value of naturalistic educational research that is agnostic as to the nature and forms of learning being studied, likewise, most individuals who identify themselves as learning scientists generally find more affinity with the Learning Analytics community.

This study used methods and perspectives from the Learning Sciences in an effort to break new ground on an enduring challenge in online learning. Many researchers have documented what most students who have taken online courses have found: the discussions in forums can be downright awful. The same students who engage in endless threaded discussions with others in interest-driven social networks often only post whatever they need to earn a grade (see Hara, Bonk, & Angeli, 2000). This study aimed to explore the extent to which conjecture-driven learning analytics might be used inform an ongoing effort to enhance the quality and quantity of student peer engagement in online courses.

**Related developments.** Five recent developments encourage educational researchers to move beyond survey research and data mining and into more interventionist efforts in the Learning Analytics tradition. The first is the growth of “interest-driven” (as opposed to “friendship-driven”) digital knowledge networks. Such networks are defined by disciplinary knowledge and increasingly participatory knowing and learning of the current generation of students. This motivates a search for more discipline-specific measures of engagement within networked contexts. The second development is the increased use of learning management systems like Sakai and Canvas in conventional and online courses, and the corresponding access to information about actual engagement in learning of every student. While many current “data mining” efforts appear quite scattershot and assumption-free, this information clearly has the potential to yield more discipline-specific measures of networked engagement.

A third development that motivates a search for new assessments is the growing consensus around the notion of “productive disciplinary engagement” among many learning scientists. As introduced by Engle and Conant (2002), “PDE” concerns social discourses and cultural practices associated with widely acknowledged disciplines of knowledge. According to Engle and Conant, engagement is presumed to be disciplinary when contributions are coordinated and on-task, and concern the specific languages and practices associated with recognizable “big D” Discourse communities (Gee, 1999). Disciplinary engagement is presumed to be productive when it raises new relevant questions, clarifies confusion, makes connections, and becomes more sophisticated. In addition to capturing more fine-grained evidence of engagement in actual learning, Engle and Conant introduced a set of learning design principles for directly fostering PDE. Research in the primary author’s own courses (e.g., Hickey & Rehak, 2013; Hickey, Kelly & Shen, 2014) and by others (e.g., Forman et al., 2014; Greeno & Forman, 2013) has shown that these design principles can be remarkably simple to implement in ways that directly enhance the quality and amount of PDE. What is remarkable about these principles is that they appear capable of readily fostering PDE without also undermining the validity of any assessments of PDE. When coupled with contemporary design-based research methods (e.g.,
Cobb et al., 2004), this presents a very compelling vision of iterative semi-automated cycles of course refinement and improvement.

The fourth relevant development is the emergence of a particularly relevant sub discipline of Social Learning Analytics (SLA, Shum & Ferguson, 2012). SLA emerged in part out of efforts at the Open University in the UK to support interactive social learning at scale (Parr, 2014). The continued growth and development of these analytics and measures suggests the value of searching for new metrics and assessments of learning, participation, and engagement. These factors motivated our investigation into the application of learning analytics to data collected in an open online course. We sought to determine and visually represent the patterns of participation and engagement in order to understand the development of central actors within the course and the formation of collaborative networks. Thus a very specific goal of this study is exploring how the methods associated with SLA might further a very similar set of goals of support “participatory” forms of social learning at scale.

The fifth development is the introduction of open digital badges. These are web-enabled digital credentials that contain specific claims of competency and detailed evidence supporting those claims. Most importantly, open digital badges can contain links to additional evidence (such as completed student work or evidence of participation) and this information can all circulate readily in social networks. This allows credentialing to take advantage of the way information gains new meaning as it circulates on the web. This in turn is expected to democratize education by moving the locus of credibility away from schools and accreditors and putting in in the hands of educators and learners. As documented by Hickey, Willis, and Quick (2015), the process of deciding what claims and evidence to place in badges and then deciding what assessments practices might generate that evidence and how to insert that evidence is badges is, at minimum, transformative for institutions. Many institutions experienced the introduction of badges as very disruptive process. Badges were introduced in part to allow educational programs to recognize and reward newer more networked forms of non-school and open learning.

**Research Context**

This study used data from an open online course offered at university in the Midwestern USA in summer 2015. The course was entitled *Introduction to Educational Data Sciences*. Building on the distinctions proposed in Piety, Hickey, & Bishop (2014), the course was intended to explore the distinctions between EDM, LA, and three other seemingly related areas (*institutional analytics* in higher education, *institutional/systemic analytics* in K-12 schools, and *individualization/personalization* as is often done with digital tutoring systems).

