Using System Log Data to Measure Opportunities for Learning

Andrew E. Krumm, Richard J. Vath, Richard J. Waddington, Steven Lonn, and Stephanie D. Teasley
The University of Michigan

Abstract
This paper describes conceptual and methodological issues concerning the use of log data captured by web-based systems. Log data can be defined as any information tracked by an online system related to a user’s engagement with that system. Conceptual and methodological issues related to system log data are explored using two online learning environments, a learning management system and an online professional development system. We demonstrate how system log data can support accurate estimates of learners’ access to and actual use of specific tools and resources within these two online systems.

Introduction
As more educational activities move online, data tracking instructors and learners’ use of web-based systems can provide a rich source of data for researchers. In recent years there has been steady growth in the use of system log data to better understand “individual learning from educational software, computer supported collaborative learning, computer-adaptive testing (and testing more broadly), and the factors that are associated with student failure or non-retention in courses” (Baker & Yacef, 2009, p. 6). In this paper, we discuss how system log data can be used to address a broader set of questions concerning the ways in which online systems provide students with specific opportunities to learn (Coley & Leinhardt, 1980).

Focusing on opportunities to learn, Rowan and Correnti (2009) observe that, “the key measurement problem is to obtain an estimate of the overall amount or rate of exposure to particular elements of instruction occurring over some fixed interval of time (e.g., a school year)” (p. 120). We analyze opportunities to learn in two different online
learning environments—a learning management system and an online professional development system. These systems operate at different scales, serve different clientele, and are designed to suit different purposes. Given these differences, this paper provides a unique opportunity to explore similarities and differences in measuring opportunities to learn enabled by diverse online learning environments. We present results in line with the following research question: What are the systematic sources of variation in instructors’ implementation and students’ appropriation of various opportunities to learn supported by two online learning environments?

We define opportunities to learn as uses for an online system that are intended to mediate interactions between and among teachers, students, and some form of content. By interactions, we mean the active processes of interpretation that constitute teaching and learning. Teachers interpret and represent subject matter to students, who interpret their teachers, the content, and their classmates and then respond and act. In turn, teachers interpret their students, all of this in overlapping contexts and over time. (Ball & Forzani, 2007, p. 530).

Opportunities to learn represent both the intended and experienced use of some element within an online system, such as a given tool or resource. The ways in which specific opportunities are made available to and experienced by students can have a large impact on a student’s learning, whereby variation in a student’s uptake of various opportunities affects the meaning behind learning outcomes targeted by researchers. Thus, to measure various opportunities and their impact on student learning, researchers must address, at a minimum, four factors: (1) the tools made available to users (i.e., what is intended), (2) the degree to which tools are actually used (i.e. a component of what is experienced), (3) how those tools are appropriated or reinvented by users (Rogers, 2003) (i.e., a further component of what is experienced), and (4) how the use of any tool fits within the overall
instructional dynamic of a learning environment. As we describe below, system log data can be useful in addressing the first two factors in any research project seeking to understand the use and impact of online learning systems.

Much of the conversation concerning system log data has occurred within the growing literature on educational data mining. In general, educational data mining focuses on “discovering novel and potentially useful information from large amounts of data…. [E]ducational data mining methods are often different from standard data mining methods, due to the need to explicitly account for (and the opportunities to exploit) the multi-level hierarchy and non-independence in educational data” (Baker & Yacef, 2009, p. 4). Much of educational data mining is concerned with two broad classes of technologies: computer-based/online tutoring systems and generic online learning environments mediated by, for example, learning management systems (Romero & Ventura, 2007).

Mining of tutoring systems is premised on traditional assumptions of applied cognitive science research, which addresses such tasks as modeling the problem space of a given domain and a student’s movement through that space (e.g., Corbett, Koedinger, & Hadley, 2001). A tutoring system’s log data is particularly attractive because it offers researchers a bounded context for understanding student behaviors. The context is bounded both in terms of the content covered by the system and also in terms of a user’s inability to improvise with the system beyond the designed intentions of the system. A bounded context for making sense of system log data is particularly important when considering such generic factors as “time.” Morris, Finegan, and Wu (2005), for example, found positive correlations between how long students remain on a learning management
system course site with a student’s likelihood of persisting in an online course. The utility of this finding, however, can be called into question given the lack of context for making sense of the importance of time or how students were to use that time. Time spent on a well-defined learning task housed within an online system, on the other hand, can be a meaningful variable to researchers in that it there is a context for interpreting its effect. System log data tracked by learning management systems, in general, do not have built in contextual markers for understanding students’ use of the system. Therefore, in the case of less bounded systems, contextual markers often need to be constructed using other sources of data, such as observation or questionnaire data (Macfadyen & Dawson, 2010; Romero, Ventura, & Garcia, 2008).

Below, we present two cases of generic online learning environments, discuss the data used to construct various contextual markers, and outline specific statistical modeling strategies used to accurately measure both intended and experienced aspects of online leaning system use.