This university had recently switched to the Canvas learning management system. Instructure Inc., the vendor, was in the process of designing a data definition dictionary that would organize efforts by instructors, staff, and administrators at client universities to request data sets. The university had organized a small grant funding program to encourage instructors, staff, and administrators to carry out innovative learning analytics with this soon-to-be available data. This paper is a report of one such effort carried out in the offered course.

**The Intro to EDS course.** Forty individuals registered for and started the Intro to EDS course. This included eight students who enrolled and completed the course for three graduate-
level credits in a special topics seminar. Of the 32 open (i.e., non-credential) participants, 6 completed the course. Some, but not all of the students who did not complete the course were relatively active until they withdrew.

This course was designed to encourage personalized disciplinary engagement through the construction of “wikifolios” that enabled commenting and discussion from students, instructors, and guest discussants (see Figure 1). The assigned readings were selected in discussion with the authors and consultation with experts from the field. Each week students drafted and posted a wikifolio. Technically speaking, each wikifolio was actually the header of a new discussion forum in the Canvas learning management system. This was a necessary work-around because Canvas (rather inexplicably) did not include the option of adding threaded discussions to student-generated pages. In the first wikifolio, students read Piety, Hickey, & Bishop (2014) and then drafted an “EDS challenge” that they embodied their professional experience, interests, and aspirations. In the remaining wikifolios, they engaged with the assigned paper by summarizing the relative relevance of key aspects of the assigned papers to their EDS challenge, in order of decreasing relevance. They then searched for other articles the referenced the assigned reading but was also relevant to their EDS challenge. They then posted summaries of the three most relevant external references.

Perhaps half of the students’ time and most of the instructor’s time was committed to participation in discussions that took place in discussion threads posted directly on student wikifolios with the instructor and eight guest discussants (usually the author of the assigned reading). As expected, these discussions were quite lively and extensive. Many of the guest discussants continued to return to the threads and other outside experts were sometimes asked to comment as well. This resulted in 1,419 comments posted as parts of threaded discussion on the 11 weekly wikifolios. These comments were analyzed to explore the research questions presented next.

**Research Questions**

Our first three research questions follow from the design principles that inform the design of the course. One of the principles is that having students discuss their work in threaded comments directly on their wikifolios (rather than removed to discussion forums) will result in more contextualized and robust engagement. This leads to the first research question: *Can we use Learning Analytics to empirically document what we believe to be remarkably levels of engagement focused on the topics of the weekly assignments?*

Another course design principle is that assigning students to “networking groups” will encourage robust engagement by functioning as an “affinity group” for like-minded professionals. The group assignment was relatively informal and so it was possible that new networking groups would emerged beyond the self-identification to the five “EDS areas” used to organize the course. In prior courses where this methods was used, we had observed that the majority of interactions occurred within groups, but we also observed highly productive exchanges that took place across groups. This led to our second question: *Can we use Learning Analytics to examine the extent to which these networking groups shaped interaction in the EDS course?*
A third course design principle was that productive disciplinary engagement should be publically rewarded. Open (non-credential learners) were offered evidence-rich digital badges for completing each module, and that badge contained all of the work that students did, including their discussion with their peers. We were interested in exploring ways of using learning analytics to further document the participation of highly engaged students. This led to the third research question: Can we use Learning Analytics to uncover new ways of readily documenting and rewarding particularly engaged participants? A specific challenge with the existing badges is that they only showed participation on one’s own wikifolio; they did not show all of the work that students did interacting with others by posting comments on their wikifolio. As this was a particularly important form of course engagement, we were searching for new ways to recognize and reward it.

Our fourth research question follows from our strong commitment to using design-based research methods to directly enhance participation in productive disciplinary engagement and indirectly enhance the understanding and achievement that results from that engagement. Our prior refinements had been driven by relatively subjective analyses of the changes in participation following modifications and validated analysis of learning gains on course achievement tests. Our effort to advance and perhaps automate these efforts leads to our fourth research question: Can we use Learning Analytics to further inform our efforts to iteratively refine this course design framework and the features used to enact it?

Analysis and Results

The data were pre-processed to ensure that all listed participants were accounted for and no duplicate records existed. Next, a network graph (Figure 2) was generated from the participants’ discussions and descriptive measures of the network were computed to identify relevant features of the participation within the courses. While these analytics are largely exploratory, some interesting initial findings were found. For instance, the identification of central actors within the courses was of particular interest to our research questions of searching for patterns of engagement. With that in mind, we computed several centrality measures and modularity detection algorithms. Unsurprisingly, the instructor played a central role in the course. As expected, this network displays the high degree of interactivity and participation of members of the course. However, at this time, we do not have any bases for comparing these findings to other online courses.

Regarding the second question, the network graph shown in Figure 2 reveals that the label of “primary EDS group” that students selected in the second week of the course do not appear to have contributed much to the structure of the social interaction. Much of the interaction occurred between networking groups throughout the course. Quite simply put, were it not for the color of the nodes, it would be impossible to identify from the network graph that there were any networking groups at all in terms of patterns of discussion. This appears to confirm two observations made during the course: these five areas did overlap substantially, and these students were all relatively new to EDS and still did not have a particular affinity with the networking groups they selected.

Regarding the third question, the Learning Analytic confirmed that one open learner, Sandy Stringfellow, was particularly deeply engaged with the course and with her peers and the
guest discussants. The network graph revealed that Ms. Stringfellow (who indicated she wanted to be identified in this paper) was the most active participant in the course, rivaled only by the instructor. In addition to the extensive discussions that are documented on her own wikifolios, this image shows that she was equally active commenting on the wikifolios of others.

This led us to define an additional version of the course badge specifically for Engaged Participation that indicated that the badges was awarded to a single participant for their exceptional engagement in the Intro to EDS Course (Figure 3). It turns out Ms. Stringfellow is an EdD student in instructional technology who wants to pursue a career in EDS, but her program does not offer any courses on that topic. Given that the guest discussants were many of the leaders in the field, the badge served to summarize her very deep professional engagement as a participant in the course. Consistent with the ideal of digital credentials, viewers of the credential can “drill down” into the evidence by clicking the embedded course badge, clicking each of the five embedded module badges, and then clicking on each of the evidence links in the badges to see the part of her engagement appeared on her page; the graph dynamically displays the intensity of the other half.

Regarding the fourth question, this initial analysis does not seem to have yielded any actionable information for further refining the Intro to EDS course so far. We are now turning our attention to SLA methods and text-analytics to search for additional strategies for improvement.

Qualifications and Reflection on the Use and Process of Learning Analytics

Several issues should be noted in the interpretation of these findings, however. First, no control measures were taken to subset the data according to the discourse type (i.e., whether a post displayed substantive content relevant to the assignment and discussion content). The identification of relevant discursive activities within the comments represents a future goal of our research and one we will be pursuing with the assistance of leading experts in educational data science as consultants. Second, these results are inherently descriptive indicators of the network and do not reflect its evolution and development across time. As such, our current results represent a very limited and exploratory set of findings. This is largely due to several issues we encountered in our acquisition of the data.

Initially, we expected to perform multiple, converging, and socially-driven learning analytics that focused on determining the engagement and participation of students within two BOOCs. Our proposed methods were to involve performing and developing natural language processing tools to develop detectors for levels of engagement and participation. Prior to participating in learning analytics research, we developed a coding process to identify levels of disciplinary engagement of students, which we expected to use in the development of a more computationally-driven and efficient process of detecting levels of engagement and patterns of student participation in two BOOCs designed and intended to promote peer driven participatory networks. These methods were to be developed in consultation with leading experts on learning analytics and educational data mining. As our participation in the research continued, however, several hurdles were encountered that shifted our methods to the development of social network analytics to identify patterns of students’ participation.
Chief among these hurdles was a noticed lack of infrastructure in getting our researchers access to relevant data. While this is understandable considering said infrastructure was being developed concurrently with our research, we did find the lag of getting access compromised some of our initial expectations of developing time-sensitive analytics that could inform the development and direction of the courses while running. Such a constraint presents a substantive issue in using learning analytics as decisions based off such analytics are likely to change with each offering of the course. Further, such a lag presented difficulty in performing more intensive and detailed content analyses due to the lack of available data from the LMS and the logistics of scheduling time for consultation with experts. This resulted in us examining and selecting less intensive methods of analysis. Currently, we intend to conduct more extensive analyses on this and other courses designed in this framework.