Case 1: Learning Management System

Learning management systems represent a class of technologies that are widely diffused across universities in the United States (Smith, Salaway, & Caruso, 2009). Lonn and Teasley (2009) describe learning management systems as web-based systems that allow “instructors and learners to share instructional materials, make class announcements, submit and return course assignments, and communicate with each other online” (p. 686). The specific learning management system that we analyzed is based on the Sakai architecture, a community-sourced product used at over 200 educational
institutions. For this proposal, we focus our analysis on courses taught by instructors who responded to an annual questionnaire at a large and research-intensive Midwestern University across the 2008-2009 academic year. Analyses of system log data revealed significant variation in learning opportunities presented to and appropriated by students. Below, we identify specific sources of variation related to the implementation of learning opportunities using hierarchical linear regression techniques.

Methods

To quantify the variation in instructors’ implementation of the learning management system, we analyzed variation in tool use across the entire university, considering course-to-course differences, differences across courses taught by the same instructor, and courses taught within the same university school or college. For these analyses we developed multiple 3-Level hierarchical linear regression models. Within each model, the total level of activity for a tool on a course site was treated as a dependent variable where courses were modeled at Level 1, instructors at Level 2, and an instructor’s respective school or college at Level 3. The number of learners on a course site was controlled for and modeled as a Level 1 predictor. Figure 1 illustrates the percentage of variance accounted for at each level specified, above. The same model structure was used for each of seventeen tools presented in Figure 1.

---

1 http://sakaiproject.org/whos-using-sakai
**Selected Results**

As Figure 1 demonstrates, a majority of variation in tool-use resides at the course level. This finding indicates that instructors often vary their use of the learning management system from course to course, thus providing learners different opportunities across multiple courses taught by the same instructor. However, there are specific tools, such as “Polls” and “iTunes”, where a large proportion of the variation resides at the instructor level. A large proportion of variation at the instructor level may signify that instructors who use these tools use them to the same degree across multiple
courses; though, these particular tools represent some of the least implemented tools, on average, across the entire university. Furthermore, these results indicate that there are not large differences between schools and colleges, or said differently, unique patterns in use for some academic units over others.

Case 2: Online Professional Development System

In order to address the growing need for high-quality professional development (PD) of in-service teachers (e.g., Committee on Science and Mathematics Teacher Preparation, 2001), researchers and teacher educators have begun to explore the use of online PD environments (Fishman, Marx, Best, & Tal, 2003). In our work with these online systems, we are examining how to provide scalable online PD support for teachers of different backgrounds and in different contexts. As part of these efforts, we have been exploring how to use system log data to characterize teachers’ cognitive engagement with the learning activities and resources in an online PD environment. In general, we are interested in how teachers appropriate opportunities to learn that are imbedded within our system.

Data Sources

Using system log data to understand teachers’ cognitive engagement can pose significant challenges. For example, while teachers cluster into different behavior profiles based on their use of the system, questions remain about what these profiles actually mean in terms of their learning. To help in understanding learning outcomes, we combined relevant teacher-level data (i.e., changes in teacher knowledge and beliefs via
survey data; analysis of classroom teaching via video; student knowledge assessment) with system log data. By examining the relationships between activity (operationalized through the system log data) and these relevant teacher characteristics and learning outcomes, we are better able to characterize the learning experiences of teachers in online PD. For this paper, we examined a cohort of 21 science teachers who participated in a sustained online PD environment during the 2008-2009 school year. We collected pre/post surveys about teacher knowledge, pre/post assessments of students' science knowledge, and video recordings of classroom teaching in addition to the system log data from the online PD environment.

Selected Results

Analyses examining how teachers spent their time in online PD show promise in combining multiple sources of data to understand system log data. For example, we examined the amount of time teachers spent on all workshop pages that specifically address Pedagogical Content Knowledge (PCK)—knowledge that we believe is essential for successful classroom enactment (e.g., Shulman, 1987). We compared teachers’ beliefs about their preparedness to enact reform-oriented science practices prior to beginning the PD to their time spent on PCK-focused workshop pages. Teacher’s incoming beliefs and their time spent on PCK-focused pages were negatively correlated (Spearman’s r(21) = - .62, p < 0.01). This suggests that teachers who came in feeling less prepared to engage in reform practices appeared to spend more time on appropriate pages in the online PD and those who felt more prepared spent less time on specific pages. Combining multiple data
sources with system log data along along with understanding the content provided by the system helped in making sense of the generic factor of time.

Conclusion

As the above cases demonstrate, system log data can be used to identify specific sources of variation in the implementation and appropriation of opportunities to learn where online tools are used. In the case of the learning management system, several opportunities, on average, are not presented to students through the tools provided by the learning management system. In the case of the online professional development system, teachers’ perceived ability to enact reform-oriented practices is negatively correlated to the time they spend on specific web pages. These two cases, therefore, show the ways in which system log data can provide detailed information about how instructors and students interact with online systems and what those interactions may mean for learning.

While system log data can offer new opportunities for researchers to investigate online learning environments, these opportunities present new conceptual and methodological challenges. Conceptually, the use of system log data is only meaningful within some frame of reference, either provided by the system itself when the system offers a bounded context for understanding students’ engagement with the system or provided by other sources of data outside of the system that situate students’ use within broader instructional dynamics. Methodologically, system log data is often hierarchically nested within academic units, such as classrooms, schools, and colleges (see Raudenbush & Bryk, 2002). In one example presented in this paper, system log data was nested within courses, instructors (i.e., the same instructor taught multiple courses), and the school or
college in which the instructor taught. Though system log data can present several challenges to researchers, when carefully deployed, it can provide important insights into students’ engagement with online learning systems.
References