While we believe our data was and remains appropriate to answering our research questions in both the discussed and other courses, another issue we encountered centered on whether we had the ethical rights to view and use these data. Indeed, as these data resulted from educational practices, we had several concerns on whether it would be inappropriate to make educational decisions on the basis of analytic results. This issue remains a relevant factor as we continue to negotiate how to effectively use and conduct analytic research to inform our pedagogical practices. A final challenge that we noted and are currently addressing centers on what analytics and measures can be generated to show evidence of students’ participation and engagement with the content as productive and effective learning. As stated, we are currently investigating the possibility of developing digital badges on the basis of our analytics, though such practices are largely in early development.

Despite these challenges we remain optimistic on the potential and outcomes for learning analytics in understanding the patterns of participation and engagement in this and other courses. Upon completion of our more intensive analyses and the development of the processes for sharing these results professionally through several conferences and publications. Specifically, we will be submitting our analyses to the 2017 Learning Analytics and Knowledge conference and will begin investigating potential academic journals with which to share our outcomes. Additionally, we will continue to refine and develop platforms to share these analytics through credentialing artifacts. We also intend to incorporate these more detailed analytics into the discussed courses in addition to other online contexts as feedback to the students. In previous courses, we provided rather broad indicators of student participation and engagement. These methods primarily used the survey systems built into the learning management systems. We believe that offering these more detailed analytics can enable participants to further engage with the material and content of these courses.

References
engagement: Explaining an emergent argument in a community of learners classroom.  
*Cognition and Instruction, 20*(4), 399–483.

International examinations and extensions of the productive disciplinary engagement framework.  


*Journal of the Learning Sciences, 22*(1), 163-166.


http://dx.doi.org/10.1145/2567574.2567626.


Figure 2. Networking Graph of Intro to EDS Discussions
Educational Data Sciences

Figure 3. Participatory Engagement Badge
In order to receive this final badge, participants had to earn badges in the following topics: Educational Data Mining; Learning Analytics; Learner Analytics; Academic/Institutional Analytics; Systemic/Instructional Improvements.

Educational Data Sciences

Daniel Hickey: Indiana University

Issued Jul 29, 2015

The recipient of this badge participated in the Big Open Online Course Introduction to Educational Data Sciences. This was a graduate-level course offered by the School of Education at Indiana University taught by Professor Daniel Hickey. In the context of this course, this individual demonstrated the ability to articulate and discuss the following aspects of Educational Data Sciences in relevant professional contexts.

In order to receive this badge, participants had to earn the following badges:
* Educational Data Mining
* Learning Analytics
* Learner Analytics
* Academic/Institutional Analytics
* Systemic/Instructional Improvements

The evidence of these competencies are the completed course wikifolios linked below, including the reflections and threaded discussions with peers, instructor, author, and/or guest discussants. This individual also demonstrated the ability to locate additional articles or other resources about learning analytics, social learning analytics, and social learning that were directly relevant to their professional context and goals. The evidence of this competency are the references included in the wikifolios.

Evidence

Figure 4. Course Badge Earned by Earning all Five Module Badges
Educational Data Sciences

In order to receive this final badge, participants had to earn badges in the following topics: Educational Data Mining; Learning Analytics; Learner Analytics; Academic/Institutional Analytics; Systemic/Instructional Improvements.

Learning Pathway Steps

Educational Data Mining

Demonstrate the ability to articulate and discuss the following aspects of Educational Data Mining (EDM) in relevant professional contexts: Primary EDM Methods, Primary EDM Applications, Ethical Principles in EDM, Considerations for EDM as a Moral Practice.

Educational Data Mining

Daniel Hickey: Indiana University

Issued Jul 29, 2015

The recipient of this badge participated in the Big Open Online Course Introduction to Educational Data Sciences. This was a graduate-level course offered by the School of Education at Indiana University taught by Professor Daniel Hickey. In the context of this course, this individual demonstrated the ability to articulate and discuss the following aspects of Educational Data Mining (EDM) in a relevant professional contexts including his own. • Primary EDM Methods • Primary EDM Applications • Ethical Principles in EDM • Considerations for EDM as a Moral Practice The evidence of these competencies are the completed course wikifolios linked below, including the reflections and threaded discussions with peers, instructor, author, and/or guest discussants. This individual also demonstrated the ability to locate additional articles or other resources about EDM and EDM ethics that were directly relevant to their professional context and goals. The evidence of this competency are the references included in the wikifolios.

Verify
Evidence

Figure 5. Educational Data Mining Badge with Link to Wikifolio Evidence